ERJ open research THE BEST IN OPEN ACCESS BASIC, TRANSLATIONAL & CLINICAL RESPIRATORY RESEARCH

Review

Blueprint for harmonizing non-standardized disease registries to allow federated data analysis – prepare for the future

Johannes A. Kroes, Aruna T. Bansal, Emmanuelle Berret, Nils Christian, Andreas Kremer, Anna Alloni, Matteo Gabetta, Chris Marshall, Scott Wagers, Ratko Djukanovic, Celeste Porsbjerg, Dominique Hamerlijnck, Olivia Fulton, Anneke ten Brinke, Elisabeth H. Bel, Jacob K. Sont

Please cite this article as: Kroes JA, Bansal AT, Berret E, *et al*. Blueprint for harmonizing non-standardized disease registries to allow federated data analysis – prepare for the future. *ERJ Open Res* 2022; in press (https://doi.org/10.1183/23120541.00168-2022).

Blueprint for Harmonizing Non-Standardized Disease Registries to Allow Federated Data Analysis – prepare for the future

Johannes A. Kroes[1], Aruna T. Bansal[2], Emmanuelle Berret[3], Nils Christian[4], Andreas Kremer[4], Anna Alloni[5], Matteo Gabetta[5], Chris Marshall[6], Scott Wagers[7], Ratko Djukanovic[8], Celeste Porsbjerg[9], Dominique Hamerlijnck[10], Olivia Fulton[10], Anneke ten Brinke[11], Elisabeth H. Bel [12], Jacob K. Sont[13]

Affiliations

1 Department of Clinical Pharmacy and Pharmacology, Medical Centre Leeuwarden, Leeuwarden, The Netherlands
2 Acclarogen Ltd, Cambridge, UK
3 European Respiratory Society, Lausanne, Switzerland
4 ITTM S.A., Esch-sur-Alzette, Luxembourg
5 Biomeris SRL, Pavia, Italy
6 Metaseq Ltd
7 BIOSCI Consulting, Maasmechelen, Belgium
8 NIHR Southampton Respiratory Biomedical Research Unit, Faculty of Medicine, University Southampton, Southampton, UK.
9 Dept of Pulmonology, Bispebjerg Hospital, University of Copenhagen, Copenhagen, Denmark
10 Patient Advisory Group, European Lung Foundation (ELF), Sheffield UK
11 Department of Pulmonology, Medical Centre Leeuwarden, Leeuwarden, The Netherlands
12 Amsterdam Medical Centers, Location AMC, University of Amsterdam, The Netherlands
13 Department of Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Center, Leiden, The Netherlands

Corresponding author:
Elisabeth H. Bel, MD, PhD
Amsterdam University Medical Centers
Location AMC
Department of Pulmonology
ORCID-ID 000-0002-6105-3574
E-mail: e.h.bel@amsterdamumc.nl

Word count abstract:	200
Word count body text:	2770 (excluding Table)

Abstract

Real-world evidence from multinational disease registries is becoming increasingly important not only for confirming the results of randomized controlled trials, but also for identifying phenotypes, monitoring disease progression, predicting response to new drugs, and early detection of rare side effects. With new open access technologies, it has become feasible to harmonize patient data from different disease registries and use it for data analysis without compromising privacy rules. In this article, we provide a blueprint for how a clinical research collaboration can successfully use real-world data from existing disease registries to perform federated analyses. We describe how the European Severe Asthma Clinical Research Collaboration SHARP fulfilled the harmonization process from non-standardized clinical registry data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) and built a strong network of collaborators from multiple disciplines and countries. The blueprint covers organizational, financial, conceptual, technical, analytical and research aspects and discusses both the challenges and the lessons learned. All in all, setting up a federated data network is a complex process that requires thorough preparation, but above all, it is a worthwhile investment for all clinical research collaborations, especially in view of the emerging applications of artificial intelligence and federated learning.

Introduction

Targeted biologic therapies have significantly improved the lives of many patients with chronic inflammatory diseases such as rheumatoid arthritis, ulcerative colitis and asthma.[1-3] Unfortunately, biologic therapies are expensive, while it is often unclear which patients benefit most from a particular biological agent.[4-6] National disease registries have therefore been set up in many countries at the initiative of governments, insurers or medical associations to monitor the effectiveness, costs and side effects of biologics.[7]

In the case of severe asthma, individual national registries have yielded interesting publications, although many important research questions including rare adverse effects or comparative effectiveness of different biologics could not be answered due to a lack of sufficient statistical power and reproducibility.[8-12] In addition, real-world evidence from multinational disease registries became increasingly important not only for confirming the results of randomized controlled trials, but also for identifying phenotypes, monitoring disease progression, and targeting the right biologic to the right patient.[13]

Meanwhile, the European Respiratory Society (ERS) had encouraged and financially supported the establishment of a clinical research collaboration (CRC) called SHARP (Severe Heterogeneous Asthma Research, Patient-centered).[14] The ambition of SHARP was to connect all existing severe asthma registries in Europe. To that end, patient data from different registries had to be harmonized to allow data-analyses in such a way that would not compromise the privacy of patients. Because some registries were reluctant to transfer patient data outside the institution where it was collected, SHARP opted for a federated analysis approach, which uses patient-level data from different sources without actually pooling the data together in a central database.

