2025 4th Annual OHDSI Open Source Developers Conference Highlights

The code of Science is Open Source

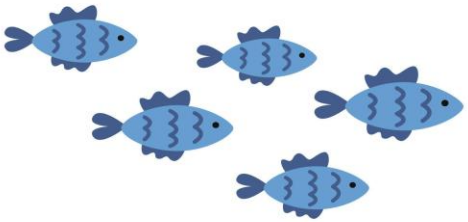
- Katy Sadowski
- Sean O'Reilly
- Robert Miller
- John Gresh
- Daniel Smith
- Jared Houghtaling
- Hayden Spence
- Kyle Zollo-Venecek
- Clark C. Evans
- Paul Nagy

Special thanks to our organizers

Open source is fundamental to science, which makes scientific research transparent, accessible, and reproducible for the benefit of both the scientific community and society as a whole

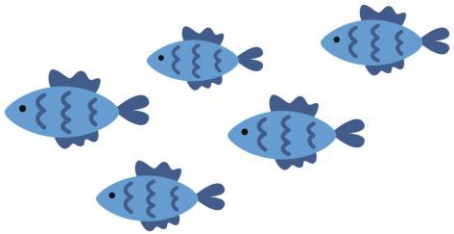
OHDSI Software Ecosystem

Data Management
Getting data into the OMOP CDM

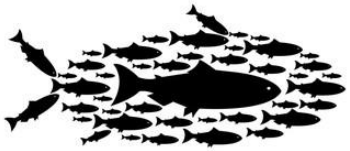


Atlas

Data Analysis
Performing analysis on the CDM



Hades



OHDSI Vocabulary

OMOP Common Data Model

2025 State of the OHDSI Open Source Ecosystem

	2012 - 3/2024	2012 - 3/2025 Diff
Repositories	284	
Developers	611	
Issue Submitters	1,082	
Issues Submitted	10,812	
Commits	50,575	

Data analysis of Github contributions for the OHDSI Organization: Source github via bitergia

Why contribute to Open Source

- **Learning and Skill Improvement:** Contributing to open source projects provides a fantastic platform to hone coding skills, learn new technologies, and gain practical experience working on real-world projects.
- **Sense of Accomplishment:** Seeing their code used by others and contributing to a successful project provides a tangible sense of achievement and pride.
- **Being part of a community**: Open source communities provide a space for developers to connect with like-minded individuals, collaborate, and gain recognition for their work.
- **Building a Professional Reputation:** Contributing to well-regarded open source projects enhances a developer's visibility and credibility within the tech community.

How do we fund Open Source Development?

- **Industry Partners:** Open-source projects partner with companies who use our software providing joint development efforts.
- **Grants:** Receiving funding from non-profit foundations and government.
- **Services**: Companies that provide services to help sites OMOPify their data and/or run network studies.
- **Open Core:** Core software under an open-source license with commercial licenses for additional proprietary features.
- **Consortia:** Multiple sites invest to add core beneficial features.

Open Source activity in OHDSI Teams

Martijn Schuemie

Anthony Sena
Chris Knoll

Frank DeFalco

HADES

Atlas/WebAPI

Technical Advisory Board

Package	Version	Maintainer(s)	Availability	Open issues	Open pull-requests	Build status	Coverage	PaRe
Achilles	v1.7.2	Frank DeFalco	CRAN	42	10	R check passing	codecov 2%	Report
Andromeda	v0.6.6	Martijn Schuemie	CRAN	13	0	R check passing	codecov 89%	Report
BigKnn	v1.0.2	Martijn Schuemie	GitHub	1	0	R check passing	codecov 96%	Report
BrokenAdaptiveRidge	v1.0.0	Marc Suchard	CRAN	2	0	R check passing	codecov 96%	Report
Capr	v2.0.7	Martin Lavallee	GitHub	6	0	R check passing	codecov 80%	Report
Characterization	v0.2.0	Jenna Reps	GitHub	6	1	R check passing	codecov unknown	Report
CirceR	v1.3.3	Chris Knoll	GitHub	3	1	R check passing	codecov 89%	Report
CohortDiagnostics	v3.2.5	Jamie Gilbert	GitHub	63	6	R check passing	codecov 90%	Report
CohortExplorer	v0.1.0	Gowtham Rao	CRAN	0	0	R check passing	codecov 100%	Report
CohortGenerator	v0.8.1	Anthony Sena	GitHub	26	3	R check passing	codecov 98%	Report
CohortMethod	v5.2.1	Martijn Schuemie	GitHub	15	0	R check passing	codecov 89%	Report
Cyclops	v3.4.0	Marc Suchard	CRAN	18	0	R check passing	codecov 84%	Report
DatabaseConnector	v6.3.2	Martijn Schuemie	CRAN	21	1	R check passing	codecov 60%	Report
DataQualityDashboard	v2.6.0	Katy Sadowksi	GitHub	44	8	R check passing	codecov 85%	Report
DeepPatientLevelPrediction	v2.0.3	Egill Fridgeirsson	GitHub	22	2	R check passing	codecov 99%	Report
EmpiricalCalibration	v3.1.2	Martijn Schuemie	CRAN	1	0	R check passing	codecov 84%	Report
EnsemblePatientLevelPrediction	v1.0.2	Jenna Reps	GitHub	5	0	R check passing	codecov unknown	Report
Eunomia	v2.0.0	Frank DeFalco	GitHub	10	0	R check passing	codecov 72%	Report
EvidenceSynthesis	v0.5.0	Martijn Schuemie	CRAN	3	0	R check passing	codecov 78%	Report

