

OHDSI EDUCATION WORKING GROUP: DRAFT OHDSI EDUCATION STRATEGY

Federated learning for federated research

Introduction & Context

The OHDSI community has grown and expanded over the last decade to be a formidable force for open science research based on the use of harmonised data at scale and via federated/distributed networks globally of data custodians and researchers, with standardised analytical tools, skills, and methods. OHDSI incorporates multiple disciplines, from epidemiologists, clinicians, data scientists, statisticians, informaticians, from academia, research institutions, government agencies and industry. It is now in a phase of rapid growth and expansion.

OHDSI has evolved as an open science community, predicated on collaboration and in the main voluntary studies involving those with curated, harmonised data, and/or those with a research interest in analysing this data via federated analysis. As such, it has become one of the most dynamic collaborations, with representation and activity in most geographical regions and across diverse therapeutic areas, inclusive of global pandemic response to COVID-19. It has demonstrably contributed to scientific discovery and development of using observational data at scale across multiple therapeutic areas and in the evolution of methodological science.

An OHDSI framework has matured over the last decade, incorporating the OMOP common data model, standardised analytical pipelines, methods and tools, all used to analyse real world, observational data and to generate evidence. Most studies are in three categories:

- Characterisation
- Population effect estimation
- Patient-level prediction

To support studies there are essentially three approaches to implementing them and conducting analysis:

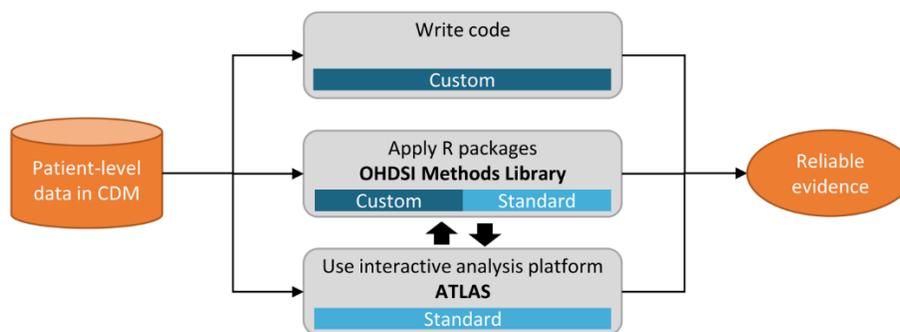


Figure 1: Implementation of studies on the OMOP CDM (Book of OHDSI, 2019)

Conducting research within the OHDSI framework requires an educational, skills and experiential background to be able to generate questions, design appropriate studies, implement methodological approaches as for figure (1), and evaluate analyses prior to dissemination of evidence generated dynamically via publications and online. As well as knowing the underpinning epidemiological theory,

researchers have a learning curve to be able to optimally utilise the increasing suite of tools used in the OHDSI framework, such as ATLAS or in the Health Analytics Data-to-Evidence Suite (HADES) in generating reliable evidence.

To ensure quality of the research method and generated evidence it is critical that researchers are well versed in epidemiological and allied principles, as well in the OHDSI framework, bearing in mind that the tools and methods continue to be updated, evolve, and are added to over time.

The OHDSI community could be segmented into the following approximate categories:

- Novice
- Intermediate
- Expert
- Mentor

There is a presumption that those who are joining the OHDSI community have diverse educational and skills background, all relevant to research, but they have a learning curve in understanding, adopting, and using the OHDSI framework. As community members progress from novice to intermediate to expert, to mentor, they will also increase their contribution to the further refinements and developments of the tools and methods used in OHDSI. This is through a continuum, with individual rates of progress within personal circumstances, support, and contribution, and the emphasis on self-directed adult learning. Each category could support the prior one, especially Novice and Intermediate).