Several harmonization and federation approaches, platforms and structures were considered.[15-20]. SHARP decided to use the open source Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), developed by the Observational Health Data Sciences and Informatics Program (OHDSI), which is currently one of the top-rated models for sharing medical data.[21] This model best meets criteria such as content coverage, integrity, flexibility, ease of retrieval, compatibility of standards and ease/scope of implementations, privacy and connectivity.[22,23] Importantly, the OHDSI/OMOP CDM is the standard used by European Health Data Evidence Network (EHDEN), which is a key initiative that sets the pace for federated analytics in Europe and the US[24]. Thus, OMOP offered great potential for connection to this fast-growing network.

In this article, we describe the harmonization process that SHARP has gone through and provide a blueprint for how to successfully use real-world data from existing disease registries to perform federated analysis. The blueprint covers organizational, financial, conceptual, technical, analytical and research aspects and discusses both the challenges and the lessons learned. The blueprint can be used as a guide for other clinical research networks with a similar ambition to link registries containing patient data.

Harmonization of severe asthma registries

SHARP's initiative to link data from disease registries from different countries was not only ambitious, but also innovative and unique, as no previous examples of this had been published before. Initially, the whole project seemed unfeasible due to the incompatibility of the local data models. Each country had its own electronic case reports of (eCRF) and database structure, in its own language. In addition, legal and regulatory requirements and strict data protection and privacy regulations (e.g., the General Data Protection Regulation (GDPR)) restricted the transfer of patient-level data outside a healthcare provider.[25] Transfer of data outside the country of origin was excluded.

With the ODHSI/OMOP CDM it seemed feasible to meet these challenges.[21] Following the initiative of the European Health Data Evidence Network (EHDEN), research studies would be conducted in a federated manner so that personal data would remain on the local sites, thus retaining full control over what happened to their data and what studies they would participate in.[24] In particular, the harmonization process would remove patient identifiers and, furthermore, only aggregated summary statistics would be exported for meta-analysis. Since aggregated data are privacy-proof by nature, federated analyses comply with the GDPR and ethical research guidelines.

Without previous examples on how to harmonize non-standardized disease registries and build a federated analysis platform (FAP) SHARP wasn't quite sure what to expect. On paper, the procedure seemed simple (Fig. 1): match the field names from the local database with concepts in the CDM; create an Extract, Transform, Load (ETL) procedure to automate the mapping of the local database to a unified format; make the translated data available for local analysis; perform an identical analysis on each registry; combine the aggregated results. However, the reality was that we had to overcome challenges at the organisational, financial, conceptual, technical, analytical and research levels.

Key learnings

In the course of the harmonization process, SHARP learned a number of important lessons, which it would like to share here with other clinical research collaborations that also have the ambition to implement such harmonization. These lessons are listed below by category.

Basic operational pre-requisites

In order for a harmonization process between existing disease registries to be successful, a number of general preconditions must be met. These concern professional project management, availability of sufficient financial resources and signed collaboration agreements between all parties. In addition, it must be ensured that the local ethics committees, the institutions and the patients have given written informed consent for the use of their medical data for scientific research.

As the first to gain experience with this complex harmonization process, SHARP was not well prepared for these preconditions. Until then, it had only collected summary data from the various European regsitries with little financial support.[26] The administrative burden quickly became a challenge for the limited support of the ERS and a dedicated, full-time project manager had to be appointed. In addition, legal services in order to establish service- and research agreements, a professional statistician and the EHDEN-trained SME's (Small and Mid-sized Enterprises) responsible for the mapping of variables in the local databases to the OMOP CDM and for the building of a FAP, were all necessary and all had to be paid. All in all, a budget of around € 200,000 per annum was required to cover these expenses.

Understanding the OHDSI/OMOP CDM

An absolute requirement for succesfully building a FAP is that every stakeholder understands the harmonization concept well and has no doubts or hesitation in participating in its implementation.

For SHARP the use of OHDSI/OMOP CDM for the harmonization of patient-level data was new and conceptually different from the traditional use of such data for scientific research.[27] Time and again, SHARP encountered lack of confidence in the OHDSI/OMOP concept.

This was mainly due to insufficient familiarity with the concept and lack of knowledge and understanding. Clinicians were concerned that patients' privacy was not sufficiently guaranteed. Local legal officers were unsure whether the data handling was secure enough, registry owners were unsure about data ownership, researchers were concerned that their data could be misused by competitors, and IT administrators were reluctant to give third parties access to their servers, due to regulatory concerns or internal IT procedures. Only intensive and repeated education and communication allowed the various parties and partners to ultimately be convinced and enthusiastically take part in the project.

Mapping registry data to the OMOP CDM

A key part of the harmonization process is the mapping of source data to the OMOP CDM. Due to diversity of format and language of the SHARP registries, this had to be manually conducted for each registry, one at a time. The process required fluent and efficient collaboration between the project manager, clinical expert, source data expert, medical terminologist/mapping expert, developer/tester, and statistician.

Not surprisingly, the mapping process faced several challenges, including incomplete registering at source, for example the lack of start and stop dates of medications, and dates when various procedures had taken place. Ideally, the mapping process should be performed on the basis of a registry 'data dictionary' - i.e. a file containing variable names, data types, units of measure, etcetera, because this enables the use of existing mapping tools. In SHARP, the registries could not provide such a data dictionary. The mapping process therefore required a more "iterative" approach than expected, as there were many "mismatches" between the data types and the actual content of the source. All these issues could only be resolved by joining forces. Unfortunately for SHARP, in-person communication was severely hampered by the COVID-19 pandemic and the lock-down measures.