TAB Charter

The OHDSI Technical Advisory Board Charter (TAB) is a document that serves as the guiding principles of the TAB. It outlines the TAB's mission, responsibilities, deliverables, and more, providing guidance for the board's operations and a clear framework for its decision-making processes.

Contents

https://ohdsi.github.io/Hades/packageStatuses.html

https://github.com/OHDSI/TAB/wiki

TuftsMedicine

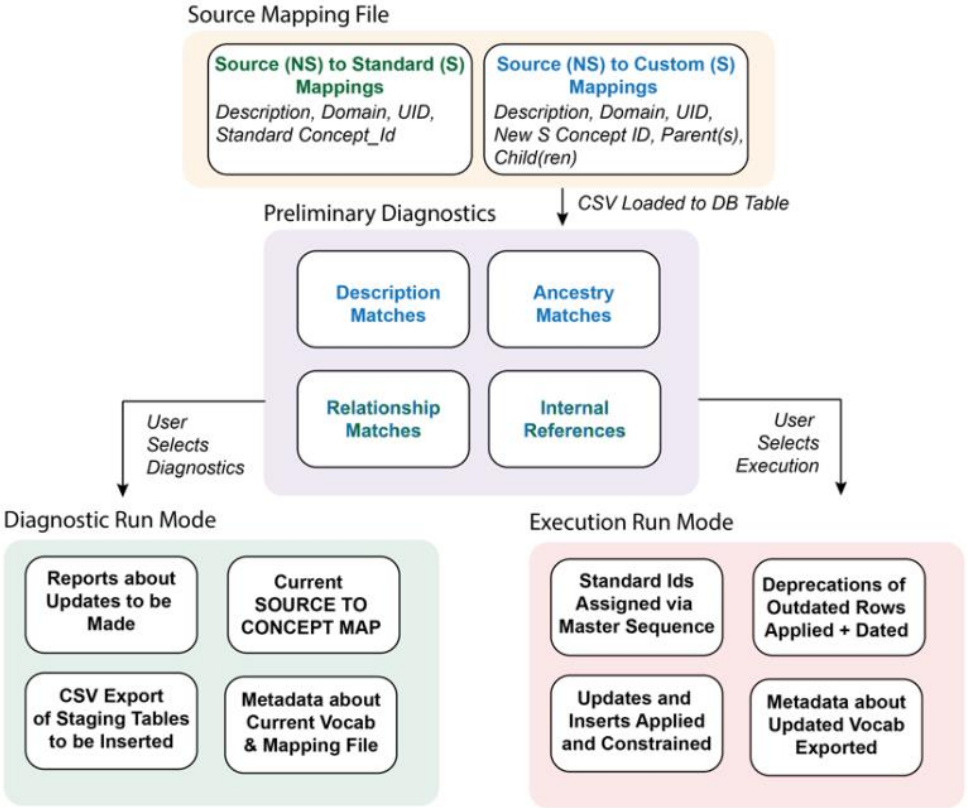
Custom Vocabulary Builder (CVB): A Multi-Platform Workflow for Terminology Management and Validation

An overview of the pipelines, processes, and tools to create custom ontologies

OHDSI Community Call
Jared Houghtaling
Asst. Prof. – Tufts Medicine
Institute for Clinical Research and Health Policy Studies (ICRHPS), Tufts Clinical and Translational Science Institute (CTSI)

