

ADVANCING FAIR RESEARCH: HEMAFAIR'S FIRST PUBLICATIONS WITH INHERENT AND HELIOS NETWORKS

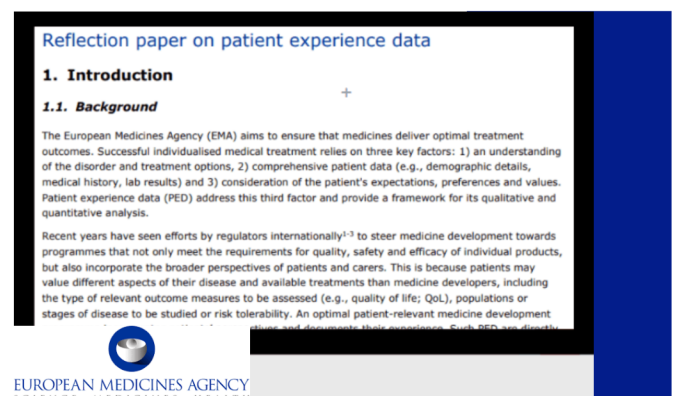


HemaFAIR is pleased to highlight its first scientific publications developed in collaboration with the INHERENT and HELIOS networks. These studies represent important milestones in advancing FAIR data practices, strengthening international research collaboration, and supporting the development of interoperable data infrastructures for hemoglobinopathy research.

Together, these publications contribute to the growing evidence base needed to improve data sharing, harmonisation, and collaborative research across the global hemoglobinopathy community.

HEMAFAIR CONTRIBUTION TO THE EMA REFLECTION PAPER ON PATIENT EXPERIENCE DATA

HemaFAIR partners contributed to the European Medicines Agency (EMA) Reflection Paper on the use of Patient Experience Data (PED) in medicines development and regulatory decision-making. This contribution supports ongoing efforts to strengthen the integration of patient perspectives into healthcare research and innovation, aligning with HemaFAIR's work on patient-reported outcomes, data interoperability, and FAIR health data practices.



HemaFAIR and HELIOS successfully delivered the Federated Data Analysis Training School and Hackathon in Ayia Napa, Cyprus, bringing together participants from multiple countries for hands-on training in FAIR data principles, semantic interoperability, OMOP, and federated data analysis. The event strengthened participants' skills in FAIR data implementation while fostering collaboration and knowledge exchange across the rare disease and hemoglobinopathy research communities.

Stay tuned for event recordings, participant testimonials, and additional highlights showcasing the impact and outcomes of this successful training initiative.



HemaFAIR training materials are now open and available!



TESS
Training
Environment for
Skills & Services

zenodo
Open access
on Zenodo



HemaFAIR
website



Explore and download the materials:
[Zenodo](#) | [TESS](#) | [HemaFAIR Website](#)



Funded by
the European Union



HEMAFAIR LECTURE SERIES CONTINUES TO GROW!

The HemaFAIR Lecture Series continues to bring together leading experts to share knowledge and foster collaboration in the fields of rare diseases, FAIR data, biomedical informatics, and hemoglobinopathy research.

Recent lectures have covered a wide range of topics, providing participants with valuable insights into emerging research, innovative methodologies, and best practices. Through these events, HemaFAIR continues to strengthen capacity building, knowledge exchange, and collaboration across the scientific community.

HemaFAIR Lecture Series Recordings available!



Missed a session? Catch up anytime!
Recordings of all HemaFAIR Lecture Series
are now available on:



HemaFAIR Website

<https://hemafairproject.eu/lectures-2/>



HemaFAIR YouTube Channel

<http://www.youtube.com/@HemaFAIR>



Explore expert talks on FAIR data, interoperability,
innovation, and more in hemoglobinopathy research.



Knowledge shared, progress amplified.
Watch now and stay informed!



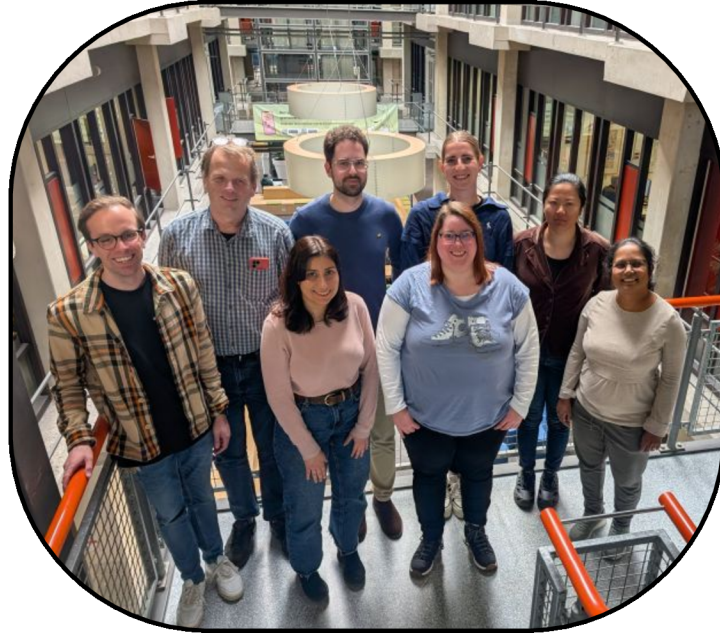
<https://hemafairproject.eu/lectures-2/>



<http://www.youtube.com/@HemaFAIR>



EXCHANGE VISITS BETWEEN HEMAFAIR PARTNERS!



