Reflecting on the use of OMOP in microbiology

xShare WP4 Workshop, IHE Europe

2026-03-24

Ankur Krishnan, University of Heidelberg and Ecraid

ECRAID-Base

Perpetual Observational Studies (POS)

The first studies to benefit from Ecraid's infrastructure are -

- **Five Perpetual Observational Studies** (POS)
- European arm of the REMAP-CAP adaptive platform trial

A POS is a **prospective, multicentre, observational clinical study** that perpetually enrolls patients

They address key clinical research gaps, including variations in clinical practices, incidence and prevalence of IDs, AMR and associated risk factors.

Clinical research backbone, **ready to concurrently or sequentially embed studies** (observational, experimental, investigator-initiated, or commercial)

Data types: Demographic, Comorbidities, Disease-specific, Sample provenance, Diagnostic testing, Microbiology and AST, Outcomes

Additionally, **site-, laboratory- and study-level metadata**

Data sources: eCRFs, LIMS, site surveys, laboratory questionnaires

Ventilator-Associated Pneumonia in **ICUs**	**POS-VAP** Started - August 2022 7117 participants enrolled Across 40 sites
Complicated Urinary Tract Infections in **Hospitals**	**POS-cUTI** Started - October 2022 6315 participants enrolled Across 43 sites
Acute Respiratory Infections in **Emergency Rooms**	**POS-ARI-ER** Started - June 2023 4328 participants enrolled Across 42 sites
Acute Respiratory Infections in **Primary Care**	**POS-ARI-PC** Started - February 2024 1772 participants enrolled Across 11 sites
Unexplained febrile illness with unusual epidemiology and/or clinical presentation in **Emergency Rooms**	**POS-Disease X** Started - December 2023 59 participants enrolled Across 8 sites
Community-Acquired Pneumonia in **Hospitals**	**REMAP-CAP (Europe)** Started - March 2018 8750 participants enrolled Across 131 sites

Why? → Harmonization and Standardization

Open Science and FAIR compliance

- ECRAID-Base operates under the maxims: *As open as possible and as closed as necessary* and *FAIR-by-design*

Harmonization within the project (between the POSs)

For what purpose?

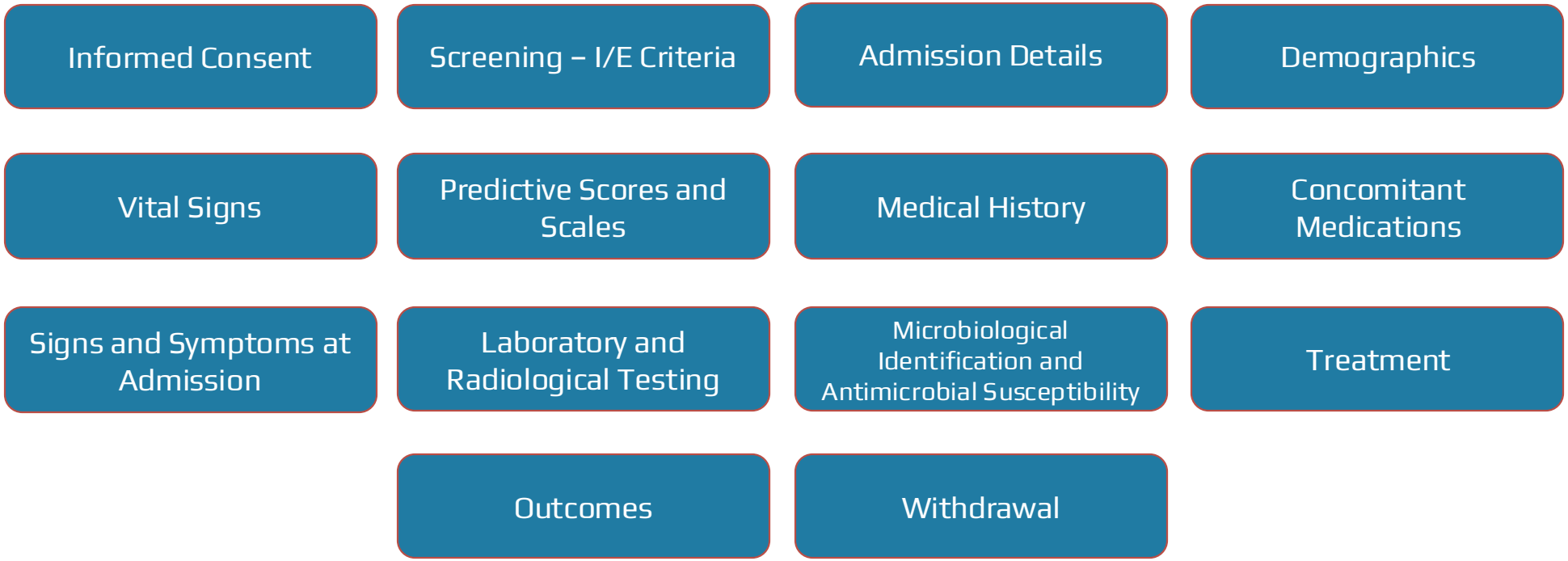
- To be compliant to regulatory submission requirements → CDISC
- To enable data reuse through federated data networks → OMOP/OHDSI
  - Privacy-by-design
  - Existing community, infrastructure, support systems, tooling and analytical packagess
  - Leverage for internal use as well

Harmonization - ‘Central library’ approach

Harmonized the data elements in the different studies

Identified common domains and data elements. Populated them with variables from CDISC template eCRF, ISARIC COVID-19 CRFs and input from study teams