29 April 2025

Background

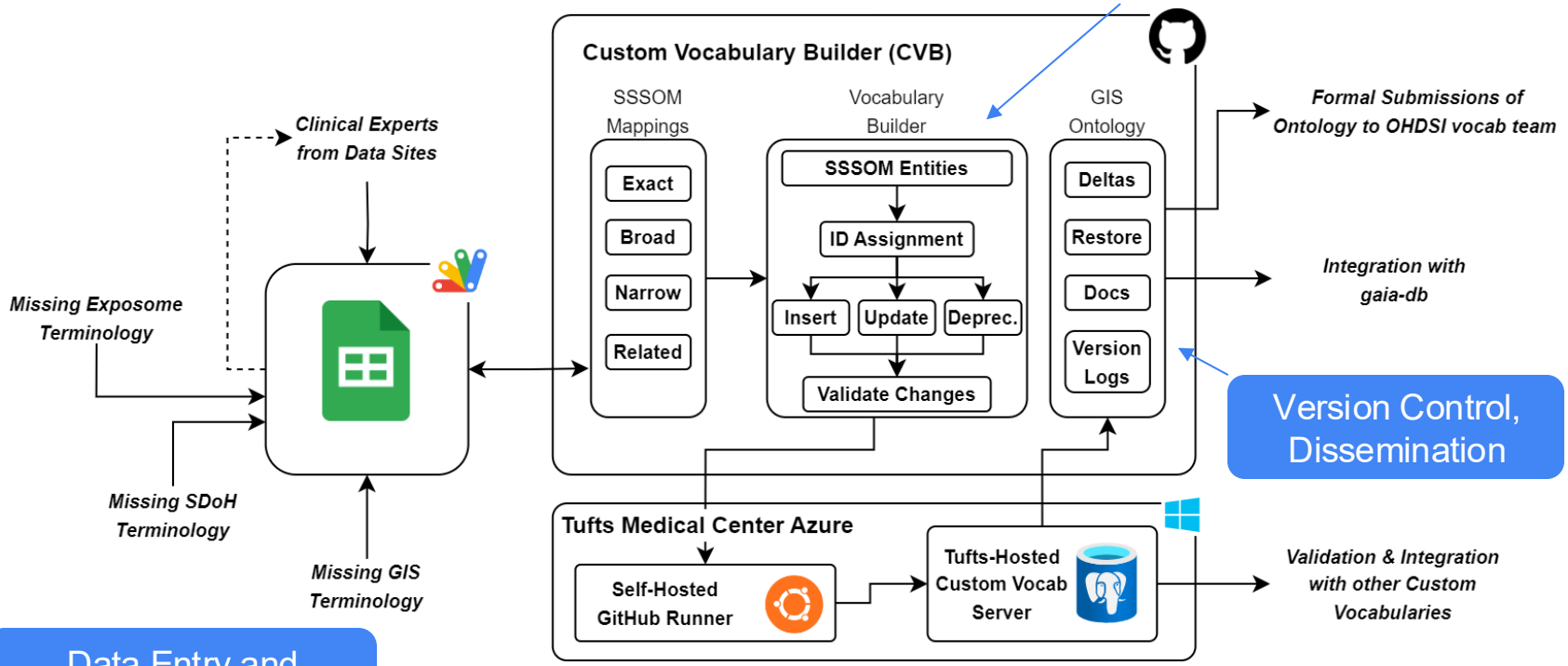


LIMITATIONS:

– Bespoke implementation with site-specific source mapping structure
– No continuous integration
– No mechanisms for delta vocab dissemination
– Poor version control