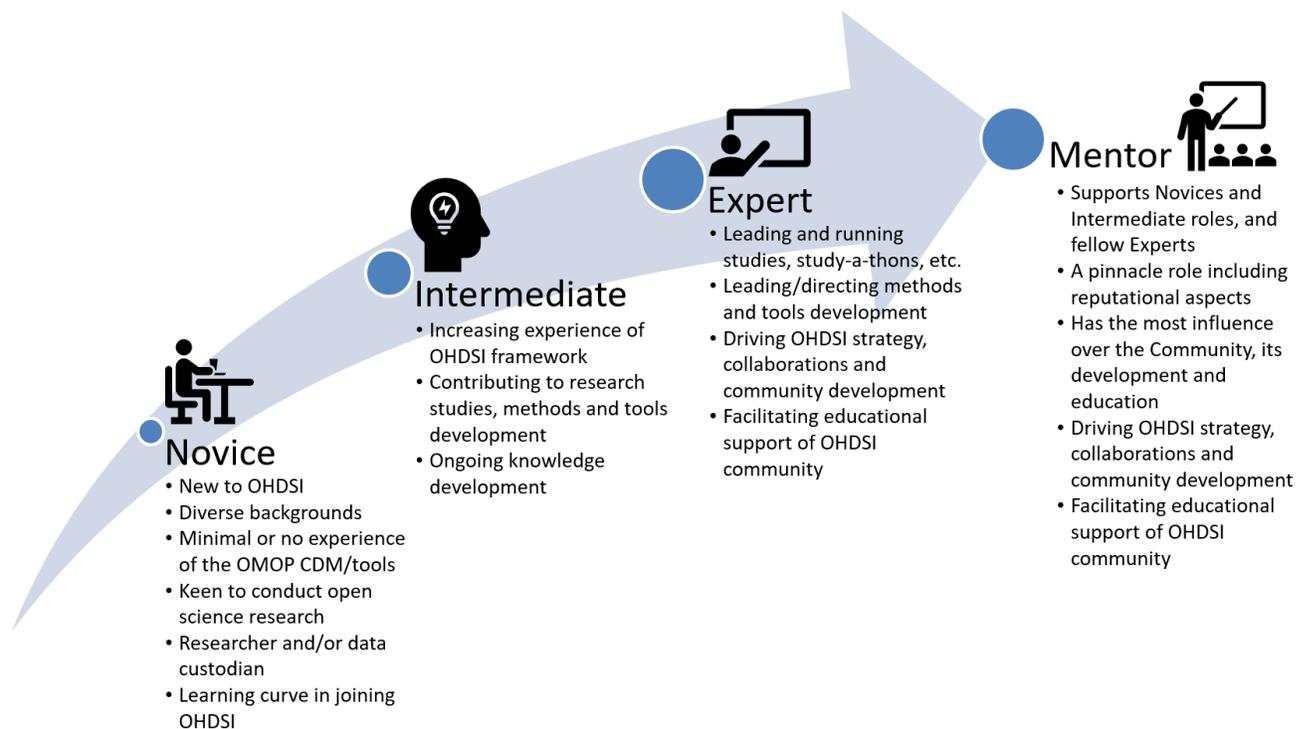


Figure 2: OHDSI Community Segmentation

Within this segmentation, various categories can be envisaged:

- Working directly on RWD, ETL
- Tool development
- Analytics, methods
- Using all the above for evidence generation
- Knowing all the prior four

Moving across from novice to mentor can be across these categories, but also there are widening knowledge across the OHDSI ecosystem. Stage gating between the segmentation could potentially be formalised, in terms of more formal recognition of moving to the next stage (e.g., novice to mentor, mentor to expert). Mapping of the community could be an initial focus.

Incorporating the segmentation will assist with the development of the education strategy for OHDSI. Currently, the provision of educational resources in OHDSI is fragmented, with the inclusion of the following resources:

- Regular internal symposia (regional)
- Video outputs from symposia and meetings
- The Book of OHDSI
- Publications, conference posters and abstracts
- Community meetings
- Community Forums
- Working Groups
- Repository development
- Hosted environments
 - OHDSI.org
 - EHDEN Academy
 - YouTube

A critical initiative was the Book of OHDSI, published in 2019, in [online](#) iterative form and in paperback via Amazon. Meanwhile all other resources are to a lesser degree accessible via OHDSI.org and related sites (as above), but there is not a comprehensive resource available to assist with triaging and signposting those in or outside of the OHDSI community to relevant educational resources.

Recently the EHDEN project conducted a survey of its 12 pharmaceutical company consortium partners (private communication), providing insights into the type of functional leads in research-related decision-making, research priorities, datasets being used and available, challenges and annoyances in conducting observational research. Most of the findings were commensurate with our broad understanding of today's challenges (and indeed the rationale for the development of the OHDSI framework, and initiatives such as EHDEN), in terms of the paucity of quality datasets representative and relevant for research needs.

Furthermore, many of the frustrations experienced by researchers in this domain were around the process of conducting research, mostly governance, versus methodological. This is exacerbated by the diversity of research conducted by researchers, and the paucity of standardised data, and methodological approaches.

A new OHDSI working-group in 2021, for early-stage researchers, also anecdotally suggested some entry challenges for novices to OHDSI:

- Understanding the language of OHDSI
- Identifying relevant resources/materials
- Further educational pathways

Essentially these are characteristic of any novice entering into a community or organisation, and a number of solutions can help address the entry challenges and learning curve for those joining and participating in the OHDSI community.

Meanwhile, the development of OHDSI centres, especially in North America, point to the emergence of focal academic institutions within the OHDSI network who will also support education, potentially in more formal settings and qualifications. Figure (3) is a schematic of the international community, resources, and pivotal nodes for OHDSI.



