# THEMIS - The Answer to Building Consistent Common Data Models Across Organizations

*Anna Ostropolets[1,2], Mui Van Zandt[1,3], Erica A. Voss[1,4,5], Asha Mahesh[4], Ron Stewart[6], Paul Petraro[7], Minnie Chou[6], Christian Reich[1,3]*

*[1] Observational Health Data Sciences and Informatics (OHDSI), New York, NY, United States; [2] Odysseus Data Services, Cambridge MA USA; [3] IQVIA, Cambridge MA USA; [4] Janssen Research & Development, LLC, Raritan, NJ, United States; [5] Erasmus University Medical Center, Rotterdam, Netherlands; [6] Amgen, Thousand Oaks, CA. USA; [7] Novo Nordisk Inc., Plainsboro, NJ, United States.*

## Background

The goal of Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is to create a standardized approach at many levels: (1) standard representation of source data, (2) unified standardized vocabularies, and (3) standardized statistical methods and tools to enable accurate, reproducible and efficient research. While there has been intensive work around the tools, vocabularies, and methods, **the Extract, Transform, & Load (ETL) policies remained opaque, giving rise to different interpretations of how to convert source data into the CDM**. This has led to the inconsistent representations of the same data sources across Observational Health Data Sciences and Informatics' (OHDSI) and inability to gather reliable and scalable evidence.

The experience of recent network studies such as FDA BEST initiative has shown that the variety of the data sources, discrepancies between different ETL approaches, and ambiguousness of OHDSI data standards may make the results of a research non-credible or non-comparable. During the project, multiple standard convention questions arose. For example, how does one classify Hospice? Some site classified this Place of Service as 'Inpatient' while others classified it as 'Outpatient'. The different interpretation of this term resulted in discrepancies in the counts and analytics.

## THEMIS Roadmap and Current Work

In order to ensure the quality and consistency across all OMOP CDM databases, an initiative was started to create a system of CDM conversion and maintenance policies. This lays the groundwork to establish a foundation of CDMs that can be used in network research studies.

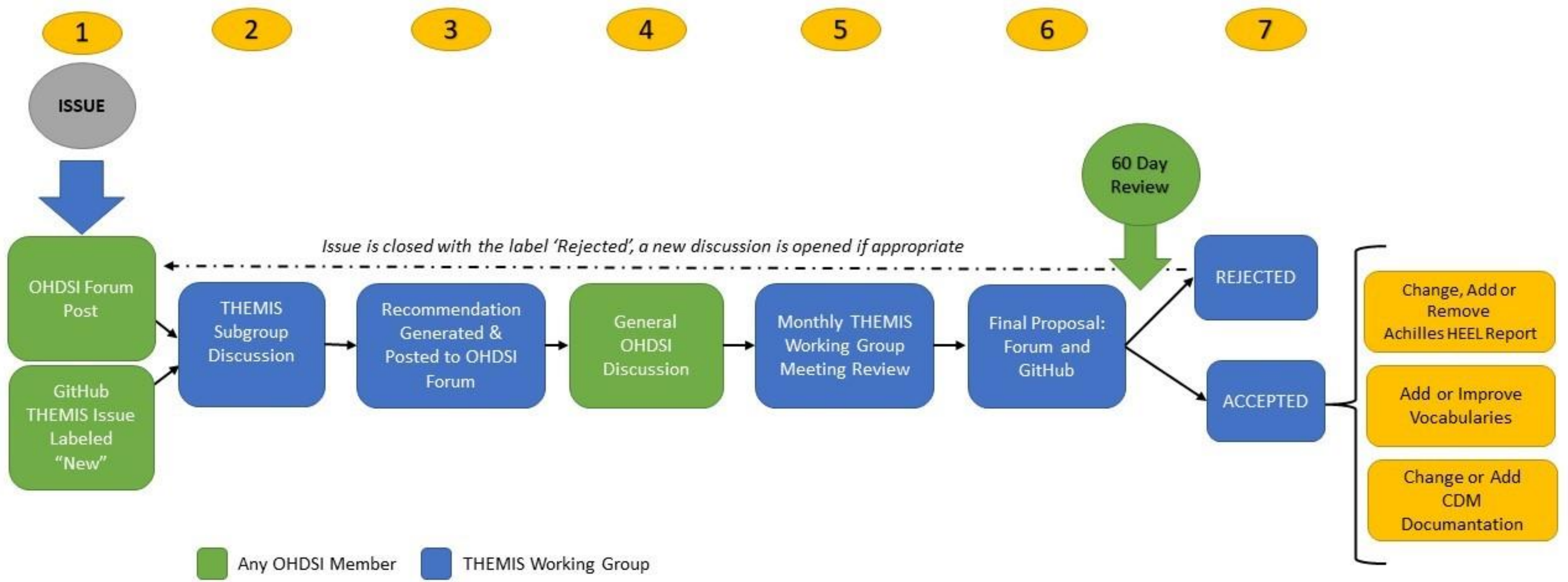
1 2 3 4 5 6 7
ISSUE
60 Day Review
Issue is closed with the label 'Rejected', a new discussion is opened if appropriate
OHDSI Forum Post
GitHub THEMIS Issue Labeled "New"
THEMIS Subgroup Discussion
Recommendation Generated & Posted to OHDSI Forum
General OHDSI Discussion
Monthly THEMIS Working Group Meeting Review
Final Proposal: Forum and GitHub
REJECTED
ACCEPTED
Change, Add or Remove Achilles HEEL Report
Add or Improve Vocabularies
Change or Add CDM Documantation
🟩 Any OHDSI Member 🟦 THEMIS Working Group