IT requirements and data access

The mapping of source data to the OHDSI/OMOP CDM is automated in an ETL (Extract, Transform, Load) procedure, reading the source data and writing the harmonized data into an OMOP CDM compatible database. Smooth operation requires a server located in the registry's data center (or in a cloud environment, if local IT regulations allow) for taking snapshots of de-identified source data. The server can also host the analytical tools (R environment, OHDSI tools), alternatively these tools can be hosted in a dedicated environment. Of course, the local servers should be accessible by the SME, but for SHARP this proved to be difficult in some cases due to local IT regulations. Nevertheless, it is highly recommended to establish access for the SME, since otherwise local IT teams have to be trained to fulfil the job.

Data quality assessments

In order to obtain the best quality of harmonized data and minimal loss of original data, it is important that source data comply with the rules of the data dictionary, which was not always the case. For a successful mapping between registry data and OMOP CDM, it is therefore important to test and validate the data quality. To this end, SHARP deployed a professional statistician who could form a bridge between the clinicians and the mapping and source data expert. This statistician wrote R scripts for descriptive statistical analysis that could be performed automatically by all local registers. Due to the diversity of the registry structures and the different levels of completeness of each variable considered, the R script at each stage had to include checks on the numerical range and to account for high levels of (or complete) missing data. The local registries were then presented with their own data overviews in well-arranged tables and graphic displays. Ideally such checks should eventually be performed on all variables of each registry before finalising the mapping.

At SHARP, quality checks revealed unexpected missing data codes, impossible values and some mismatches due to the use of free text fields by the clinicians who had entered data. Where necessary, changes were made to the mapping schema and in some cases to the source data in the local registry database. Again, these solutions required time, close collaboration between clinicians, source data experts, mapping experts and data analysts.

Data analytical aspects

Using a FAP and analyzing real-world data from different disease registries in different countries requires strong analytical skills. In fact, the person in question must unite epidemiological, biostatistical and observational data science expertise, be a confident programmer, and be willing to learn the ins and outs of the OHDSI/OMOP CDM Also, the statistician should be able to perform an appropriate meta-analysis of summary statistics to draw conclusions from all participating registers. Of course, and luckily, more than one person may fulfil different aspects of this role in the studies.

While processing data from the SHARP registries, it became clear that a statistician be engaged at the outset of the project and be involved in the writing of all protocols and analysis plans. This helps to ensure that the necessary data is available and mapped across all relevant registries, and that any local categorization of data does not preclude the planned analysis.

Recommendations and blueprint

Table 1 shows the blueprint with recommendations for an optimal harmonization process between disease registry data and OMOP CDM for multinational federated analyses.

Table 1. Blueprint for harmonising disease registries using OMOP CDM.

Topic	Recommendation
Basic conditions	- Selection of a legal body for clinical research collaboration (CRC) - Securing of sufficient financial resources for ≥3 years - Appointment of a full-time dedicated project manager - Establishment of a contract with an SME specializing in OHDSI, OMOP CDM and mapping - Establishment of contract with a hands-on statistician with programming skills - Written confirmation from each registry that patients have given written consent to use their medical data for (international) clinical research - Identification for each local registry of named individuals in the following roles: - Registry owner - Legal officer - Clinical expert - Source data expert - IT contact/administrator - Translator of medical terminology - Platform/System user - Conclusion of collaboration agreements between CRC and registries
Conceptual aspects	- Production of a document and a Power Point presentation explaining the OMOP CDM and the federated approach to all stakeholders - Organization of a plenary kickoff meeting with all stakeholders - Organization of regular team meetings for each registry to monitor progress
Technical aspects	- Provision/hire of a dedicated Linux server for each registry (local data center or cloud environment) for the installation and setup of the FAP, with access to a local copy of the source database; - Provision to all required parties of access to the Linux registry

	servers - Testing of the functioning of the FAP on local Linux servers by SME
Mapping aspects	- Checks source data quality - Provision of registry data dictionary to SME by source data experts - Provision of a representative, but anonymized registry data sample by local team to smoothen ETL process and avoid "black box mapping" - Assistance by clinical experts in optimizing the mapping - Provision by SME to statistician(s) of a codebook of the variables mapped
Analytical aspects and Quality control	- Learning by statistician(s) on the principles of OHDSI and OMOP comon data model - Provision by SME of access to FAP for statistician(s) - Creation by statistician of scripts in R (or OHDSI tools for the production of descriptive summary statistics - Execution by local analyst in each country of the pre-written R-script via the FAP - Checks by clinical on the validity of the output and provision of feedback to statistician and SME - Revision by source data expert and SME of any mapping issues. - Creation of a second round of data summaries and a repeat of the quality control process - Production of final OMOP CDM tables
Research studies	- Creation of research protocol and approval by CRC, local clinical experts and registry owners - Identification of dedicated local teams for each registry, comprising clinical experts, source data experts and data analysts. - Creation of a formal analysis plan by a statistician, for review and approval by representatives of all participating registries - Creation by statistician of analysis scripts in R (or OHDSI tools) - Execution by local data analysts of pre-written scripts in R (or ODHSI tools) using the FAP.

	- Fostering of collaboration between best practices for statisticians and data analysts via workshops to discuss issues like imputation rules, filters and exclusions - Production of final statistical tables and graphics for each registry singly, according to the analysis plan - Meta analysis by statistician of summary statistics from all registries - Writing and submission of manuscript

CDM: Common Data Model; CRC: Clinical research Collaboration; ETL: Extract, Transform, Load; FAP: Federated Analysis Platform; IT: Information Technology; OHDSI: Observational Heath Data Sciences and Informatics; OMOP: Observational Medical Outcomes Partnership; SME: Small and Medium-sized Enterprise.

A schematic summary of required steps for harmonizing disease registries using OHDSI/OMOP CDM is given in Figure 2. An estimate of the time required per item is given in Table E1 and Figure E1. Registries that are currently connected or in the process of being connected are listed in table E2.