Created master lists of controlled terminologies mapped to existing ontologies



Planning and decision-making

Proactive

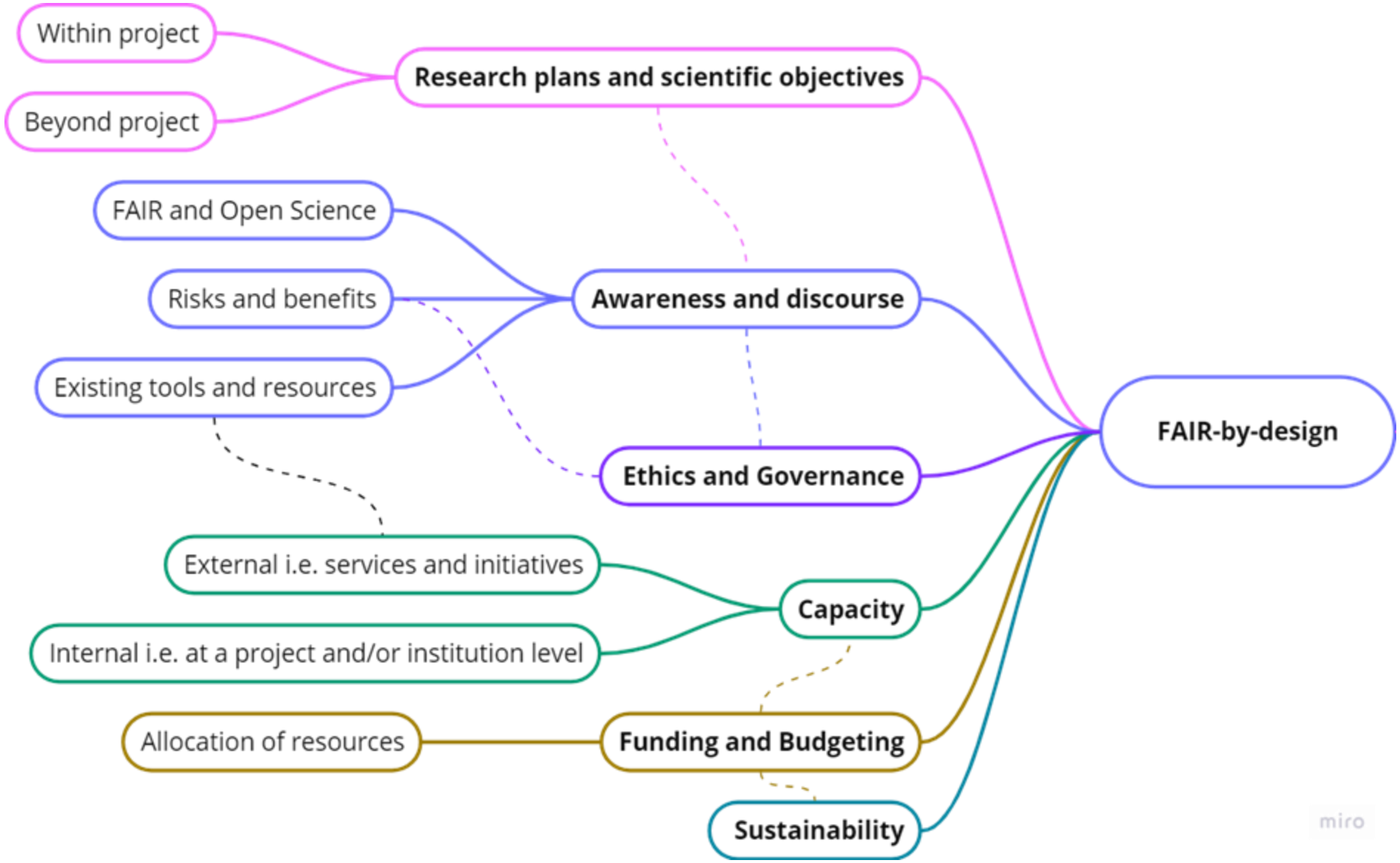
Prospective

Purpose-driven

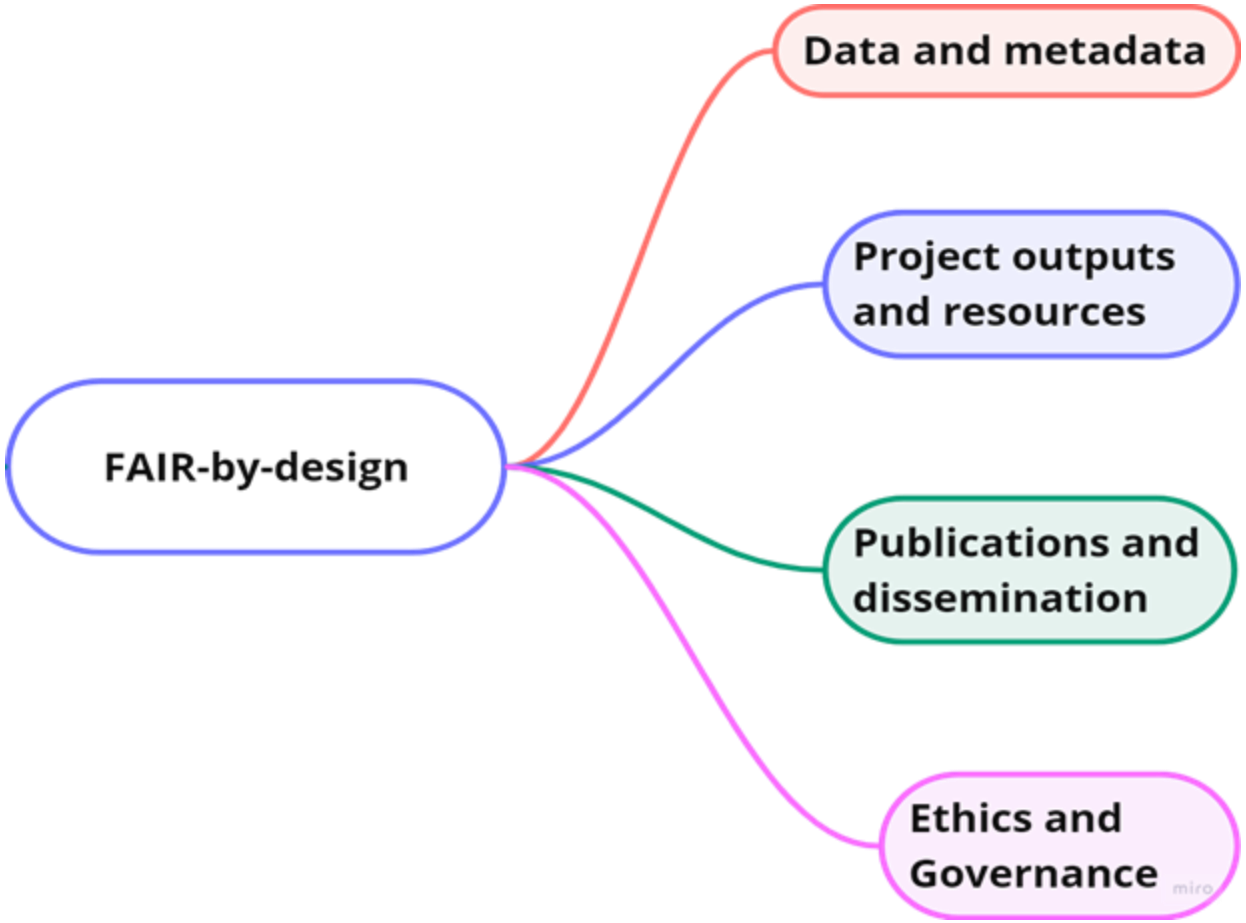
Representing different stakeholders, expertise and perspectives

