# Real4Reg Use Case 1: Lessons Learned on Exploring Real-World Data - Workshop Protocol

The Real4Reg consortium aims to ensure high visibility of the project and the promotion of active interaction with key stakeholders. As stated in the Dissemination, Exploitation and Communication Plan, Real4Reg *"will organise three workshops for real-world data (RWD) experts to address the needs from early carrier researchers to experienced professionals, and also the lay public, including patients. Registration will be free-of-charge to encourage the participation and training of young and junior team members in the partner organisations*."

The third workshop took place in January 2026 and the details are provided below.

**Workshop Title: "Real4Reg Use Case 1: Lessons Learned on Exploring Real-World Data"**

Date: 27 January 2026

Time: 11:30 - 13:00 CET

Online Platform: WebEx

**Agenda:**

| | |
|---|---|
| 11:30-11:35 | **Introduction** |
| 11:35-12:15 | **Presentation: Use Case 1 Results and Challenges in OMOP Implementation** – Vera Ehrenstein (Aarhus University) and Mohamed Aborageh (Fraunhofer) |
| 12:15-12:55 | **Roundtable session** – Discussion about Real4Reg's Use Case 1 results and how RWD can be helpful in regulatory decision-making and health technology assessment (HTA)<br>**Moderation**: Britta Haenisch (BfArM)<br>**Participants**:<br>Piia Rannanheimo (Finnish Medicines Agency, Fimea)<br>Lucía Ortíz (Plataforma de Organizaciones de Pacientes, POP)<br>Daniel Morales (European Medicines Agency, EMA)<br>Ruben van Eijk (University Medical Center Utrecht) |
| 12:55-13:00 | **Closing** |

The workshop was publicised through email invitations sent to a list of stakeholders, as well as via publications on Real4Reg's website, social media, and promotion by Real4Reg's partners on their websites and social media channels.

**Participants:**

The workshop had a total of 218 participants from 17 countries (Belgium, Cape Verde, Denmark, Finland, France, Germany, Hungary, Moldova, Netherlands, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States of America, and Wales). Participants came from different types of organisations: regulatory/HTA agencies, universities, industry, payers, healthcare institutions, patient associations, among others. Participants held a wide range of professional roles including scientists/researchers, market access specialists, project managers, pharmacovigilance and medical affairs specialists, professors, epidemiologists, pharmacists, biostatisticians, patient advocates or physicians.

The audience also included members of the Real4Reg consortium and Advisory Board, members of other sister projects of the MetReal Cluster, and related institutions.

**Workshop Recording and Materials:**

The recording of the workshop is available on Real4Reg's website, as well as the speakers' presentations. (Click here)

**Summary of The Activities:**

**Presentation - Real4Reg Use Case 1 Results and Challenges in OMOP Implementation**

Speakers: Vera Ehrenstein (VE) (Aarhus University) and Mohamed Aborageh (MA) (Fraunhofer SCAI)

Outline:

- Introduction to Real4Reg
- Use Case 1 – Objectives, Methods, Breast Cancer and Amyotrophic Lateral Sclerosis (ALS) Results (incidence, demographic characteristics, comorbidities, treatments, disease course, and overall survival)
- Observational Medical Outcomes Partnership (OMOP) – Data Access and Extraction, Mapping, Challenges, and Mitigation Measures
- Main Findings and Lessons Learned on RWD Use

**Roundtable:**

The roundtable session was moderated by Real4Reg's principal investigator Britta Haenisch (BfArM), with the following participants: Piia Rannanheimo (PR)(Fimea), Lucía Ortíz (LO)(POP), Daniel Morales (DM)(EMA), and Ruben van Eijk (RE)(University Medical Center Utrecht).

*Role of Each Stakeholder Group in Overcoming the Barriers to RWE Use in Pre-Authorisation*

PR stated that she had experience with one prior Finnish OMOP-based project and the issues she encountered were the same as in Real4Reg. She believes that it is important for HTA agencies to thoroughly describe the evidence needs to the data management, OMOP centres, and analyses institutions, so the systems can develop and adapt to answer HTA agency's needs. She also said that Real4Reg's results on breast cancer and ALS are highly detailed, which is very important to HTA agencies when they have to make decisions on new medicines, since they need to look at these specific indications that the products are intended for.