Discussion

In this article, we described our experience in harmonizing patient data from different European severe asthma registries using the OHDSI/OMOP CDM. Based on the lessons learned, we put together a blueprint that can be used by researchers in other disease areas where there is a desire to establish federated data networks of real-world patient data already collected in non-standardized registries. The harmonization process was not without challenges, but it was above all a unique experience to connect colleagues and partners from different countries, specialties and disciplines in one large federated project.

To date, most studies on OHDSI/OMOP CDM were related to architectural concepts and tool development.[27] However, over the last couple of years, an increasing number of publications have appeared using the OMOP CDM in prospective network studies with observational patient data, in particular related to the COVID-19 pandemic.[28-32]. Other studies have used large administrative claims databases [33,34] or electronic medical records databases.[35,36] Our study is the first that used the OMOP CDM to harmonize non-standardized national disease registries.

When SHARP CRC was founded in 2017, its vision was to incrementally change the research culture across Europe, emphasizing ambitions that serve the collective needs of the asthma research community and bring people with asthma to the center of the research environment into a reality context.[14] SHARP's goals included better understanding the mechanisms of severe asthma, improving treatment for severe asthma, and exploring ways to prevent severe asthma. It wanted to achieve this by establishing a platform that would allow the integration of local national asthma registries into a pan-European multicenter registry of patients with severe asthma.[21] At the same time, the scientific community expressed the increasing need for more large-scale real-world research. Not only for confirming the results of randomized controlled trials, but also for identifying phenotypes, monitoring disease progression, predicting response to new drugs and detecting rare side effects.[38,39] However, due to concerns regarding data privacy, data security, data access rights and data ownership, some SHARP registries were reluctant to transfer patient-level data to one central database, as was the case with other international registries such as the International Severe Asthma Registry ISAR.[40]. However, in order not to lose the precious data from these existing registries, it was then decided to establish a federated data platform and use the OHDSI/OMOP CDM to harmonize the databases.[21]

At that time, the use of OMOP CDM was relatively new and had never been applied to existing disease registries. Since there was no example of how to approach the harmonization process, it was not surprising that SHARP encountered multiple challenges and obstacles, from which it ultimately learned a lot.

In retrospect, the unfamiliarity and misunderstanding of the OMOP CDM concept among doctors, researchers, legal entities and IT administrators was perhaps the main reason why the process was sometimes unnecessarily delayed. There were concerns that data privacy would not be guaranteed, data would fall into the wrong hands and the security of data centers would be compromised. Therefore, we cannot emphasize enough the need to repeatedly explain the concept and process of harmonization to all stakeholders, through meetings, presentations and personal discussions.

Furthermore, it appeared that collaboration between clinicians, IT technicians, registration holders and legal entities was essential, and that they all should be able to devote sufficient time and attention to the project. Not only for the initial harmonization process, but also prior to any future research project, such multidisciplinary dedicated teams should be set up for each registry. Team members should be able to consult each other easily and ad hoc, preferably by mobile phone.

Investing in building the FAP and achieving the harmonization of severe asthma registries has brought many benefits to SHARP CRC. Firstly, thanks to the joint effort and overcoming adversity, it has created a strong and solid partnership between many stakeholders including patients, clinicians, researchers, pharmaceutical industries, IT technicians, data analyst and consultants. Secondly, it now features a state-of-the-art platform that allows for innovative and large-scale real-word studies with relatively little effort. Finally, and perhaps most importantly, because of its privacy-protected structure, scalability and generalization, the SHARP FAP is now perfectly equipped for the future in which artificial intelligence and federated learning will play an increasingly important role in generating evidence with real-world data.[41-43

Conclusion

We have provided a blueprint for what it takes as a nonprofit clinical research collaboration to successfully use real-world data from existing disease registries for executing federated analyses. The open access OHDSI/OMOP CDM has enabled patient data from different disease registries to be harmonized and used for data analysis without compromising privacy rules. We have learned that building a FAP to enable large-scale analysis of patient-level data from non-standardized registries is a complex process and can only be successful if all parties fully understand and support the concept. At the same time, it ensures strong collaboration and builds an enriching network that enhances the knowledge and interrelationships of all partners with the common goal of using real-word data efficiently. We believe that, especially given the increasing adoption of artificial intelligence and federated learning, the harmonization of disease registry data to a common data model is a worthwhile investment, which we can certainly recommend to other clinical research collaborations. Ultimately, the rewards of such efforts will manifest in terms of improved disease understanding and better patient care.

Acknowledgement

We wish to recognise the contribution of the National leads of the SHARP countries for their help to build the network and for the communication with all the actors involved in the process of building the SHARP Federated Analysis Platform

References

1. Olsen NJ, Stein CM. New drugs for rheumatoid arthritis. N Engl J Med. 2004;350: 2167-2179.

2. Danese S, Fiocchi C. Ulcerative colitis. N Engl J Med. 2011;365: 1713-1725.

3. Israel E, Reddel HK. Severe and Difficult-to-Treat Asthma in Adults. N Engl J Med. 2017;377: 965-976.

4. Smolen JS, Aletaha D. Rheumatoid arthritis therapy reappraisal: strategies, opportunities and challenges. Nat Rev Rheumatol. 2015;11: 276-289.

5. Ma C, Battat R, Dulai PS, et al. Innovations in Oral Therapies for Inflammatory Bowel Disease. Drugs. 2019;79: 1321-1335.

6. Wenzel SE. Severe Adult Asthmas: Integrating Clinical Features, Biology, and Therapeutics to Improve Outcomes. Am J Respir Crit Care Med. 2021;203: 809-821.