Harmonization v/s Standardization

Extent of prospective standardization possible



Planning and decision-making



cdisc
OHDSI EUROPE
Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

OpenAIRE
argos
zenodo
GitHub

OPEN ACCESS
ClinicalTrials.gov
zenodo

Broad consent for future use



CIOMS
WMA

GDPR

Implementation - OMOP/OHDSI Journey

Transforming POS data to OMOP-CDM

Funding

European Health Data & Evidence Network (EHDEN)

Joined through the EHDEN 5th Open Call for Data Partners

Received funding to transform POS-VAP data to OMOP-CDM

Expertise

Leveraged the network of EHDEN-certified SMEs

Partnered with **edenceHealth NV** to develop and deploy the Extract - Transform - Load (ETL) process

Support from EHDEN and OHDSI communities

Implementation - OMOP/OHDSI Journey

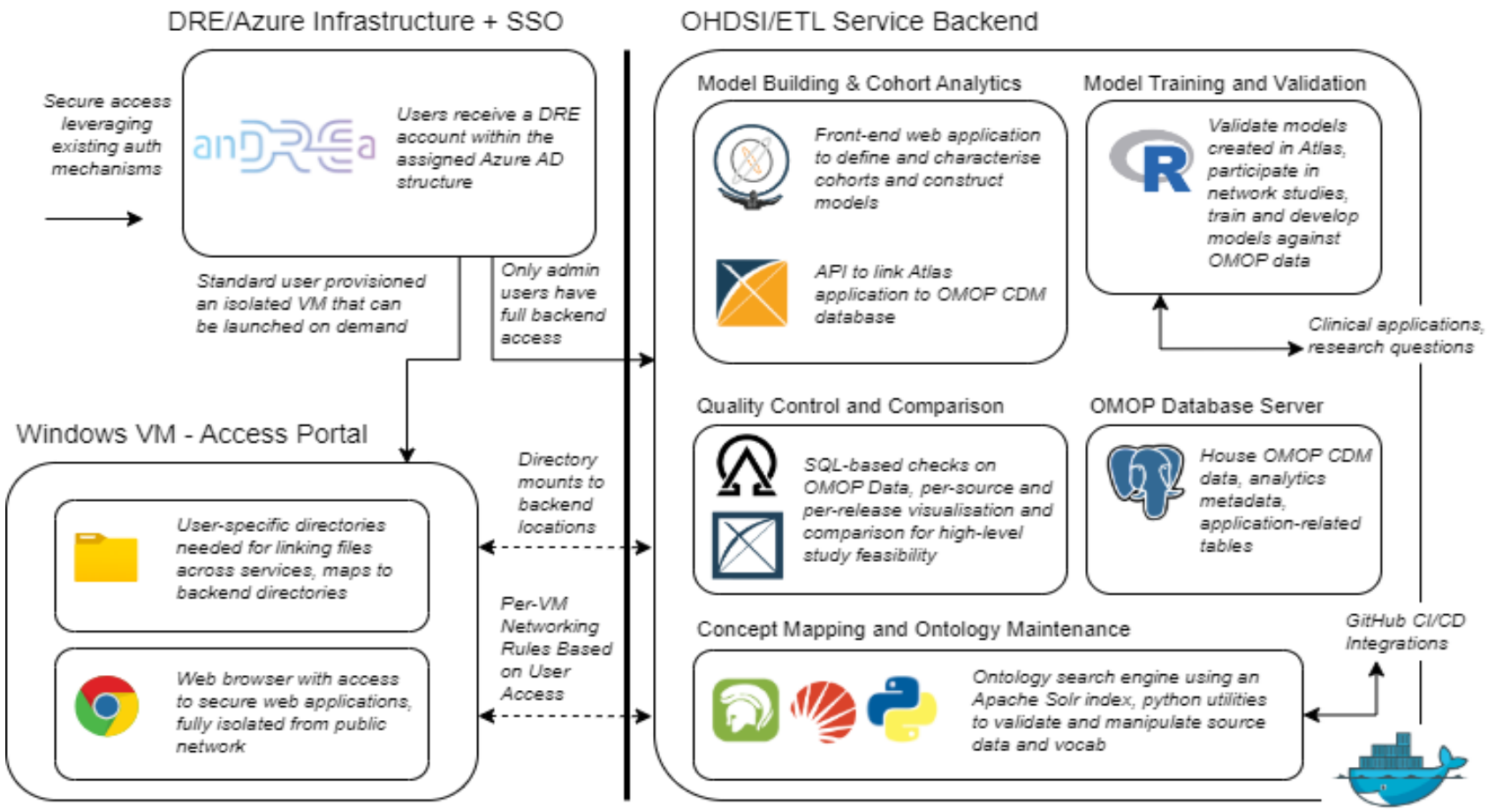
Technical infrastructure

Scalable

Sustainable

Secure

Architecture to deploy and maintain ETL pipelines and OHDSI tooling



Mapping ECRAID-Base data to OMOP-CDM

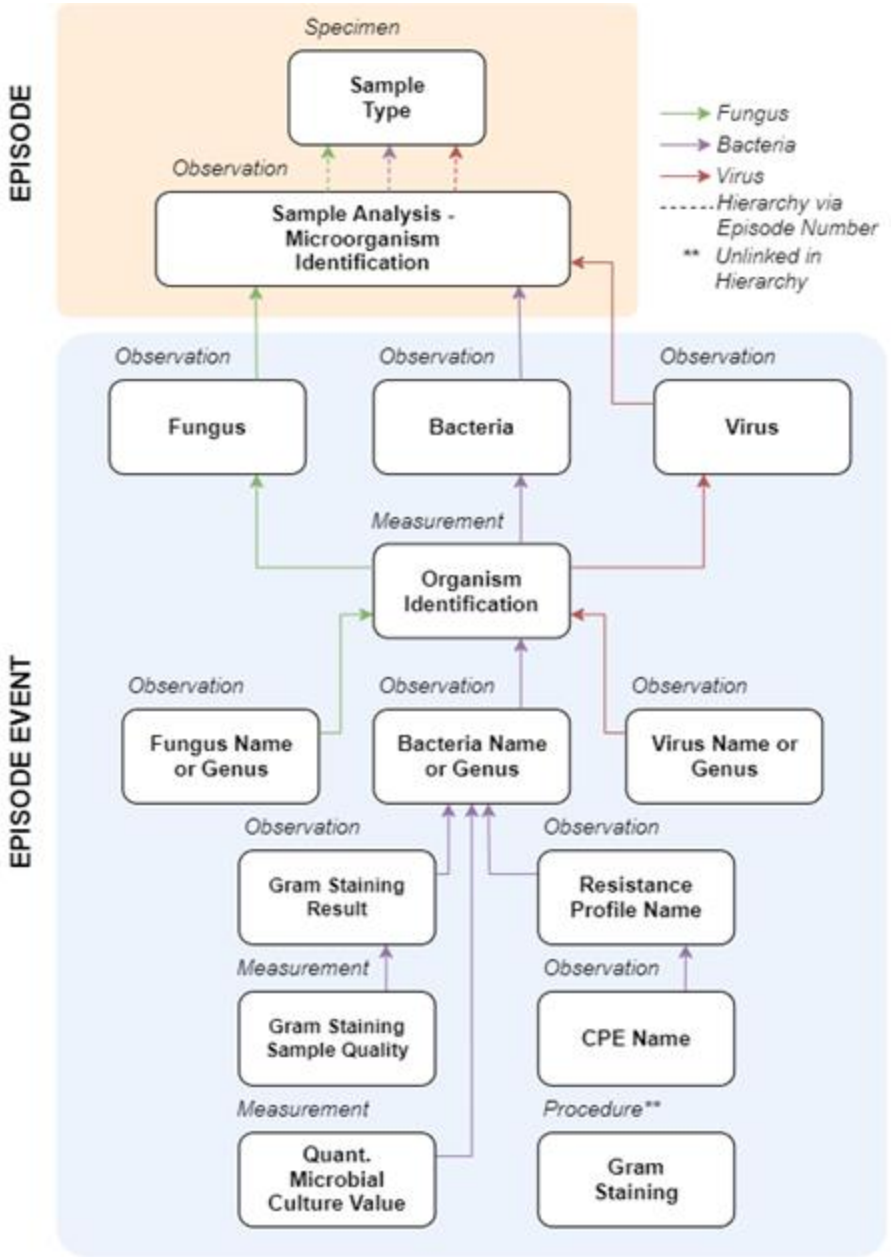
Semantic and Structural mapping

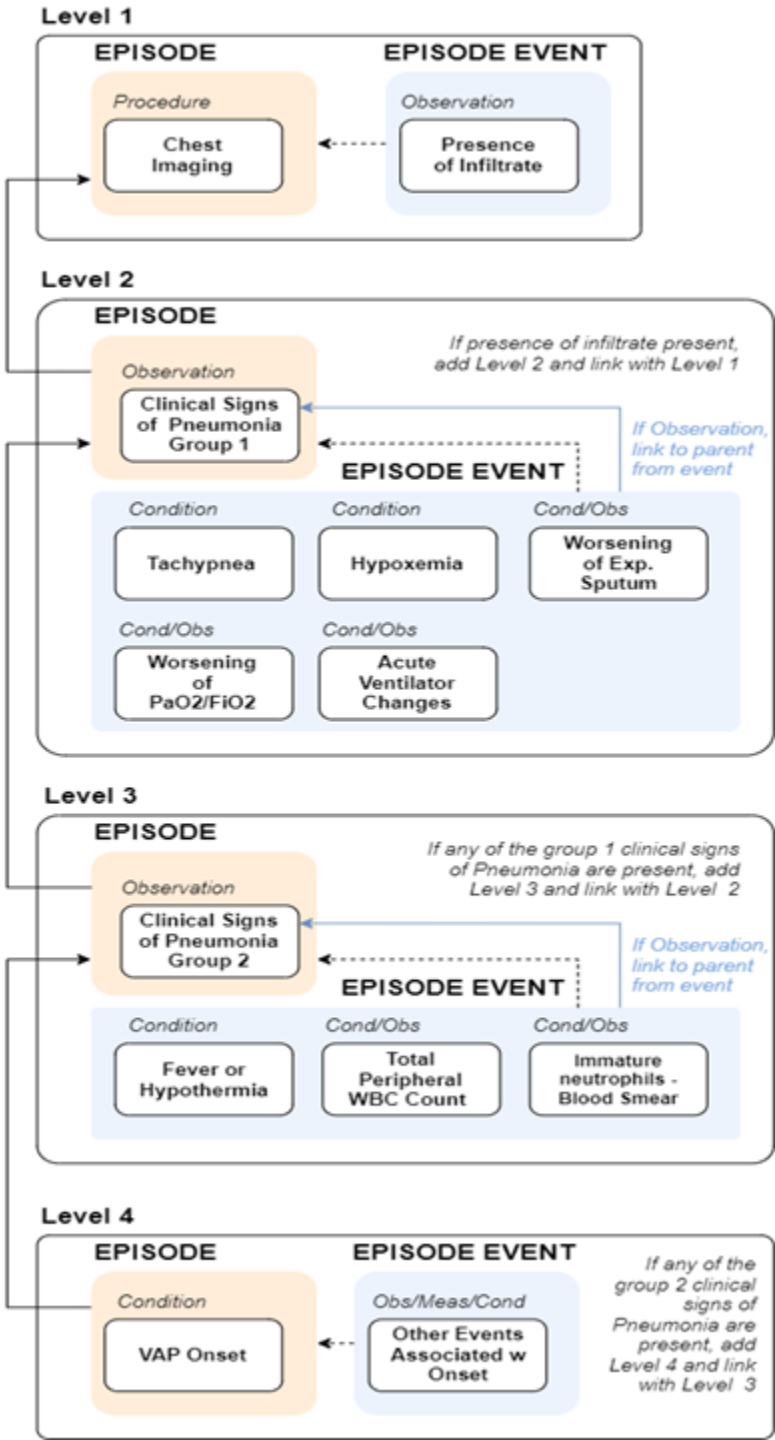
Patient-level ID/AMR data is hierarchical and inter-event dependent