LO said that one of the patients' organisations main roles is to increase trust, usability, and speed of RWE use, which can be done by improving data sources awareness since these institutions act as connecters between patients and regulatory bodies. They also help in the translation of technical documents into plain language so patients can understand every step and help them in defining endpoints/outcomes that are meaningful in their real life to help regulatory decision-making.

DM believes in a greater data sources awareness, on which the HMA-EMA Catalogues of Real-World Data Sources and Studies can contribute. The European Health Data Space (EHDS) will hopefully be helpful in providing a larger data catalogue and in facilitating data access with standardised applications and stricter timelines. Regarding data quality DM said that the EMA Data Quality Framework might be important to improve data metrics, transparency, and reporting. In addition, European Medicines Regulatory Network (EMRN) can also help by providing scientific advice when there are uncertainties about RWD use/relevance for a specific question.

RE stated that accessing patient-level data is very difficult and, therefore, this should be the highest priority. He mentioned Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) as an example of a community dataset where academic institutions, clinical trials, electronic healthcare registries, and industry can contribute, but also access the data. He believes that collecting aspects that usually are not captured in electronic health records, for example patients' quality of life, is as important as collecting diagnosis' details. One solution might be the creation of a European common registry where each country collects and tries to harmonise core outcome sets.

### *OMOP Advantages Over Local Data Models*

PR said that one of OMOP's biggest advantages was the interoperability across borders, allowing the combination of multiple datasets and providing highly detailed and quality information. Another benefit was that data stays local and the analysis can be run separately, which translates into less time required and increased trust in this system.

LO believes that OMOP solves some problems that local models have, for example scalability and comparability. She mentioned Crobi, which is the first validated well-being scale that evaluates the psychological, affective, social, and emotional aspects of people living with chronic diseases. By analysing the results with an OMOP model, the researchers can provide patients' input on invisible aspects to regulatory decision-making.

DM mentioned that OMOP benefits are essentially interoperability, scalability, and standardisation. These can accelerate federated analysis, facilitate cross borders cooperation, and reduce heterogeneity, respectively.

For RE one important aspect of OMOP is the capacity of international exchange between different datasets. However, he mentioned that one of the main risks of using it is that it can be very time consuming to set it up and, therefore, some aspects might need to be prioritised over others.

### *Implausible Findings of Use Case 1 Results*

PR reflected about how comprehensive were the Finnish datasets, since they possess three OMOP data centres, however to be able to reflect on implausible findings she would need more background information regarding the methods used. In a decision-making point of view and keeping in mind the current needs, the results might be too broad level.

LO mentioned that POP represents all patients with chronic conditions, meaning that they do not focus on specific diseases and, therefore, she had no comments regarding ALS and breast cancer results specifically. However, in her opinion implausible findings might arise from data mapping issues and early spotting can avoid wrong conclusions that can interfere with regulatory or HTA decisions.

DM believed that the implausible findings were highlighted during the presentation and appropriately recognised that were likely caused by problems in the data capturing step. DM said that the absence of psychiatric disorders in the Portuguese breast cancer analysis is highly unlikely. In his opinion, it is really important to understand the data ahead of time, characterising it and generating data metrics, all tasks that will depend on data quality. However, it is also as crucial to understand the clinical practice guidelines used in each country, that might explain, to an extent, some of the differences found.

RE said that one of the implausible findings could be the discrepancies in riluzole use across different countries, especially because the rates seem very low compared to the literature, however they could also reflect practice differences across Europe. RE believes it would be important for ALS patients to have something similar to breast cancer analysis, regarding how severely affected patients were and for how long were they sick, in order to facilitate data interpretation and ensure the date makes sense.

### *Real4Reg's Impact in Stakeholders Work*

PR recognised that RWD is important on all stages of the medicine's lifecycle and that HTA bodies really need to describe their evidence needs, event though they vary slightly across different steering groups. She also believes that it is relevant to specify the questions that the projects are aiming to serve, so they can be customised to HTA needs. In her opinion, separating OMOP data basis and the processes for data acquiring should be considered, given the development that these areas still need.

LO mentioned that Real4Reg is very useful for their work as a patient organisation, since it can provide stronger evidence for patient advocacy. In addition, the project has a really practical impact due to its' capacity to make RWD more reliable and explainable, allowing it to be used in decisions that are actually relevant for patients and that have a direct impact on their quality of life.