Figure 3: Network of OHDSI global community and resources

Proposed approach

To address the challenges outlined above in a strategic fashion, ensuring more formal structuring of educational resources and to support an expanding OHDSI community, the following recommendations should be considered:

- A guide to the OHDSI community – like a guidebook or app which provides insights and direction to participating and contributing, perhaps including do's and don'ts (which could be incorporated into the Book of OHDSI)
- The Book of OHDSI is currently in English, Korean, and Chinese, but consideration should be given to other international languages, especially Spanish, Hindi and Portuguese. Furthermore, it is an excellent medium for community members to contribute to, and a singular resource representing that community and the OHDSI research framework
- More educational resources need to be incorporated into OHDSI.org

- Elements of the Book of OHDSI (e.g., expanded glossary of terms)
- A catalogue of educational resources
 - Meetings (already included)
 - Videos and materials
 - Signposting to resources globally (e.g., institutions, courses, etc.), inclusive of within learning pathways (see later)
 - Training datasets (see later)
- The navigation of the site needs to be improved and enhanced to facilitate discovery by visitors, with a more modern interface, look and feel
- The catalogue of educational resources would support a federated approach vs. local, i.e., resources remain where they are generated and maintained:
 - Informal materials – videos, meetings, and symposia outputs, e.g., abstracts, posters, etc.
 - Formal materials – educational courses, accredited qualifications
 - Resources – Book of OHDSI, publications
 - Driven by a criterion of relevance to the OHDSI community/OHDSI framework
- The technical and scientific outputs of OHDSI need to incorporate linked education resources:
 - Each tool developed needs a course or vignettes to assist initial utilisation with an overview of key features
 - High impact publications should be accompanied by short overviews and/or videos outlining the importance, methods and implications
- Tool training resources
 - When OHDSI members are learning to use the OHDSI framework tools, training datasets, whether free to access and/or synthetic, need to be available to support hands-on use during the learning curve. Currently availability and access is fragmented and also require a degree of knowledge beyond the novice level. This could be incorporated also into the catalogue of resources
- Repository development
 - Work Group outputs (metadata driven) for a centralised resource for the wider community for multiple ends, including educational, and to link Novices and Intermediates to potential opportunities for collaboration and focused work within Work Groups
- Development of a mentorship scheme
 - Use of Mentors, and Expert (subject to resources/inclination) to support Novices and Intermediate OHDSI members is currently informal, but could be formalised longer term with appropriate support and resources. Some intermediate colleagues may also consider being mentors too. This could likely be streamlined by interest and application themes of groups
- Creation of clearly-outlined, competency-based learning pathways
 - Formal establishment of an educational curriculum for OHDSI members:
 - Introductory modules common to all (what should everyone know?)
 - OMOP CDM and standardised vocabularies
 - ETL learning pathway
 - Tools, e.g., ATLAS

- Methods
 - Phenotype definition, characterisation, and evaluation
 - Characterisation
 - Population-level effect estimation
 - Patient-level prediction
 - Federated networks, network studies at scale
 - Principles of observational research governance
- Differentiated modules per e.g., functional specialism (e.g., regulatory science, or value-based healthcare)
- Signposting within a learning pathway as to the location of modules, even internationally between different academic centres or providers
- Use of resources such as the EHDEN Academy and OHDSI Centres
- Use of the Education WG to assist in oversight of learning pathways
- Ultimately there could be accredited qualifications within the OHDSI framework, driven by the academic centres and partners in the community, alongside and/or integrated within existing accreditation pathways for relevant scientific disciplines
- Learning resource assessment and impact evaluation would be necessary and beneficial in ensuring an optimal experience for those studying, and for OHDSI in understanding what works across a community of practice
 - Alongside this will be the need for more rigorous assessment of learning resources and evaluation of educational outcomes. Exploration of more quantitative evaluation processes and tools would be important for future educational provision in OHDSI
 - Developing deeper insights are required into sustainability of OHDSI and the level of educated collaborators required for activities, research and Work Groups, inclusive of role definition and enhancing future capacity
- Funding
 - Almost the majority of activity in OHDSI is voluntary
 - To enhance the educational strategy, resources, and network and to ensure sustainability, consideration needs to be given to a central educational budget
 - Such a budget needs to be further elaborated, but even a nominal budget would already be impactful
 - Creating annual reports and metrics to measure the impact of our overall educational initiatives to substantiate future grant submissions
 - Use of OHDSI Centres to provide administrative support to organize and facilitate grant submissions or other sponsored curriculum opportunities for educational advancement (all efforts to be done in coordination with the OHDSI Steering Committee leadership)

NH251121