- Resistance profiles
- Microbiological identification → cure
- Onset → clinical cure

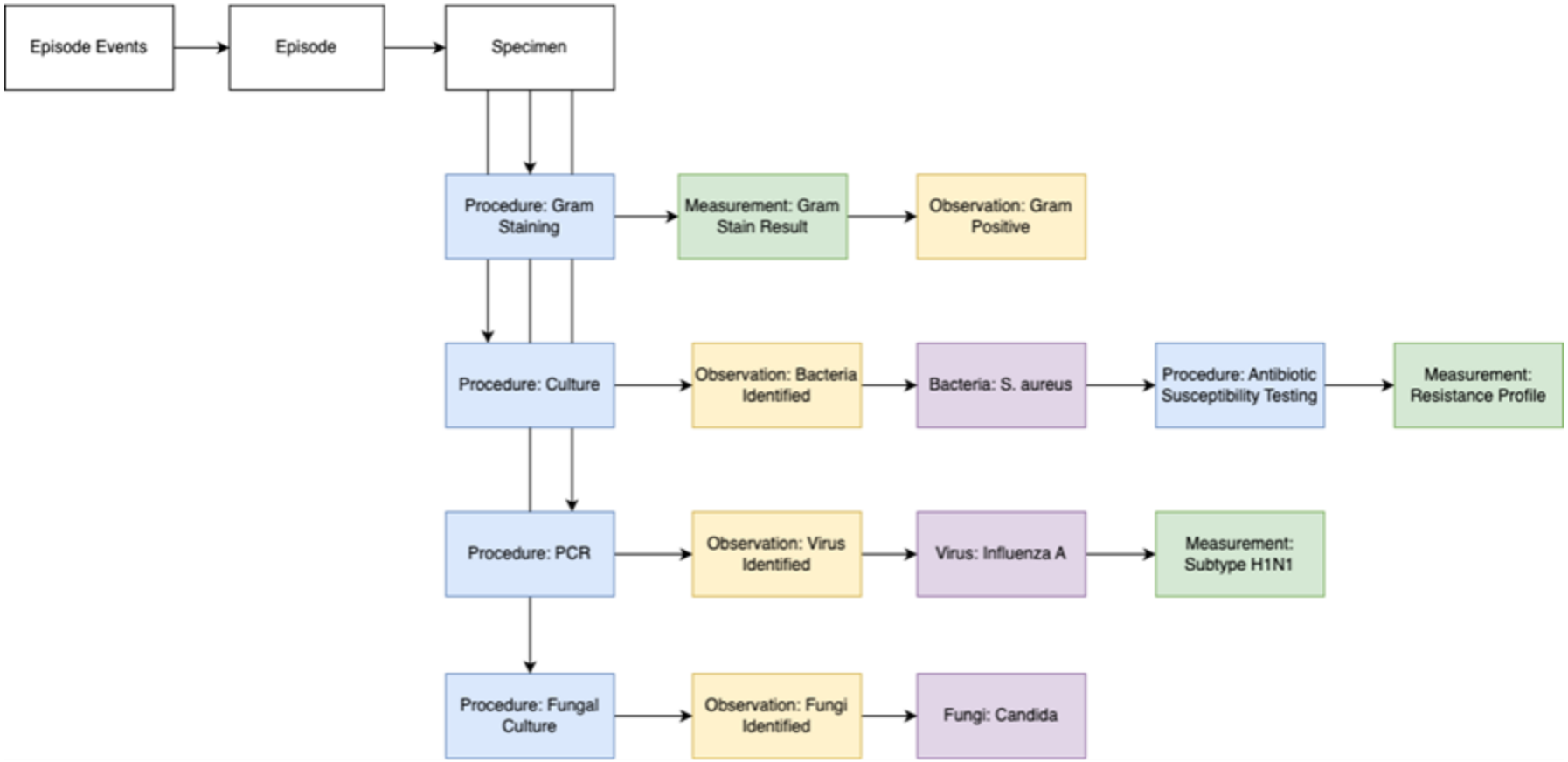
Leveraged the 'Oncology extension' (OMOP-CDM v5.4)

Does not capture all the innate nuances and complexities for AMR data and metadata (sample provenance, EUCAST/CLSI/Other guidelines, etc.)





Microbiology data: nested structure



OMOP-CDM currently does not easily accommodate nested, hierarchical structures

Vocabulary and concept mapping

<u>Disease area, healthcare setting and study-specific variables and concepts</u>

- Examples -
  - Predefined list of 'primary reason of admission'
    - 'Stroke (ischemic, hemorrhagic, subarachnoid hemorrhage)' variable we decided to map to the higher level concept 'Cerebrovascular accident'
  - Antibiogram/Resistance profiles

POS-cUTI	
Vancomycin	•S = S •I = I •R = R

POS-VAP	
Vancomycin susceptibility of S. aureus (VRSA)	•S=S •R=R •Not tested = UNK

<u>Maintaining granularity v/s Maximizing standardization</u>

- Example -
  - Mapping 'mTOR inhibitor therapy' and 'monoclonal antibodies therapy' from our study to SNOMED/OMOP vocabulary
  - We could only find the concepts for specific mTOR inhibitor therapies such as 'Sirolimus therapy' or specific monoclonal antibody therapies such as 'Infliximab therapy' in SNOMED CT
  - The parent level concept is 'Immunosuppressant drug therapy', which was found to be too broad

<u>Mapping 'Master lists'</u>

- We developed 'master lists' of bacteria, viruses, fungi (NCBI taxonomy ID) and antimicrobials (ATC)
- edenceHealth team developed a python script to auto-map the items to concepts in OMOP vocabulary

Vocabulary and concept mapping

Concepts not yet pulled into Athena

- Example - Calcineurin inhibitor therapy
- Has since been added as a standard concept in Athena

Non-standard concepts

- Example - ‘Resistance to third generation cephalosporins (cefotaxime, cefoxitin, ceftriaxone etc.), but no ESBL confirmed’

Mapping ‘No’ and ‘Not assessed’ i.e. for variables where capturing ‘No’ or ‘Not assessed’ has clinical significance ,such as for patient management

- Comorbidity (from Charlson Comorbidity Index) or clinical conditions in the VAP clinical criteria (such as tachypnea, hypoxemia, etc.) or resistance profile was 'not present, we used the ‘No evidence of’ concept.