DM believes that apart from the clinical results and evidence generation, Real4Reg can he very helpful to transparently explain the processes to get there, which really impacts stakeholders' trust. It can also be important to establish data metrics that can be run against different databases to determine fitness for purpose and to identify data gaps overtime that can be enriched. The transparency in OMOP mapping, as well as the sharing of recommendations and lessons learned can help in the acceptance and adoption of this model in certain areas.

RE saw this project as a benchmarking in datasets. From an epidemiological perspective, he mentioned Real4Reg's importance in helping diseases' understanding, for example by determining prevalence and differences between countries. From the methodologic point of view, RE believes that Real4Reg was also extremely valuable to assess the complications and to determine which joint efforts need to be done to capture missing aspects and to progress in the future.

**Questions From The Chat:**

- **Question for MA**: "*Why does OMOP not base its drug standardization on the ISO IDMP standards in Europe, but instead relies on RxNorm? Can the EMA's SPOR data model (the EU implementation of IDMP) be leveraged within OMOP mappings or vocabulary standardization?*"

  MA Answer: "OMOP started as a US-centric initiative, at a time when RxNorm was already the de-facto clinical drug terminology in the US and had robust mappings from the NDC and other code systems. The current OHDSI documentation states that many concepts are sourced from RxNorm, and drugs not from the US market are added via RxNorm Extension"

- **Question for MA**: "*As more health units harmonize to the OMOP CDM, I'm curious about the community's lessons learned regarding data sharing. In your experience, is it more effective to share a customized target dataset (fully mapped to OMOP with site-specific extensions) or to share the raw source data alongside the ETL specifications? I'm specifically interested in how could we balance the ease of immediate analysis against the need for researchers to verify the original data lineage.*"

  Answer: "I believe sharing an OMOP formatted data avoids introducing country specific nuisances and really saves a lot of time for projects that would be required to implement the ETLs. Even if the ETL is designed and ready, implementation can take time depending on the available hardware resources."

- **Question for VE**: "*You alleged that the enterprise is particular relevant to 'rare subgroups' - could you identify those reliably in the data (beyond men)*?"

  Answer: "Depends on the validity of the identification. From validation studies, positive predictive values of many database phenotypes is high meaning that you can often reliably identify patients with the condition. Whether the sensitivity is high, is more difficult to say. As always, the answer is 'it depends'."

- **Question for VE**: "*Would you be able to identify the clinical subtypes of BC such as HR+/-; HER+, TNBC?*"

  Answer: "In some data sources, yes. Depending on the data provenance (cancer registry, pathology data etc) the completeness of identification may differ."

- **Question for MA**: "*Even tough you covered RxNorm issues, it's difficult to understands how much impact does the RXNorm mapping have on the quality of data. Seems like an extra step in Europe?*"

  Answer: "I think integrating the European standards as part of the OHDSI project is a good start. For example, EMA SPOR is already being used to leverage OMOP mappings but as an auxiliary source that feeds mapping to RxNorm/RxNorm Extension."

- **Question for MA**: "*If I understood correctly, there was some manual work necessary to transform the native data to the OMOP already in the focussed Use cases 1 and 2. Would* *that aspect not be missing from federated analyses where data sources are mapped independently? Would not more heterogeneity be expected simply by the way the data was OMOPed?"*

Answer: "An independent mapping procedure would force some native sources to compromise on the specificity on their coding systems by only considering a level of coding that matches other participant partners. I believe this should be decided on, and the specificity levels and expected granularity should be described as part of the ETL design."

**Key Takeaways:**

- Key words: data access, interoperability, scalability, harmonisation, heterogeneity, data sources, evidence needs, and datasets.
- The main barriers to RWE use in pre-authorisation include data access in a timely manner, lack of trust, and data sources' unawareness. Some strategies to overcome these issues are the construction of community datasets, the exact description of data needs by HTA agencies, the establishment of stricter timelines for data access, and the fostering of trust and understanding by patients.
- OMOP presents major advantages over local data models, namely interoperability, scalability, comparability, standardisation, reduced heterogeneity, and facilitated international collaboration. However, one strong disadvantage is that it can be very time-consuming.
- Real4Reg's results can be impactful from two perspectives. From the methodological point of view the project can make RWD use more reliable, explainable, and trustworthy. It is also helpful to determine fitness for purpose and to identify gaps in datasets that can be enriched. From an epidemiological perspective, these results are also important to understand diseases courses and to identify differences between countries.

