Bridging communities: Transforming MIMIC-IV to the OMOP CDM

👤 PRESENTER: **Michael** Kallfelz

INTRO:

The PhysioNet institution provides researchers access to several very valuable datasets, among which MIMIC is the most prominent. It requires however some technical background to perform analyses and **allows research only within siloes**. This project aimed at, in a collaborative approach, making the latest MIMIC content available as OMOP CDM so that the research community can be extended into OHDSI, using the OHDSI provided tools. It was also meant to provide a template for transforming rich ICU derived data, including biosignal data as waveforms.

METHODS

1. A Google Cloud Platform® (GCP) environment with a BigQuery instance was provided. Access to the MIMIC IV BigQuery repository was established
   a) The ETL process design was inspired by previous projects [1] and based on a combination of Python and SQL scripts
   b) The delivery of the OMOP CDM and operation was done within the GCP
2. Waveform data was processed from the original MIMIC repository and a process designed to extract structured information and integrate the biosignal data with other clinical data.
3. An ATLAS instance was created for reviewing ACHILLES results, performing concept coverage counts and cohort analysis as User Acceptance Testing

RESULTS

The entire ETL logic, mappings and documentation are available in an OHDSI github repository [2] and the transformed data can be retrieved from PhysioNet. As found in UAT, some areas of source data still need to be included to achieve full coverage.

Combining the power of **two research organizations** to enable extended **research opportunities** and explore handling of **biosignal data.**

Take a picture to access the showcase repository or follow **this link**

Findings

- The collaborative approach that was taken between the PhysioNet and OMOP teams helped in mutually understanding opportunities as well as limitations and addressing them.
- Work with the waveform material in wfdb format, its conversion and extraction of additional knowledge and integration with clinical data in the OMOP CDM generated new insights for the PhysioNet team.
- For integrating the waveform, an approach was chosen making use of the visit detail table. While it is debatable if this approach can be applied widely, it supports the discussion around a general approach for integrating non OMOP objects (such as Imaging data) with classic OMOP CDM tables.
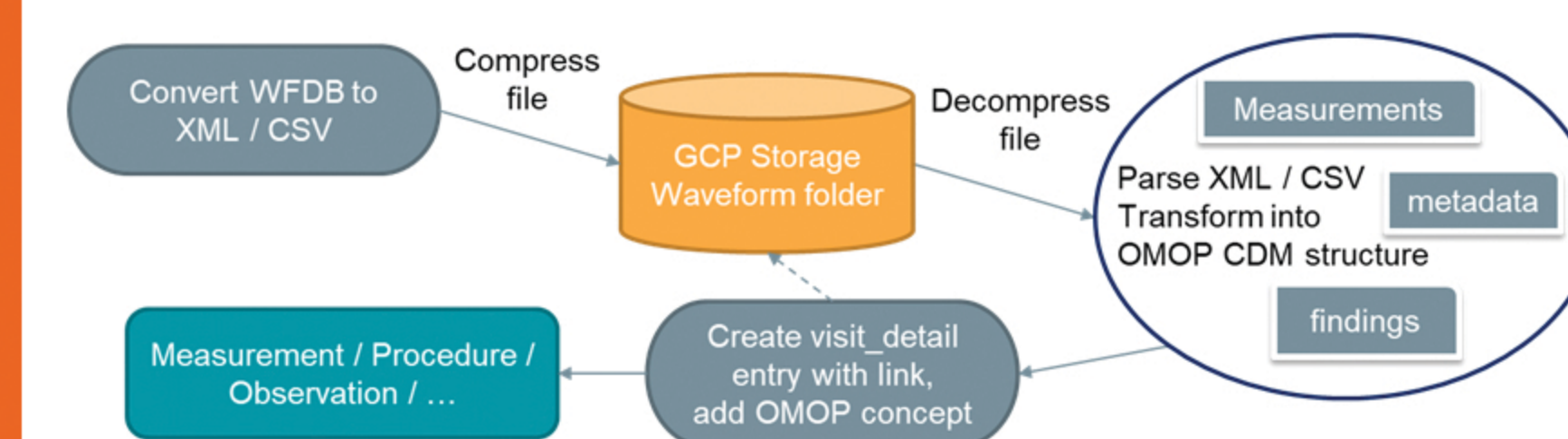
Conclusions

- Borrowing strength from each other has led to a very fruitful collaboration between the communities, resulting in a deeper understanding of the dataset and better results
- More ICU datasets at PhysioNet should be explored to achieve the ability to reproduce cohort analysis in multiple datasets
- These next steps are suggested:
  - Improve coverage of source data (such as inputevents)
  - Implement ways of properly integrating objects into the OMOP CDM
  - Collaborate with PhysioNet to improve waveform data processing

References:
[1] Paris N, Parrot A.
[2] MIMIC github

👤 AUTHORS: Michael Kallfelz[1], Andrew Williams[2], Tom Pollard[3], Gigi Lipori[1], Anna Tsetkova[1], Manlik Kwong[2], Jeff Osborn[5] , Vojtech Huser[6]
[1] *Odysseus Data Services* | [2] *Tufts University* | [3] *Massachusetts Institute of Technology* | [4] *Univ. Florida Health* | [5] *Endpoint Health* | [6] *National Institute of Health*