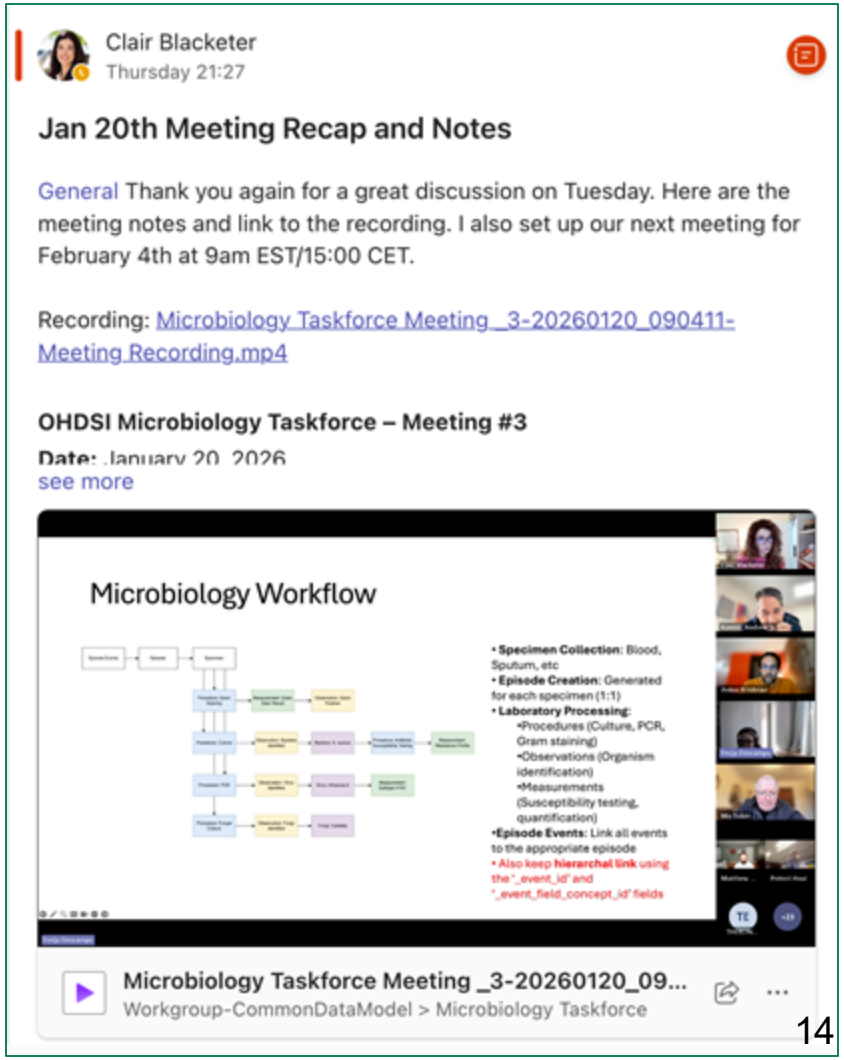
- VAP clinical criteria, microbiological screening or test for a particular resistance profile was not assessed/performed, we used the ‘Not performed’ concept

- Particular drug therapy (antibiotic or immunosuppressant) was not prescribed, we used the ‘No drug therapy prescribed’ concept

Metadata - Clinical guidelines (breakpoints), site/lab-specific, diagnostics-related, Discovery-Access-Provenance,

Microbiology data: the OHDSI task force

- Multiple teams and projects face the difficulty of fitting microbiology data in OMOP-CDM
- Multiple solutions have been presented
    - Extending OMOP-CDM with additional, dedicated tables
    - Using OMOP-CDM 5.4 linking columns to allow hierarchy
    - Adding additional logic using episode and episode_event tables to cluster
- Microbiology task force created to discuss and design global solution together
    - Teams are presenting and discussing their solutions



Conclusion

A few cliches -

'*We do not need to reinvent the wheel*

- OHDSI community, tooling, analysis packages, CDM, WGs, governance, agreements and cross-walks with other SDOs and their communities
- Leverage other open-science communities, platforms and tools

'*Harmonizing the harmonization approaches*'

- OHDSI Microbiology taskforce → WG; conventions, guidelines and tooling; governance
- Align with similar efforts in HL7 FHIR community and so on…

'Being mindful of going from a smaller bubble to a bigger bubble but a bubble nonetheless'

- Cross-walks, coordination and alignment
- No standard or ontology (and its community) can do this by itself

Discussion points

Benefits of communication and cross-walks between standards, ontologies and formats

- Coordination and alignment of interoperability conventions and practices within and across the communities and stakeholders
  - It is not data exchange v/s analysis v/s compliance i.e. it is not FHIR v/s OMOP v/s CDISC

- Driven by and enabling wider set of use-cases by connecting various data types and sources at necessary levels of granularity and complexity

- Leverage existing cross-walks between FHIR -> OMOP and FHIR -> CDISC
  - HL7 FHIR as a the middle ground to enable analytical layer and/or compliance (regulatory) layer

- Efficient and judicious data collection
  - By reducing the need to duplicate information from hospital EHR data into CRFs
  - Only study-specific data elements will need to be collected thereby, reducing the burden of data collection for the sites

Thank you for your attention!