Overview – CVB

ID Assignment & Relational Logic



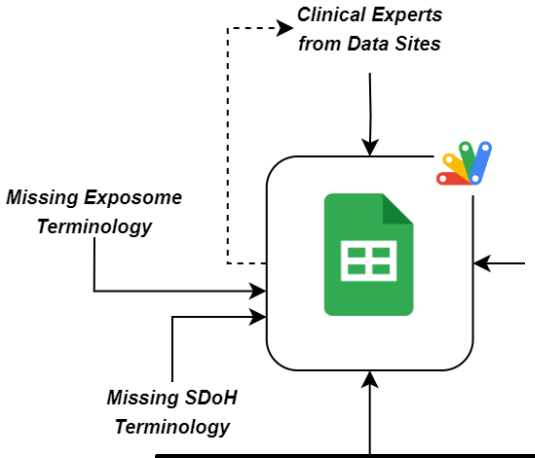
Version Control, Dissemination

Data Entry and Mapping Validation

Continuous Integration

https://github.com/TuftsCTSI/CVB

Expert Data Entry



GIS Vocabulary Package - Version 2.0EXT

File Edit View Insert Format Data Tools Extensions Help

A1 | source_code

	A	B	C	D	E	F
1	source_code	source_concept_	source_vocabula	source_description	relationship_id	predicate_id
2	EPAA_2NDMAX	0	OMOP SDOH	2nd Highest 1-Hour Measurement Of Carbon Monoxide (CO) In The Year (PPM)	Has relat contex	skos:relatedMat
3	EPAA_2NDMAX	0	OMOP SDOH	2nd Highest 24-Hour Average Of PM10 In The Year	Has relat contex	skos:relatedMat
4	EPAA_2NDMAX	0	OMOP SDOH	2nd Highest 24-Hour Average Of Sulfur Dioxide (So2) In The Year	Has relat contex	skos:relatedMat
5	EPAA_2NDMAX	0	OMOP SDOH	2nd Highest Daily Max 1-Hour Measurement Of Ozone (O3) In The Year (PPM)	Has relat contex	skos:relatedMat
6	EPAA_2NDMAX	0	OMOP SDOH	2nd Highest Non-Overlapping 8-Hour Average Of Carbon Monoxide (CO) In The Year (PPM)	Has relat contex	skos:relatedMat
7	EPAA_4THMAX	0	OMOP SDOH	4th Highest Daily Max 8-Hour Average Of Ozone (O3) In The Year (PPM)	Has relat contex	skos:relatedMat
8	EPAA_98PR_PM	0	OMOP SDOH	98th Percentile Of The Daily Average Measurements In The Year Of PM2.5 (Ug/M3)	Has relat contex	skos:relatedMat
9	EPAA_98PR_NO	0	OMOP SDOH	98th Percentile Of The Daily Max 1-Hour Measurements Of Nitrogen Dioxide (No2) In The Year (Pp	Has relat contex	skos:relatedMat
10	EPAA_99PR_SO	0	OMOP SDOH	99th Percentile Of The Daily Max 1-Hour Measurements Of Sulfur Dioxide (So2) In The Year	Has relat contex	skos:relatedMat
11	GIS1000000	0	OMOP SDOH	Access To Clean Water	Subsumes	skos:narrowMat
12	GIS1000000	0	OMOP SDOH	Access To Clean Water	Subsumes	skos:narrowMat
13	GIS1000000	0	OMOP SDOH	Access To Clean Water	Subsumes	skos:narrowMat
14	GIS1000000	0	OMOP SDOH	Access To Clean Water	Subsumes	skos:narrowMat

Apps Script GIS Vocab Integration

Files | SendToGithub.gs | Libraries | Services

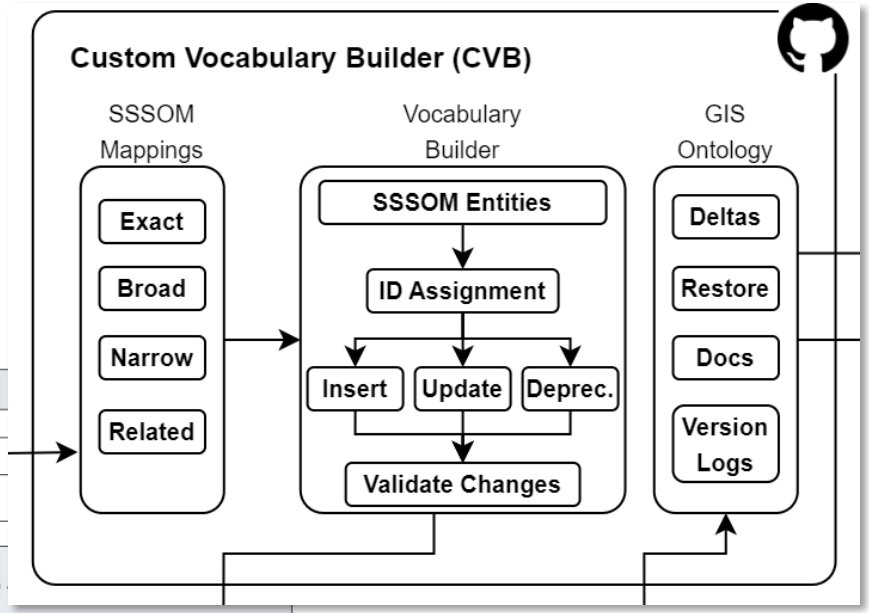
Run Debug createJwt Execution log

```
const rsa_load = ScriptProperties.getProperty("rsakey");

const createJwt = ({ privateKey, expiresInMinutes, data = {} }) => {
  // Sign token using RSA with SHA-256 algorithm
  const header = {
    alg: 'RS256',
    typ: 'JWT',
  };

  const now = Date.now();
  const then = new Date(now);
```

Use AppsScript to integrate GoogleSheet with GitHub

Build Pipeline

SSSOM-CSV Files from GoogleSheets



CVB Public

main | 4 Branches | 0 Tags | Go to file

omop-vocab-builder[bot] Mapping Update: vocabulary_delta - 20250... f7d78f9

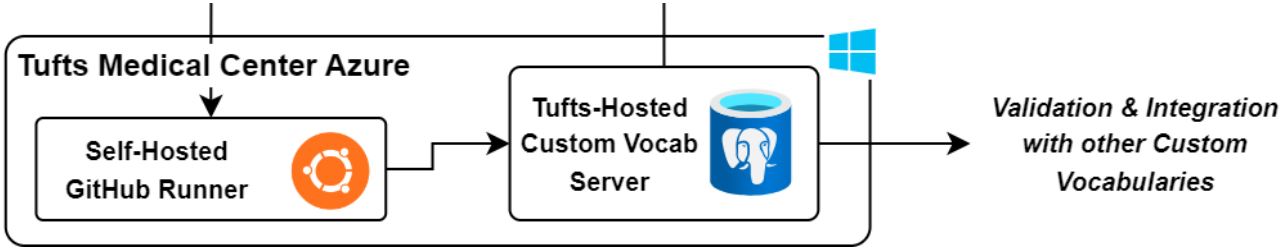
.github/workflows	Launch first run	3 weeks ago
GIS	Mapping Update: vocabulary_delta - 20250418	last week
MIMIC	Mapping Update: vocabulary_delta - 20250321	last month
PSYCHIATRY	Mapping Update: vocabulary_delta - 20250401	3 weeks ago