*Figure 1. THEMIS workflow*

The **THEMIS Working Group** (WG), open to anyone in the OHDSI community, works to review and analyze the structure of the CDM in order to improve the utility of the database model and to develop solutions for ETL conversions based on OHDSI member's research experiences. THEMIS works in collaboration with the CDM WG and the teams developing OHDSI tools (e.g. ACHILLES) ensuring the integration of all three aspects of standardization in OHDSI. ***Figure 1*** shows:

- The THEMIS process starts by the WG **extracting OHDSI community feedback** for areas where policies or more guidance is needed. It also allows direct requests from community members through GitHub repository (https://github.com/OHDSI/Themis).
- The THEMIS group then **internally discusses what is believed to be the best solution**.
- This **recommended solution is then posted to the OHDSI Forum** (http://forums.ohdsi.org/) to gain initial community feedback and is duplicated on the repository in the structured format.
- Once the THEMIS group believes the community has landed on the **recommended policy or guidance description it is once again posted to the community, this time under a 60-day GitHub review process.**
- Afterwards, unless there is still community disagreement, the policy and guidance changes being recommended by THEMIS will a**ugment the CDM specifications, add necessary enhancement to the OMOP Vocabularies or make updates to OHDSI tools** (e.g. improve ACHILLES HEEL warnings).

## THEMIS Roadmap and Current Work (continued)

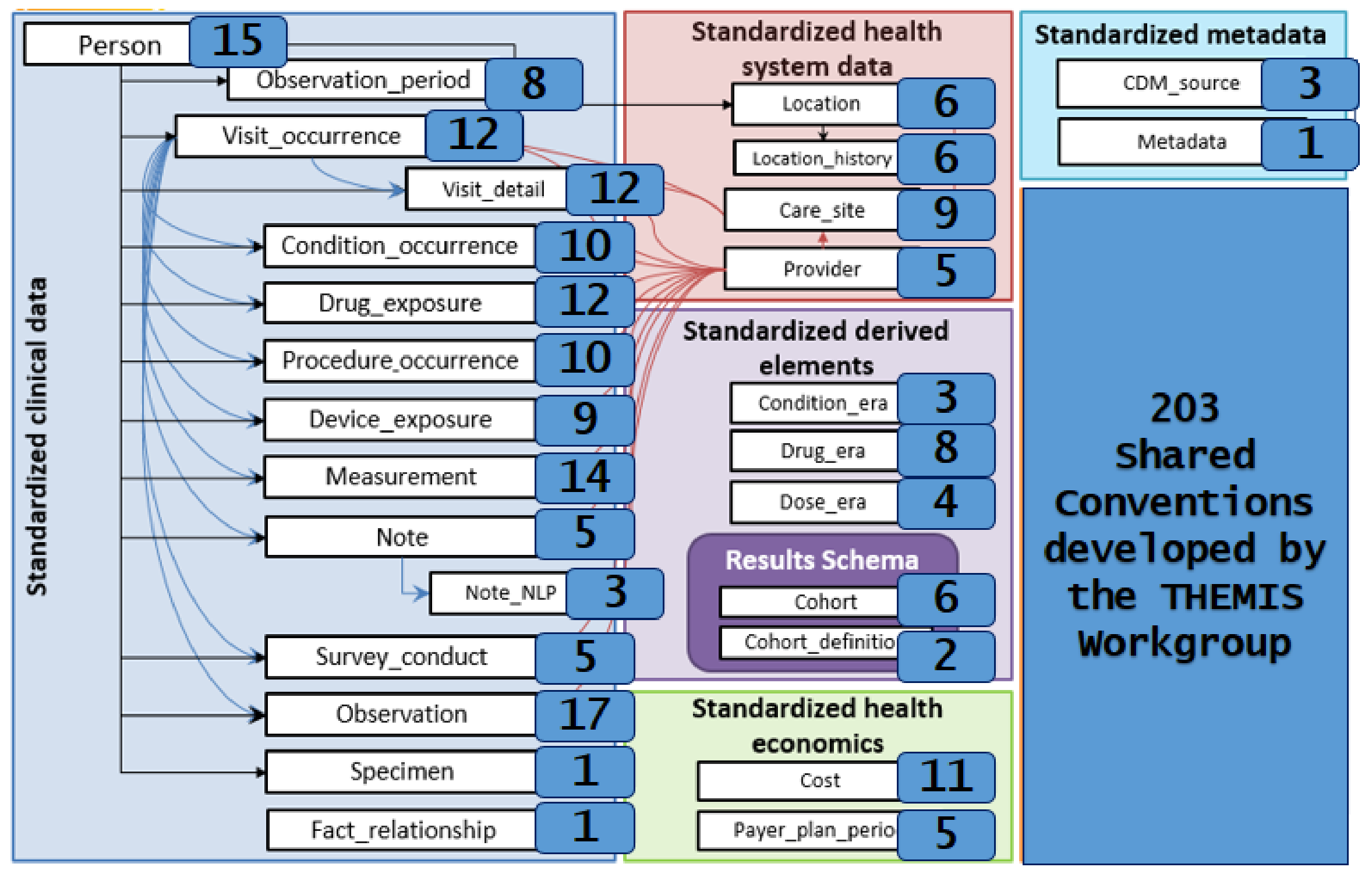
Standardized clinical data
Person 15
Observation_period 8
Visit_occurrence 12
Visit_detail 12
Condition_occurrence 10
Drug_exposure 12
Procedure_occurrence 10
Device_exposure 9
Measurement 14
Note 5
Note_NLP 3
Survey_conduct 5
Observation 17
Specimen 1
Fact_relationship 1
Standardized health system data
Location 6
Location_history 6
Care_site 9
Provider 5
Standardized derived elements
Condition_era 3
Drug_era 8
Dose_era 4
Results Schema
Cohort 6
Cohort_definitio 2
Standardized health economics
Cost 11
Payer_plan_perio 5
Standardized metadata
CDM_source 3
Metadata 1
203 Shared Conventions developed by the THEMIS Workgroup