## Appendix – Comments From the Chat

- VE - Good morning, would it be possible to test slide sharing
- BF - Bom dia
- VS - Bom dia
- AV - Bom dia
- Infarmed - Good morning! Feel free to use the chat to share your questions and comments! While we may not be able to address everything during the webinar, we will take notes for future opportunities.
- HR - Bom dia
- PSP - How are OHDSI, Concept iDs etc. related to ICD, SNOMED ...?
- PSP - sorry, you have showed it presently
- PSP – Aren't free text field an area for AI-use?
- SR - Last year at the OHDSI Europe symposium there were many posters of researchers using AI to extract OMOP concepts from unstructured free text fields
- SR - OHDSI is an international network of researches working with OMOP, they also make tooling such as Athena available. Concept ID is the unique identifier within the OMOP common data model. OMOP further standardizes existing standards, it brings together SNOMED, ICD, LOINC, ... into one standardized and central overview. Existing standardization often has a relatively niche application: SNOMED = clinical concepts; LOINC observations and measurements; ICD = classification/reporting ... They have grown broader over time but have their own original domain. For example existing SNOMED codes have a concept ID on top to make sure there is an unique identifier across all existing standards. Type 2 diabetes mellitus is SNOMED 44054006 but in 201826 in OMOP. In OMOP it is guaranteed unique while the SNOMED code may also exist in ICD.
- PR - Why does OMOP not base its drug standardization on the ISO IDMP standards in Europe, but instead relies on RxNorm? Can the EMA's SPOR data model (the EU implementation of IDMP) be leveraged within OMOP mappings or vocabulary standardization?
- FB - As more health units harmonize to the OMOP CDM, I'm curious about the community's lessons learned regarding data sharing. In your experience, is it more effective to share a customized target dataset (fully mapped to OMOP with site-specific extensions) or to share the raw source data alongside the ETL specifications? I'm specifically interested in how could we balance the ease of immediate analysis against the need for researchers to verify the original data lineage.
- PSP - You alleged that the enterprise is particular relevant to 'rare subgroups' - could you identify those reliably in the data (beyond men)?
- VE - Depends on the validity of the identification. From validation studies, positive predictive values of many database phenotypes is high meaning that you can often reliably identify patients with the condition. Whether the sensitivity is high, is more difficult to say. As always, the answer is 'it depends'.
- FNW - Would you be able to identify the clinical subtypes of BC such as HR+/-; HER+, TNBC?
- VE - In some data sources, yes. Depending on the data provenance (cancer registry, pathology data etc) the completeness of identification may differ.
- MA - OMOP started as a US-centric initiative, at a time when RxNorm was already the de-facto clinical drug terminology in the US and had robust mappings from the NDC and other code systems. The current OHDSI documentation states that many concepts are sourced from RxNorm, and drugs not from the US market are added via RxNorm Extension

- PR – Even though you covered RxNorm issues, it’s difficult to understands how much impact does the RXNorm mapping have on the quality of data. Seems like an extra step in Europe?
- MA - I think integrating the European standards as part of the OHDSI project is a good start. For example, EMA SPOR is already being used to leverage OMOP mappings but as an auxiliary source that feeds mapping to RxNorm/RxNorm Extension.
- FH - Real4Reg saw the implementation of EHDS (at least its start). Other projects report EHDS has started to facilitate greatly the access to data. Wasn’t it the case for the 2 Real4Reg case studies?
- CA - Regarding the prospective European database (which will collect BC diagnosis and lifestyle of the patients), will you also take in account comorbidities, such as like diabetes, hypertension and obesity that are related to BC incidence?
- CA - Considering that these diseases can result from the lifestyle of breast cancer patients.
- FNW - If I understood correctly, there was some manual work necessary to transform the native data to the OMOP already in the focussed Use cases 1 and 2. Would that aspect not be missing from federated analyses where data sources are mapped independently? Would not more heterogeneity be expected simply by the way the data was OMOPed?
- MA - I believe sharing an OMOP formatted data avoids introducing country specific nuisances and really saves a lot of time for projects that would be required to implement the ETLs. Even if the ETL is designed and ready, implementation can take time depending on the available hardware resources.
- MA - An independent mapping procedure would force some native sources to compromise on the specificity on their coding systems by only considering a level of coding that matches other participant partners. I believe this should be decided on, and the specificity levels and expected granularity should be described as part of the ETL design.
- Infarmed - Thank you for all the questions and comments! We tried to address most of them!
- AV - obrigado