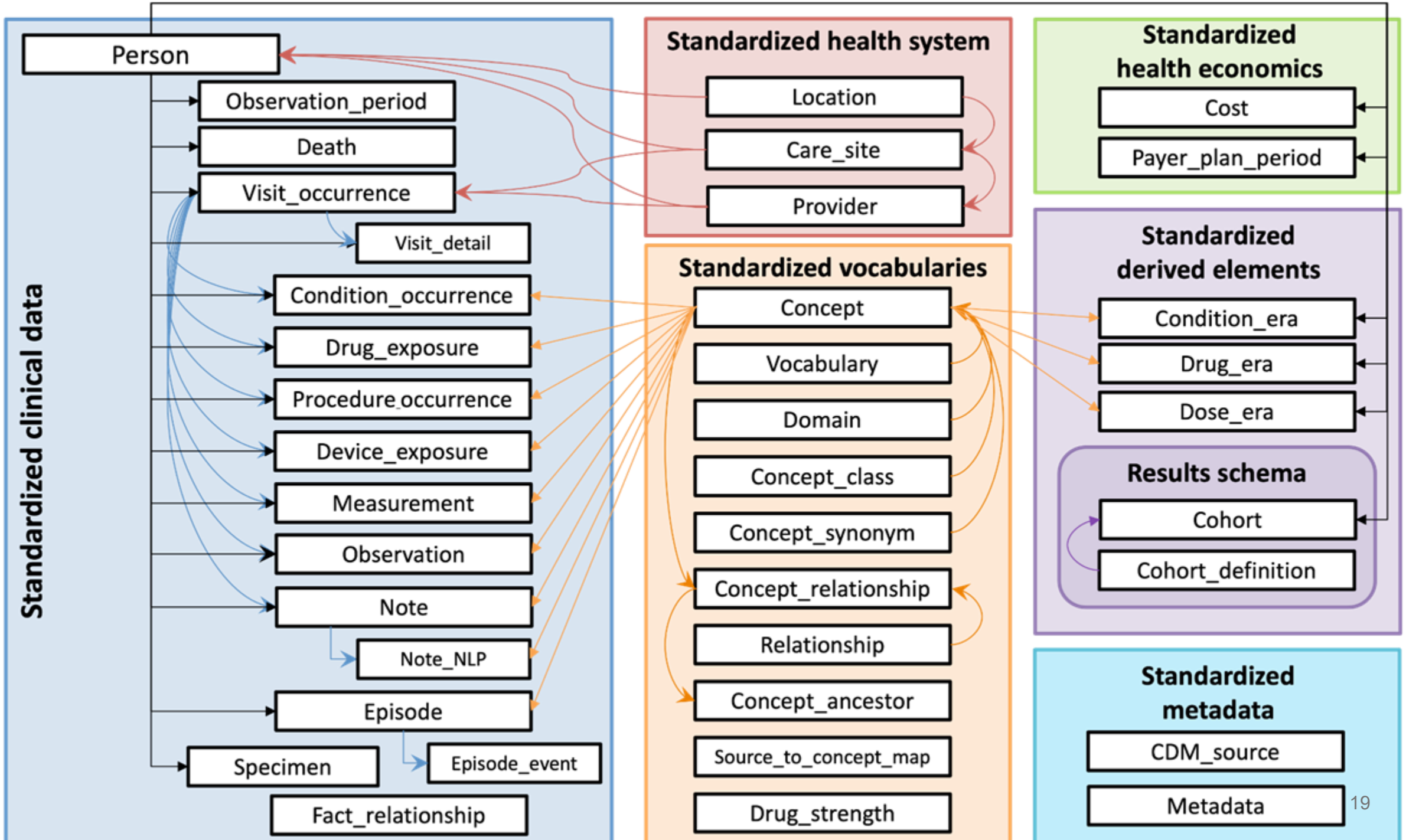
For any questions, please contact:

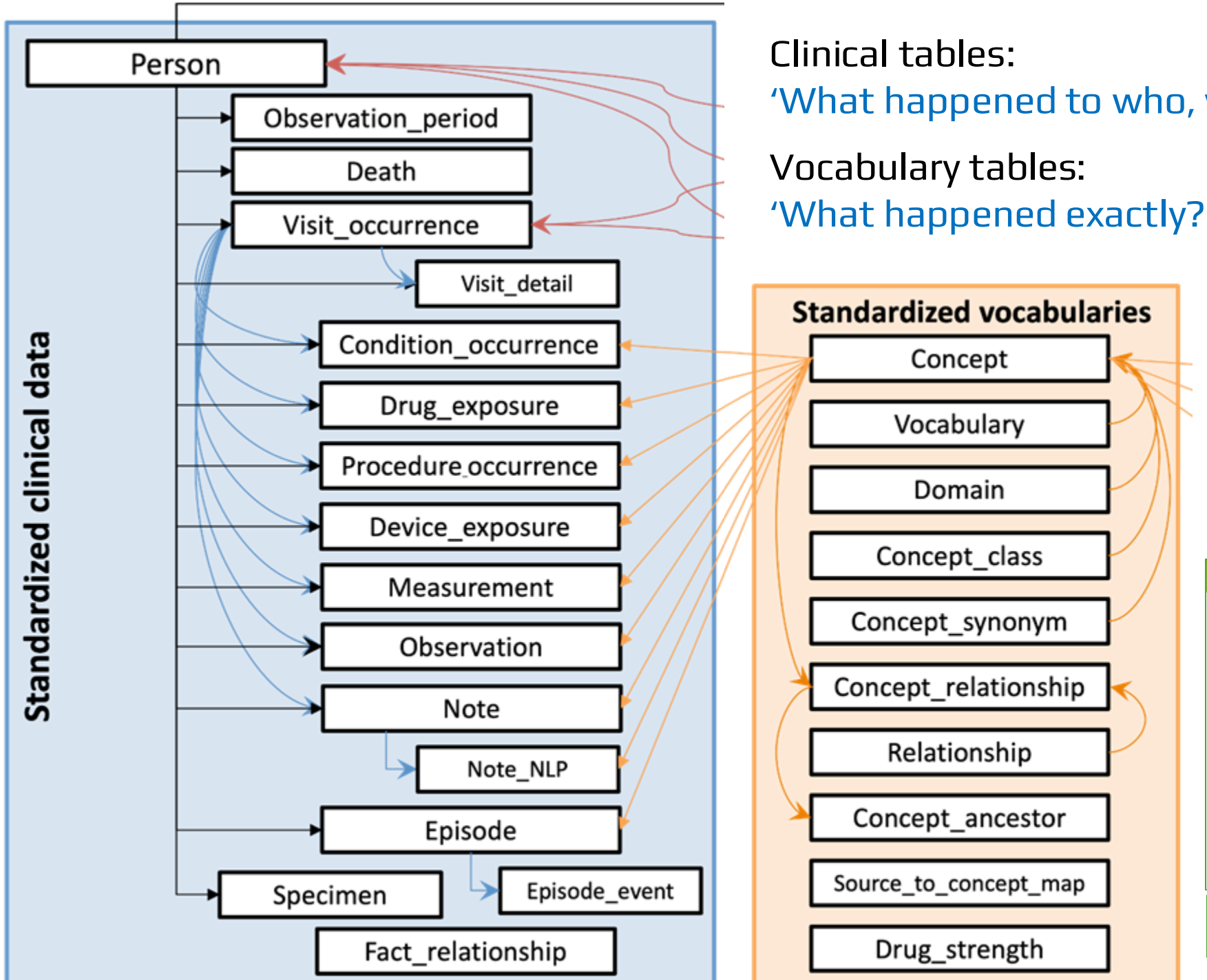
Ankur Krishnan - ankur.krishnan@uni-heidelberg.de