GitHub Action launches build logic on self-hosted runner on Tufts Azure

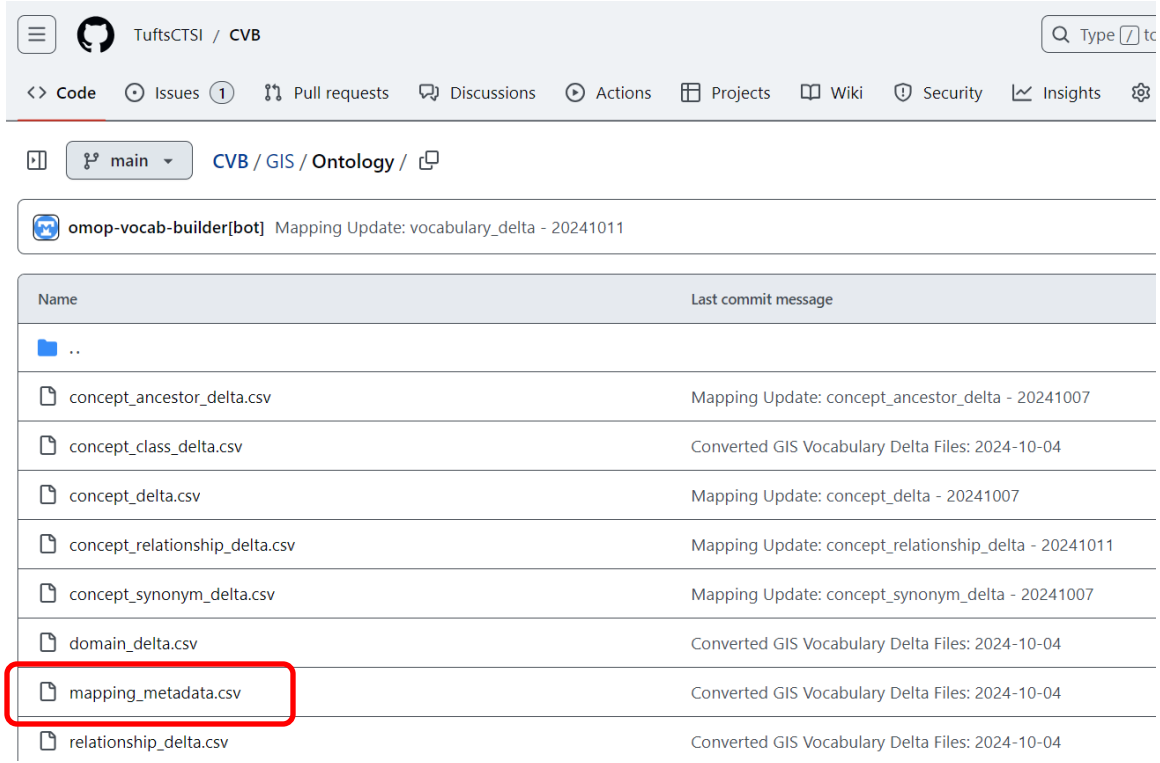
Build Execution

```
Connecting...
2024-10-21T13:49:40.65445  Connecting to the container 'gh-runner'...
2024-10-21T13:49:40.69721  Successfully Connected to container: 'gh-runner'
2024-10-17T02:17:36.719439188Z √ Connected to GitHub
2024-10-17T02:17:37.222142073Z
2024-10-17T02:17:37.222160319Z # Runner Registration
2024-10-17T02:17:37.222163286Z
2024-10-17T02:17:37.333014991Z
2024-10-17T02:17:37.333194076Z
2024-10-17T02:17:37.333891407Z
2024-10-17T02:17:37.570368242Z √ Runner successfully added
2024-10-17T02:17:38.353188812Z √ Runner connection is good
2024-10-17T02:17:38.353201812Z
2024-10-17T02:17:38.353205144Z # Runner settings
2024-10-17T02:17:38.353207622Z
2024-10-17T02:17:38.356242475Z
2024-10-17T02:17:38.356253926Z √ Settings Saved.
2024-10-17T02:17:38.356256428Z
```

Self-Hosted Runner in Azure Container App launches build and validation logic against private Postgres server with OMOP vocabulary tables



