

# Scaling OHDSI open source community projects, lessons learned by Oncology WORKGROUP

Shilpa Ratwani<sup>1</sup>, Asieh Golozar<sup>1</sup>, Michael Gurley<sup>1</sup>, Andrew Williams<sup>1</sup>, Robert Miller<sup>1</sup>, Christian Reich<sup>1</sup>, Dmytry Dymshyts<sup>1</sup>, Michael Kallfelz<sup>1</sup>, Rimma Belenkaya<sup>1</sup>

<sup>1</sup>OHDSI Oncology WORKGROUP

## Background

The value of OHDSI as an open source research community has a direct correlation to the size of the community and its ability to influence observational research by improving both the quality of healthcare data analytics as well as the usefulness of healthcare data to these stakeholders.

The OHDSI community's success lies in strong thought leadership that is continually expanding, valuing and respecting the input of the community for innovation, adoption, market acceptance and providing structured ways for community member to participate and contribute. The inclusive membership model of the OHDSI community enabled OHDSI to produce open software, standards, data models, and analytics in the domains of participating researchers, programmers and engineers, system architects, and other technical contributors.

However, it takes immense commitment to build and coordinate this unique ecosystem to make it aligned with community research needs. Extending OMOP CDM, Vocabulary and analytic toolset to support observational cancer studies represents a large community effort that requires active leadership, product and project management alongside with technical and research contributions to be successful. In this report, we are sharing lessons learned from this ongoing effort.

## Methods

- (1) From the beginning, the Oncology Workgroup has had a strong **leadership** consisting of determined, likeminded, technically strong collaborators with a clear vision of the group mission and goals. For the first two years the same technical and strategic leaders performed project management that proved to be ineffective in addressing the scope of the effort.
- (2) Four **subgroups** specialize in primary development areas:
  - a. The CDM/Vocabulary Subgroup focuses on conceptual modeling of cancer data in OMOP CDM, extension of OMOP CDM and development of specialized cancer vocabularies, and producing guidelines for the use of the Oncology Module
  - b. The Development Subgroup focuses on the development of ETL and post-ETL .
  - c. The Genomic Subgroup focuses on the harmonization and representation of the Genomic Somatic variants in the OMOP Standardized Vocabularies.
  - d. The Outreach/Research Subgroup focuses on collaboration with other informatics groups, dissemination and adoption of the Oncology Module, and research opportunities.
- (3) A designated **project manager** joined the Workgroup in 2018 focusing on:
  - Tracking a project's mission, goals, and objectives, setting concrete goals enables the

- participants in a project to discuss and align around a single narrative that can guide their activities and inform decisions on things that work and addressing things that don't.
- Facilitating/coordinating the WORKGROUP activities to ensure decisions are documented, assigned, and completed. With individuals volunteering their time to attend WORKGROUP discussion, it's important to make sure the time contributed is used effectively to progress on aspects of the project that bring the most value to stakeholders.
- (4) **Documentation** is an area that can easily be sidelined in a large community setting with focus on developing open source products/models. The Oncology WORKGROUP follows a structured approach to technical and user documentation available and makes it available to the community via Github.
- (5) **Dissemination** efforts of the group include tutorials, workshops, conference presentations, publications, proactive outreach to standard and research organization, support for adapters of the OMOP Oncology Module.
- (6) **Self-governance and sustainability.** Oncology WORKGROUP like many other initiatives in the OHDSI community attracts individuals that are passionate about contributing to research. Our approach has been channeling this passion by enabling individuals self-govern and undertake individual areas of development.

## Results

Utilizing the methods described above, the WORKGROUP was able to implement extension of the OMOP CDM and Vocabulary, develop ETL and post-ETL processes and enable bringing cancer data into OMOP CDM and enable cancer research at OHDSI. Specific outcomes include:

### CDM and Vocabulary

- Integration of ICDO-3, NAACCR, CAP, HemOnc, NCIt
- Creation of the Cancer Modifier vocabulary
- Integration and harmonization of 6 genomic variant databases
- Extension of CDM with the Episode model
- Conventions for defining cancer disease and treatment episodes

### ETL and post-ETL

- Vocabulary driven ETL from tumor registry
- Post-ETL treatment regimen extractor

### OMOP Oncology Module adoption

MSK, IQVIA, Columbia, Tufts, Northwestern, University of South Whales, University of South Korea

### Education and Dissemination

- OHDSI Symposium Tutorial
- European OHDSI Oncology Workshop
- Papers in major oncology informatics journals<sup>1,2,3</sup>

### Network Studies

- Treatment pattern and outcomes of patients with metastatic bladder cancer using the Oncology Module and Vocabulary
- Long-term Outcomes of Prostate Cancer Patients Undergoing Non-Interventional Management (i.e., Watchful Waiting) and the Impact of Comorbidities and Life Expectancy

- Characterizing patients with metastatic Non-Small Cell Lung Cancer (NSCLC) with and without liver metastasis at the time of diagnosis with metastatic NSCLC

## Conclusion

Product and project management positively affected productivity and efficiency of the Oncology Workgroup efforts demonstrating that even relatively simple changes in the operational model can have a significant impact on outcomes.

Coordination of activities allowed the WORKGROUP to implement 6 new vocabularies during the last two years as opposed to 3 vocabularies during the first two years. Project management and operationalization lessened the load on technical contributors and better leveraged their expertise. This produced impactful improvements to the earlier versions of the OMOP Oncology Module by better reflecting research needs, and being thoroughly tested, and transparent.

Breaking the Workgroup into several sub-specialties, ensuring strong leadership in each sub-specialty and close collaboration between the sub-specialties provided better focus on different areas of development and at the same time broader coverage.

Targeted dissemination efforts have garnered the interest of various research entities like NCI and NIH, FDA, JAX, N3C, VICC, Mitre, CISNET, European Innovative Medicines Initiative that are presently collaborating with the Oncology Workgroup on their development efforts and research.

To further stakeholders' interest in conducting cancer research by organizations old and new to the OMOP framework, a strategy of scaling Oncology WORKGROUP efforts is necessary. Present Workgroup resources do not fully address the need of further development and adoption of the OMOP Oncology Module and conducting network research. The strategy of scaling should include seeking grants and other funding opportunities to operationalize this program.

## References

1. Warner JL, Dymshyts D, Reich CG, Gurley MJ, Hochheiser H, Moldwin ZH, Belenkaya R, Williams AE, Yang PC. HemOnc: A new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. *J Biomed Inform.* 2019 Aug;96:103239.
2. Belenkaya r. et al. Extending the OMOP Common Data Model and Standardized Vocabularies to Support Observational Cancer Research. *JCO Clin Cancer Inform.* 2021 Jan;5:12-20.
3. Jeon H, You SC, Kang SY, Seo SI, Warner JL, Belenkaya R, Park RW. Characterizing the Anticancer Treatment Trajectory and Pattern in Patients Receiving Chemotherapy for Cancer Using Harmonized Observational Databases: Retrospective Study. *JMIR Med Inform.* 2021 Apr 6;9(4).