*Figure 2. THEMIS Release 1 Conventions using v5.3.1*

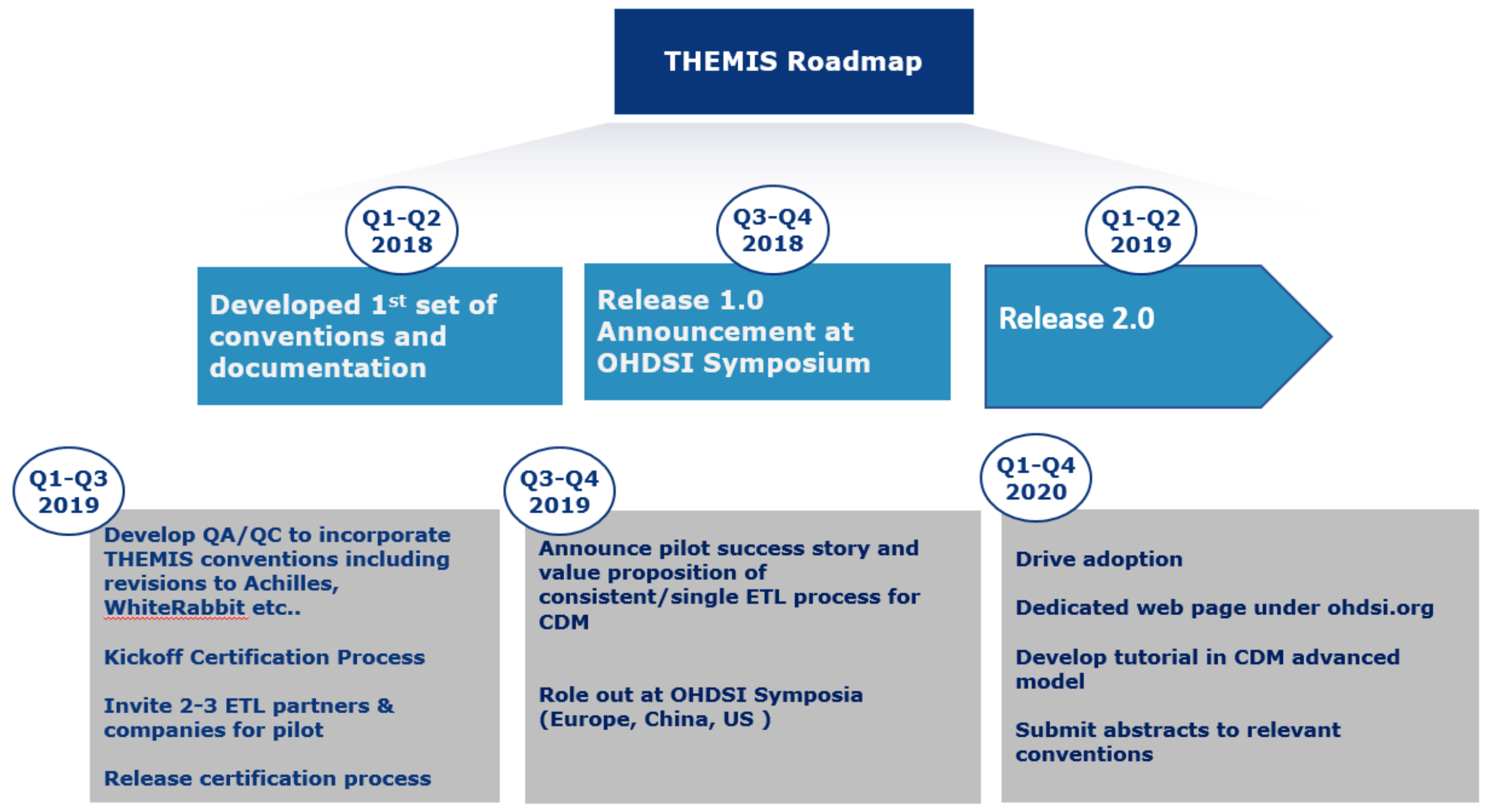
THEMIS Roadmap
Q1-Q2 2018 Developed 1st set of conventions and documentation
Q3-Q4 2018 Release 1.0 Announcement at OHDSI Symposium
Q1-Q2 2019 Release 2.0
Q1-Q3 2019
Develop QA/QC to incorporate THEMIS conventions including revisions to Achilles, WhiteRabbit etc..
Kickoff Certification Process
Invite 2-3 ETL partners & companies for pilot
Release certification process
Q3-Q4 2019
Announce pilot success story and value proposition of consistent/single ETL process for CDM
Role out at OHDSI Symposia (Europe, China, US )
Q1-Q4 2020
Drive adoption
Dedicated web page under ohdsi.org
Develop tutorial in CDM advanced model
Submit abstracts to relevant conventions