Versioned Delta Output



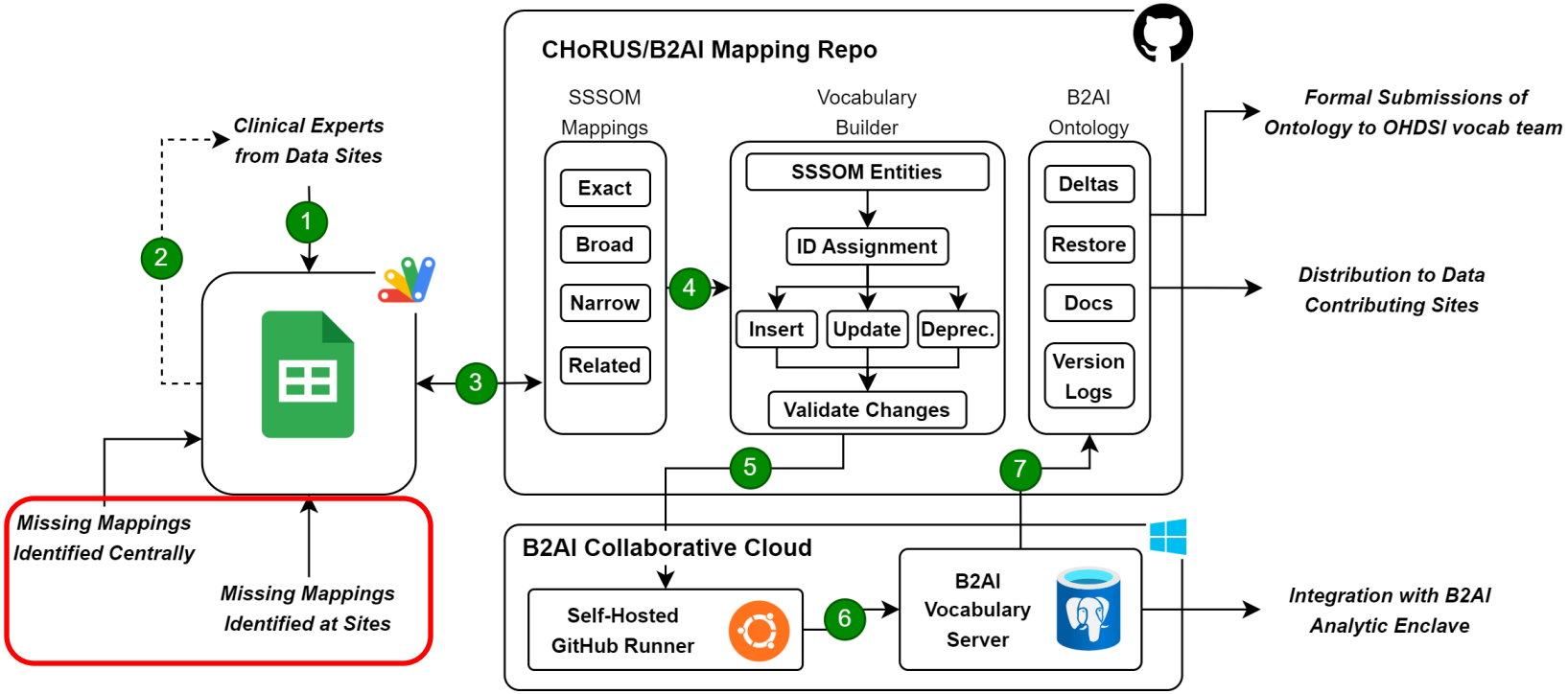
User can download delta terminology and insert directly into existing vocabulary tables