7. van Bragt JJMH, Adcock IM, Bel EHD, et al. Characteristics and treatment regimens across ERS SHARP severe asthma registries. Eur Respir J. 2020;55(1):1901163

8. Jackson DJ, Busby J, Pfeffer PE, et al. Characterisation of patients with severe asthma in the UK Severe Asthma Registry in the biologic era. Thorax. 2021;76: 220-227.

9. Graff S, Brusselle G, Hanon S, et al. Anti-Interleukin-5 Therapy Is Associated with Attenuated Lung Function Decline in Severe Eosinophilic Asthma Patients from the Belgian Severe Asthma Registry. J Allergy Clin Immunol Pract. 2022;10: 467-477.

10. Eger K, Kroes JA, Ten Brinke A et al. Long-Term Therapy Response to Anti-IL-5 Biologics in Severe Asthma-A Real-Life Evaluation. J Allergy Clin Immunol Pract. 2021;9: 1194-1200.

11. Heffler E, Detoraki A, Contoli M, et al. COVID-19 in Severe Asthma Network in Italy (SANI) patients: Clinical features, impact of comorbidities and treatments. Allergy. 2021;76: 887-892.

12. Sá-Sousa A, Fonseca JA, Pereira AM, et al. The Portuguese Severe Asthma Registry: Development, Features, and Data Sharing Policies. Biomed Res Int. 2018;2018: 1495039.

13. Pavord ID, Hanania NA, Corren J. Choosing a Biologic for Patients with Severe Asthma. J Allergy Clin Immunol Pract. 2022;10:410-419.

14. Djukanovic R, Adcock IM, Anderson G, et al. The Severe Heterogeneous Asthma Research collaboration, Patient-centred (SHARP) ERS Clinical Research Collaboration: a new dawn in asthma research. Eur Respir J. 2018;52:1801671

15. Open Algorithms (OPAL). https://www.opalproject.org/. last accessed 7 June 2022

16. DataSHIELD. Secure Bioscience Collaboration. https://www.datashield.org/ last accessed 7 June 2022

17. I2b2 TRANSMART. An Open-Source—Open-Data Community. https://i2b2transmart.org/ last accessed 7 June 2022

18. The Personal Health Train Network. https://pht.health-ri.nl/. last accessed 7 June 2022

19. Clinerion Real-World Solutions. https://www.clinerion.com/. last accessed 7 June 2022

20. TriNetX | Real-world data for the life sciences and healthcare. last accessed 7 June 2022

21. Observational Health Data Sciences and Informatics. OMOP Common Data Model. Available from: https://www.ohdsi.org/data-standardization/the-common-data-model . Date last updated: February 14, 2022. Date last accessed: February 18, 2022.

22. Wirth FN, Meurers T, Johns M, Prasser F. Privacy-preserving data sharing infrastructures for medical research: systematization and comparison. BMC Med Inform Decis Mak. 2021;21:242

23. Garza M, Del Fiol G, tenenbaum J, Walden A, Zozus MN. Evaluatig common data models for use with a longitudinal community registry. J Biomed Infrom.2016:64:333-341

24. EHDEN. European Health Data & Evidence Network (EHDEN). Available from: http://www.ehden.eu Date last updated: February 10, 2022. Date last accessed: February 18, 2022.

25. Bentzen HB, Høstmælingen N. Balancing Protection and Free Movement of Personal Data: The New European Union General Data Protection Regulation. Ann Intern Med. 2019;170: 335-337.

26. van Bragt JJMH, Hansen S, Djukanovic R, et al. SHARP: enabling generation of real-world evidence on a pan-European scale to improve the lives of individuals with severe asthma. ERJ Open Res. 2021;7: 00064-2021.

27. Reinecke I, Zoch M, Reich C, et al. The Usage of OHDSI OMOP - A Scoping Review. Stud Health Technol Inform. 2021;283: 95-103.

28. Reyes C, Pistillo A, Fernández-Bertolín S, et al. Characteristics and outcomes of patients with COVID-19 with and without prevalent hypertension: a multinational cohort study. BMJ Open. 2021;11: e057632

29. Recalde M, Roel E, Pistillo A, et al. Characteristics and outcomes of 627 044 COVID-19 patients living with and without obesity in the United States, Spain, and the United Kingdom. Int J Obes (Lond). 2021;45: 2347-2357.

30. Prats-Uribe A, Sena AG, Lai LYH, et al. Use of repurposed and adjuvant drugs in hospital patients with covid-19: multinational network cohort study. BMJ. 2021;373: n1038.

31. Tan EH, Sena AG, Prats-Uribe A, et al. COVID-19 in patients with autoimmune diseases: characteristics and outcomes in a multinational network of cohorts across three countries. Rheumatology (Oxford). 2021;60: SI37-SI50.

32. Lane JCE, Weaver J, Kostka K, et al. Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study. Lancet Rheumatol. 2020;2: e698-e711.

33. Williams RD, Reps JM, OHDSI/EHDEN Knee Arthroplasty Group, et al. 90-Day all-cause mortality can be predicted following a total knee replacement: an international, network study to develop and validate a prediction model. Knee Surg Sports Traumatol Arthrosc. 2021. (online ahead or print)

34. Nestsiarovich A, Reps JM, Matheny ME, et al. Predictors of diagnostic transition from major depressive disorder to bipolar disorder: a retrospective observational network study. Transl Psychiatry. 2021;11: 642.