ecraid Base

The ECRAID-Base project has received funding from the European Union's Horizon 2020 Research and Innovation programme, under the Grant Agreement number 965313.

Backup slides


Standardized clinical data
Person
Observation_period
Death
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Observation
Note
Note_NLP
Episode
Episode_event
Specimen
Fact_relationship
Standardized health system
Location
Care_site
Provider
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_synonym
Concept_relationship
Relationship
Concept_ancestor
Source_to_concept_map
Drug_strength
Standardized health economics
Cost
Payer_plan_period
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized metadata
CDM_source
Metadata



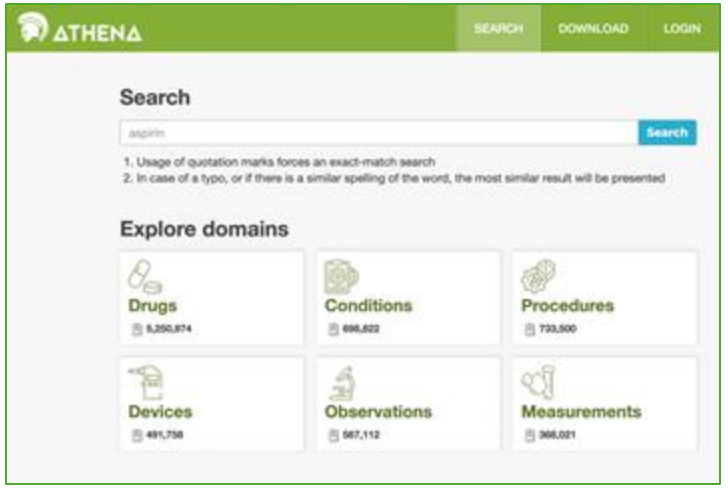
Clinical tables:

'What happened to who, when and in what context?'

Vocabulary tables:

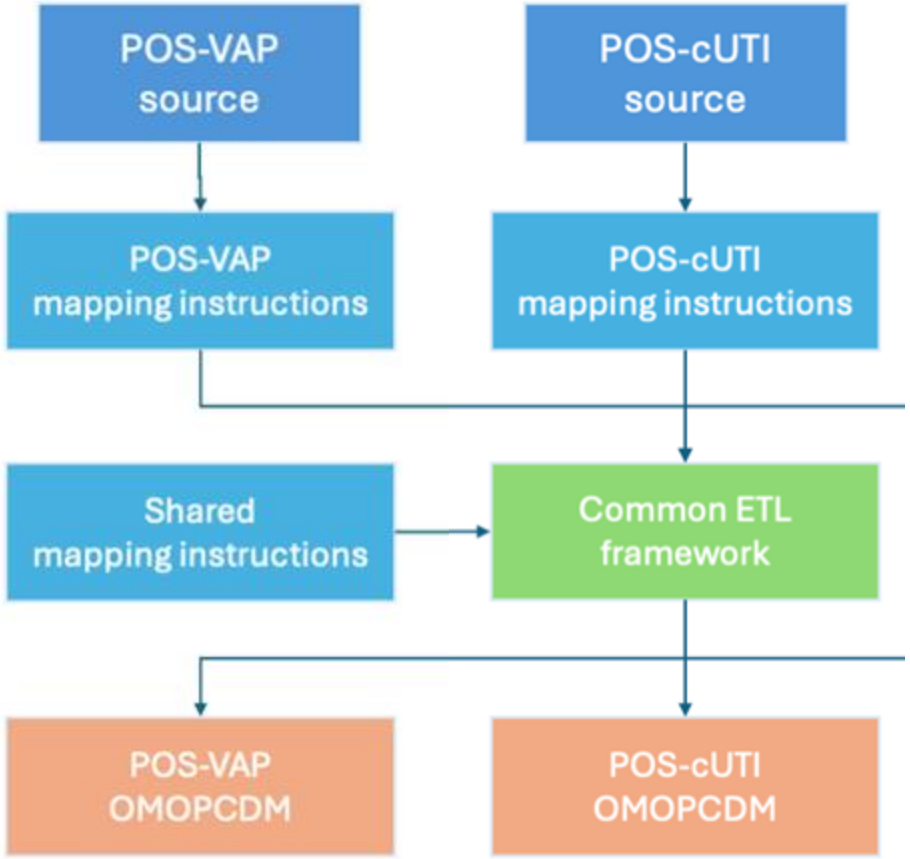
'What happened exactly?'

Example: methicillin (RxNorm) has OMOP concept_id 1361385



https://athena.ohdsi.org

POS-VAP and POS-CUTI



ETL framework:

- load source data
- load pivot/mapping instructions
- fill pre-pivot tables
- pivot and write to OMOP

Advantages:

- no need to update ETL code to add/update/correct mapping
- Can be re-used for different registries
- Common mappings can be shared

Federated studies on OMOP-CDM network