Current Use Cases

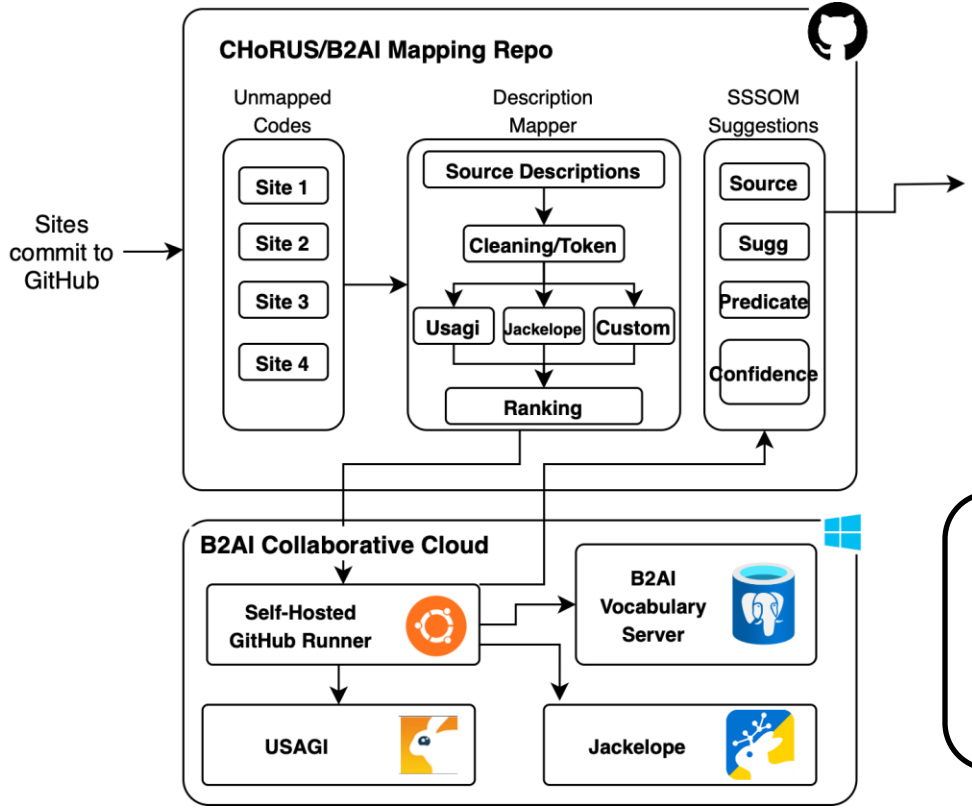
- GIS
- Psychiatry
- MIMIC IV
- CHoRUS Bridge2AI (diff. repo)
- Possibly yours?

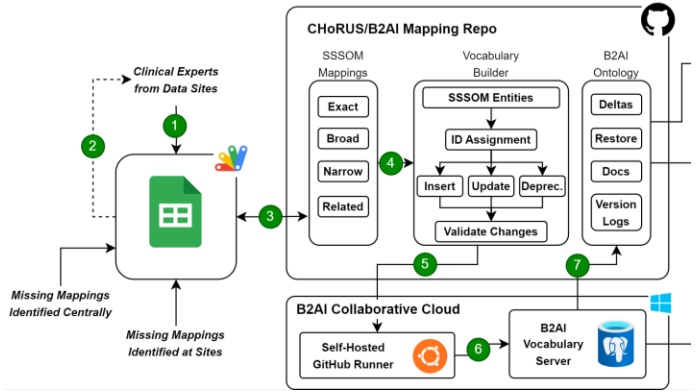
Check out Brian Gow's PR updating the MIMIC IV -> OMOP ETL with new vocab!

Closing the Loop – Mapping Pipeline



Closing the Loop – Mapping Pipeline





Mapping workflow populates a SSSOM formatted GoogleSheet that can be distributed for validation

Other Next Steps

- Tighter integration with the OHDSI vocab team

- Templating and parameterizing for quicker expansion to other use cases

- Evaluation of scalability, ease of use, and quality of resulting delta vocabularies

- Validation and expansion of underlying builder logic to cover additional edge/use cases

Acknowledgments

- Polina Talapova
- Kyle Zollo-Venecek & GIS WG
- Andrew Williams
- Robert Miller
- Brian Gow & Physionet Team
- Piper Ranallo & Psychiatry WG
- Freija Descamps, Lars Halvorsen & edenceHealth Team
- OHDSI members and developers
- CHoRUS Research Consortium

Useful Links

- github.com/TuftsCTSI/CVB
- github.com/chorus-ai

Developers Dialogue Panel Recap

- Invest in DevOps
  - Version control & standard repo structure
  - CI/CD & containerization
- Copilots in, vibe coding out
  - LLMs excel at documentation & writing repetitive/boilerplate code
- Open source is a virtuous cycle
  - Look for win-win scenarios for devs, their employers, and the OHDSI community