35. Kim GL, Yi YH, Hwang HR, et al. The Risk of Osteoporosis and Osteoporotic Fracture Following the Use of Irritable Bowel Syndrome Medical Treatment: An Analysis Using the OMOP CDM Database. J Clin Med. 2021;10: 2044.

36. Mun Y, You SC, Lee DY, et al. Real-world incidence of endophthalmitis after intravitreal anti-VEGF injection: Common Data Model in ophthalmology. Epidemiol Health. 2021: e2021097.

37. Biedermann P, Ong R, Davydov A, et al. Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases. BMC Med Res Methodol. 2021;21: 238 30. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence - What Is It and What Can It Tell Us? N Engl J Med. 2016;375: 2293-2297.

38. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence - What Is It and What Can It Tell Us? N Engl J Med. 2016;375: 2293-2297.

39. Bartlett VL, Dhruva SS, Shah ND, et al. Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence. JAMA Netw Open. 2019;2: e1912869.

40. International Severe Asthma Registry ISAR. https://isaregistries.org/. last accessed 20-5-2022

41. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019;6:94-8

42. Rubinger L, Gazendam A, Ekhtiari S, Bhandari M. Machine learning and artificial intelligence in research and healthcare☆,☆☆. Injury. 2022 Feb 1:S0020-1383(22)00076-6.

43. Sadilek A, Liu L, Nguyen D, et al. Privacy-first health research with federated learning. NPJ Digit Med. 2021 Sep 7;4(1):132.

Figure 1. Architecture of the Federated Analysis Platform

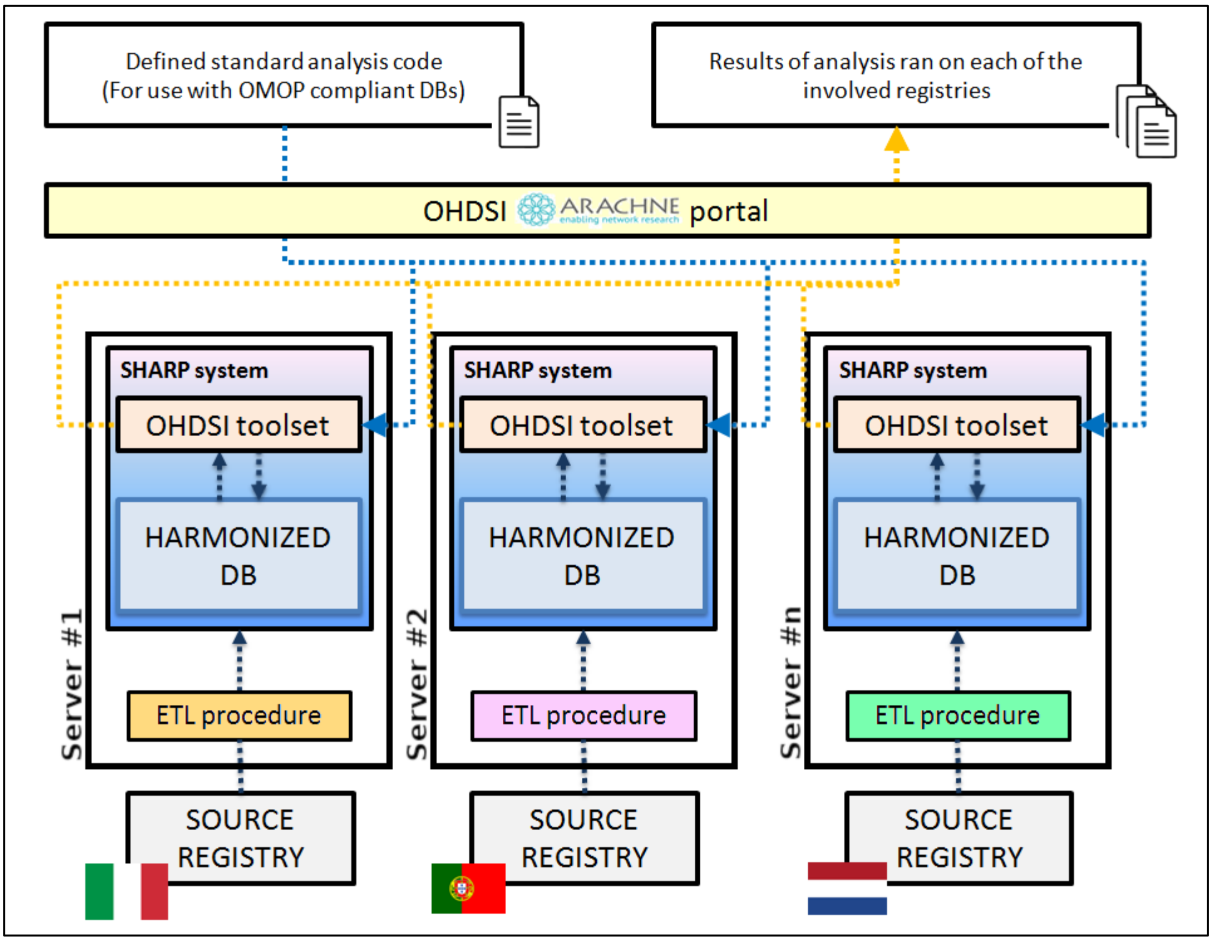

Defined standard analysis code
(For use with OMOP compliant DBs)
Results of analysis ran on each of the involved registries
OHDSI ARACHNE portal
enabling network research
SHARP system
OHDSI toolset
HARMONIZED DB
ETL procedure
SOURCE REGISTRY
Server #1
Server #2
Server #n

FIGURE 1. Field names of the different national registries are mapped to concepts in the common data model. An ETL procedure is created to automate the mapping from the local database into a unified format; the harmonized data are made available for local analysis using the OHDSI toolset or R-code; an identical analysis is run on each registry; the results are combined using federated analysis tools. DB: database; ETL: Extract, Transform, Load; OHDSI: Observational Health Data Sciences and Informatics; OMOP: Observational Medical Outcomes Partnership; SHARP: Severe Heterogeneous Asthma Registry – Patient Centred.