*Figure 3. THEMIS Roadmap*

***Figure 2*** shows how many conventions are reviewed and managed by the THEMIS working group.

*The issues solved include:*

- ***Events that fall out of OBSERVATION_PERIOD***
  The decision was to allow events to fall out OP, it is not recommended to use the events that are outside the observation period to identify a cohort. Payer plan period should be used to capture coverage (including partial e.g. Medicare Part D).
- ***Duplicates in DRUG_EXPOSURE, DEVICE EXPOSURE and PROCEDURE_OCCURRENCE***
  ETL should not dedupe multiple records of the same device, drug or procedure that occurred on the same day unless there is a reason to believe the item is a true data duplicate. The source of the data, modifiers, claim types etc. should be considered.
- ***Multiple Deaths***
  We recommend to use only one death date per individual If a patient has clinical activity (e.g. prescriptions filled, labs performed, etc.) for a time period that is longer than sixty days after death, it is possible to drop the death record.
- ***Negative values in measurements***
  Negative values are not allowed in MEASUREMENT table, except for the measurement of base excess and QRS axis. If source data contains negative values for positive measurements, then set it to NULL.

## THEMIS Next Steps

ACHILLES is part and parcel of the tools that used in CDM dataset quality assurance including the notifications about the OMOP CDM violations and warnings that allow conversion enhancement and improvement. From this point of view, **THEMIS regulations will be embedded into ACHILLES report indicating the rules data vendors should obey**. To ensure the transparency, these reports will be published on the forum, so that the community can evaluate them and consider appropriate or inappropriate to use in network studies. This Certification will be obligatory for the participants in the research projects and will be performed on a regular basis. The next step will be to publish the list of certified and validated sites in order to facilitate cross-dataset research.
In addition, the THEMIS working group will work through OHDSI to engage with key vendors of observational data to obtain general agreement of the OMOP CDM such that the data vendors acquire sufficient expertise as to complete the OMOP CDM ETL prior to providing the data to OHDSI members.

***Figure 3*** highlights the THEMIS road map.

## Conclusions

**THEMIS is OHDSI initiative that serves the purpose of standardizing the ETL process**. It enables transparent and reproducible research by creating policies and conventions that regulate the structure, quality, and content of converted CDM datasets. Future steps include certification and validation of sites through publishing their ACHILLES reports, as well as enhancement of community involvement and knowledge spreading. By having such standardization, studies across multiple organizing using the OMOP CDM will result in more accurate and consistent results.