Panelists

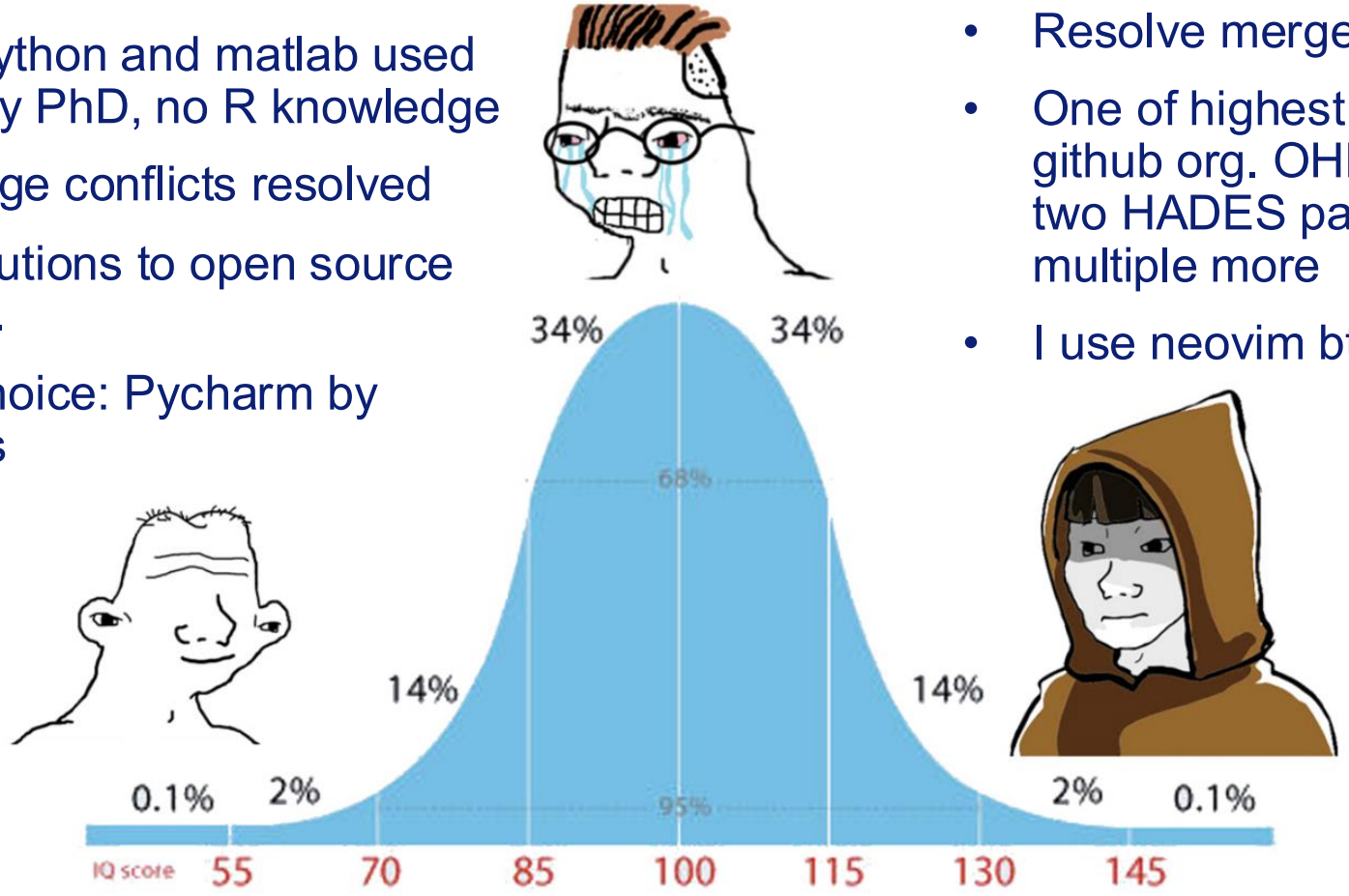
- **Martin Lavallee**, Boehringer Ingelheim
- **Eduard Korchmar**, EPAM
- **Egill Fridgeirsson**, Erasmus MC
- **Lawrence Adams**, Artificial Intelligence Centre for Value Based Healthcare

Developers Dialogue Introduction

Egill Fridgeirsson
Postdoc Medical Informatics @ Erasmus MC Rotterdam
Patient-Level Prediction maintainer
Deep Patient-Level Prediction maintainer and creator
e.fridgeirsson@erasmusmc.nl

Me before OHDSI

- Mostly python and matlab used during my PhD, no R knowledge
- 0 git merge conflicts resolved
- 0 contributions to open source software.
- IDE of choice: Pycharm by Jetbrains



- Quite proficient in R and Python. Can hold my own (debug and hack) in c++, java, rust
- Resolve merge conflicts every week
- One of highest committer in OHDSI github org. OHDSI titan, Maintainer of two HADES packages, contributed to multiple more
- I use neovim btw (and sometimes rstudio)

Erasmus MC

How did I get there?

- Lucky to be in a position where I use and rely on OHDSI software packages for my research (deep learning models for patient-level prediction)
- Quickly after being hired I was introduced to Jenna Reps, an important collaborator of our department and main maintainer of PLP at the time.
- I learned basic R and how to use the debugger (debugonce and browser)
- I'm insanely curious about the code I'm using, to trust the code for my studies I needed to understand everything. Often meaning I stepped through the code line by line.
- Jenna is nice, so I immediately started pestering her with issues on github. Soon after I started contributing to PLP

My PLP dev highlights

- Reduce flakiness of tests and speed
  - Speed went from 30 minutes to 1hour => 90 seconds today
- Improve CI/CD
  - Test all hyperlinks in docs
  - Publish docker container automatically with every new release
- Get into CRAN
- Other small things
  - Enforce HADES codestyle strictly
  - Add docs to many internal functions
- Content wise: new models, new methods, bugfixes, optimizations and a lot more!
- Unique challenges
  - Multilingual codebase – mixing python and R.

DEMO

Develop a deep learning model using a gpu with the deep PLP docker container