Figure 2. Schematic summary of harmonization steps.

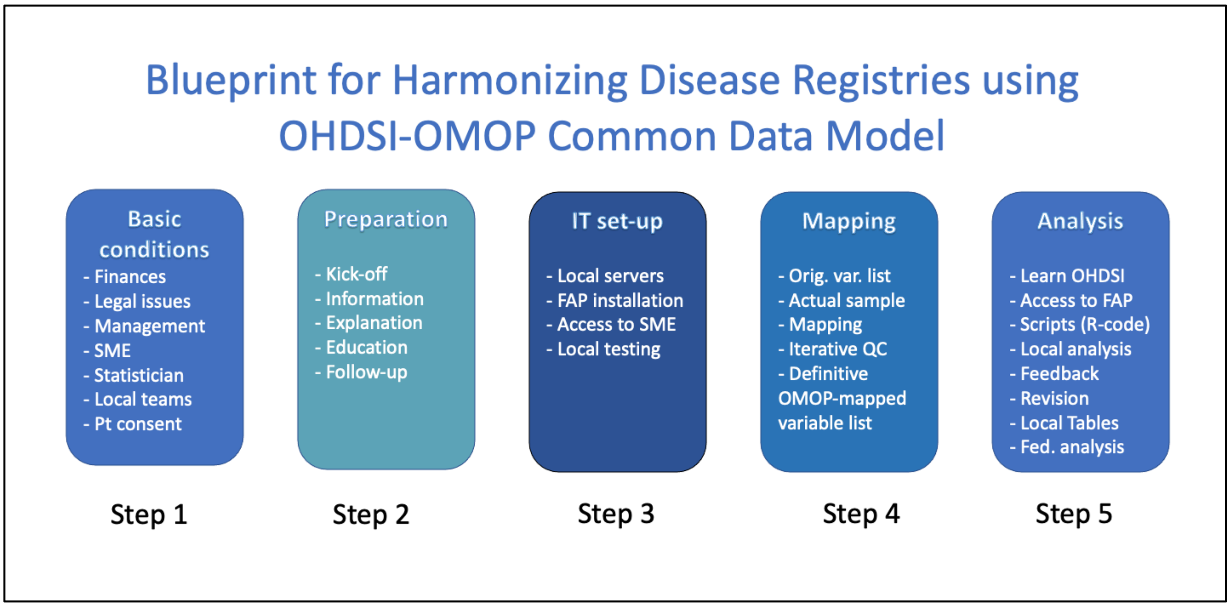


Figure 2. Schematic summary of steps to be taken for a successful harmonization process of local non-standardized disease registries to the OHDSI/OMOP Common Data Model for federated analyses. FAP: Federated Analysis Platform; Fed: Federated; IT: Information Technology; OHDSI: Observational Health Data Sciences and Informatics; OMOP: Observational Medical Outcomes Partnership; Orig: Original; Pt: Patient; QC: quality Check; SME: Small and Medium-sized Enterprise; Var: variable

Blueprint for Harmonizing Non-Standardized Disease Registries to Allow Federated Analysis – prepare for the future

ONLINE REPOSITORY APPENDIX

Content

Table E1. Estimation of the time required for building a FAP

Topic	Tasks	Estimated average time needed
Basic conditions	Setting-up a collaboration network/consortium - Writing of a protocol and governance document - Selection of a legal body (foundation/society) for a clinical research collaboration - Securing of sufficient financial resources for ≥3 years - Appointment of a full-time dedicated project manager	10 months
	- Establishment of a contract with an SME specializing in OHDSI, OMOP CDM and mapping - Establishment of contract with a hands-on statistician with programming skills - Written confirmation from each registry that patients have given written consent to use their medical data for (international) clinical research - Identification for each local registry of named individuals in the following roles: - Registry owner - Legal officer - Clinical expert - Source data expert - IT contact/administrator - Translator of medical terminology - Platform/System user - Conclusion of collaboration agreements between CRC and registries	8 months
Conceptual aspects	- Production of documents and a Power Point presentation explaining the OMOP CDM and the federated approach to all stakeholders - Organization of a plenary kick-off meeting with all stakeholders	3 months
	- Organization of regular team meetings for each registry to monitor progress	Per registry 2h/week
Technical aspects	- Provision/hire of a dedicated Linux server for each registry (local data centre or cloud environment) for the installation and setup of the FAP, with access to a local copy of the source database;	Per registry 2 months

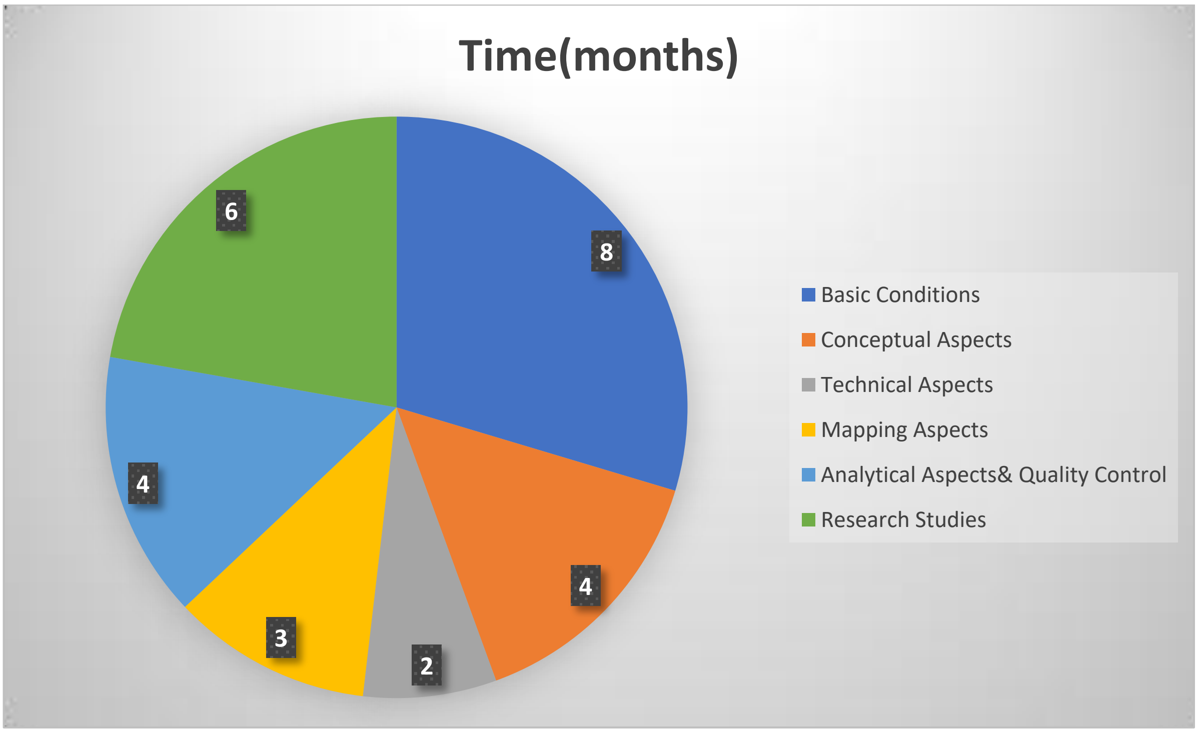
	- Provision to all required parties of access to the Linux registry servers - Testing of the functioning of the FAP on local Linux servers by SME	
Mapping aspects	- Checks source data quality - Provision of registry data dictionary to SME by source data experts - Provision of a representative, but anonymized registry data sample by local team to smoothen ETL process and avoid “black box mapping” - Assistance by clinical experts in optimizing the mapping - Provision by SME to statistician(s) of a codebook of the variables mapped	Per registry 3 months
Analytical aspects and Quality control	- Learning by statistician(s) on the principles of OHDSI and OMOP common data model - Provision by SME of access to FAP for statistician(s) - Creation by statistician of scripts in R (or OHDSI tools for the production of descriptive summary statistics	1 month
	- Execution by local analyst in each country of the pre-written Rscript via the FAP - Checks by clinical on the validity of the output and provision of feedback to statistician and SME Revision by source data expert and SME of any mapping issues. - Creation of a second round of data summaries and a repeat of the quality control process - Production of final OMOP CDM tables	Per registry 3 months
Research studies	- Creation of research protocol and approval by CRC, local clinical experts and registry owners - Identification of dedicated local teams for each registry, comprising clinical experts, source data experts and data analysts. - Creation of a formal analysis plan by a statistician, for review and approval by representatives of all participating registries - Creation by statistician of analysis scripts in R (or OHDSI tools)	Depending the magnitude and complexity of the study ≥6 months

Table E2. Countries engaged in SHARP CRC and their registry's status for the SHARP FAP

SHARP Countries	Registry Name	Status: Connexion to SHARP FAP	Comments
Austria	ASA-Net: Austria Severe Asthma Net	Not Connected	Under communication to integrate SHARP Central Registry
Belgium	BSAR: Belgium Severe Asthma Registry	Connected	
Croatia	SHARP Central	Connected	
Czech Republic	GAN: German Asthma Network	Connected	
Denmark	DSAR: Danish Severe Asthma Registry	Connection ongoing	
Estonia	SHARP Central	Connected	
Finland		Not Connected	Under communication to integrate the FAP
France	RAMSES: The French registry of severe asthma patients	Connected	
Germany	GAN	Connected	
Greece	HTS-SAR: Hellenic Thoracic Society - Severe Asthma Registry	Connected	
Hungary	SHARP Central	Connected	
Iceland		Not Connected	Under communication to integrate the FAP
Ireland	SHARP Central	Connection ongoing	
Italy	SANI: Severe Asthma Network Italy	Connected	
Latvia	SHARP Central	Connected	
Lithuania	SHARP Central	Connected	
Netherlands	RAPSODI/SHARP Central	Connected	
Poland	SHARP Central	Connected	

Portugal	RAG: Registo de Asma Grave Portugal	Connected	
Romania	SHARP Central	Connected	
Russia		Not Connected	Russian Pulmonary society declined the invite to the SHARP FAP
Serbia	SHARP Central	Connected	
Slovenia	SHARP Central	Connected	
Spain	GEMA: Spanish Asthma Guidelines	Connected	
Sweden	SHARP Central	Connected	
Switzerland	GAN	Connected	
Turkey	SHARP Central	Connected	
United Kingdom	UKSAR: UK Severe Asthma Registry	Connection ongoing	

Figure E1. Pie Chart and time required for building a FAP plus proof of principle study


Time(months)
Basic Conditions
Conceptual Aspects
Technical Aspects
Mapping Aspects
Analytical Aspects& Quality Control
Research Studies