A recent exchange visit from CING to Amsterdam UMC (AUMC) provided a valuable opportunity for collaborative work on the FAIRification of rare hematological disease registries. Through close interaction and knowledge exchange, the teams advanced activities related to data findability, interoperability, and reusability, further strengthening the partnership and supporting HemaFAIR's mission to develop FAIR and interoperable health data infrastructures.

HEMAFAIR PRESENTED AT SWAT4HCLS 2026!

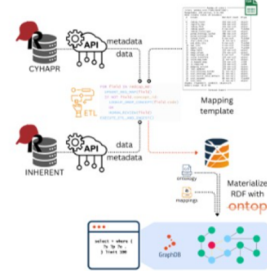
HemaFAIR was presented at SWAT4HCLS 2026 in Amsterdam through the poster "HemaFAIR: Implementing FAIR Principles in Real-World Rare Hematological Disease Datasets." The conference provided a valuable platform to showcase the project's progress in FAIRifying rare hematological disease data and to connect with researchers working on semantic web technologies, interoperability, and FAIR data solutions in health research.

HemaFAIR: Implementing FAIR principles in real-world Rare Hematological Disease datasets

Stella Tamana¹, Christina Yiangou¹, Kallia Orphanou¹, Maria Xenophontos¹, Panayiota L. Papasavva¹, César Bernabé², Marco Rosco², Julio Marchiori Diaz², Martijn G. Kerloot², Ronald Cornet², Anna Minalidou², Coralea Stephanou², Annalisa Landi², Viviana Giannuzzi², Fedele Bonifazi², Carsten W. Lederer², Petros Kountouris²
¹Department of Blood Disorder Genetics & Thalassemia, The Cyprus Institute of Neurology & Genetics, 6 Iron Avenue, 2371 Agios Dometios, Nicosia, Cyprus
²Department of Human Genetics, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands
³Amsterdam UMC location University of Amsterdam, Department of Medical Informatics, Amsterdam, 1105 AZ, The Netherlands
⁴Fondazione per la ricerca farmacologica Gianni Benzi onlus, Bari, Italy

Introduction

Rare Hematological Diseases (RHDs) are characterized by small, geographically dispersed patient populations. Research in this field is significantly impeded by data fragmentation and a lack of standardization across isolated registries. To address this, we apply the Findable, Accessible, Interoperable, and Reusable (FAIR) principles to two key use cases: the national Cyprus Haemoglobinopathy Patient Registry (CYHAP) and the INHERENT platform [3]. Our objective is to transform these sites into a patient-centered data ecosystem capable of supporting secure, cross-registry research.



```

PREFIX hema: <http://hemafair.eu/ontology/hemafair-ontology.owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?diagnosis_label ?variant_label ?status (COUNT(DISTINCT ?p) AS ?n_patients)
WHERE {
  ?p a hema:Patient ;
    hema:hasDiagnosis ?diagnosis ;
    hema:hasTreatmentStatus ?s ;
    hema:hasAgeGroup ?a ;
    ?diagnosis rdfs:label ?diagnosis_label ;
    ?p hema:composedOf ?variant ;
    ?variant a hema:Variant ;
    ?variant rdfs:label ?variant_label ;
    ?p rdfs:label ?status .
}
GROUP BY ?diagnosis_label ?variant_label ?status
ORDER BY DESC(?n_patients) ?diagnosis_label
LIMIT 5
  
```

Figure 2. SPARQL Query & Results: This table presents the results of a SPARQL query executed against the RDF graph in GraphDB, which retrieves the distribution of thalassemia genotypes and their association with transfusion status. The query counts the number of patients for each combination of diagnosis, genetic variant, and transfusion status, ordered by the number of patients in descending order. Only the top 5 results are displayed here.

Results

Our evaluation confirms that standardized schemas are sufficient for preserving necessary data granularity for RHD. We established ETL pipelines for ingesting data for CDS to the OMOP CDM and for CDM to CARE-OM and have successfully materialized RDF triples in GraphDB, with preliminary SPARQL queries validating the graph structure and content (see Figure 2).

To enhance findability, a FAIR Data Point (FDP) [5] was deployed to expose dataset-level metadata. This FDP was assigned a persistent identifier (<https://doi.org/10.26434/chemrxiv-2024-01-01>) and is indexed by the central FAIR Data Point Index.

Both CYHAP and INHERENT were integrated into the European Rare Disease Registry Infrastructure (ERDI). This integration encompassed registration in ERDI and the deposition of metadata in ERDI-UI. Furthermore, the ERDI patient pseudonymization service was implemented for CYHAP to enable GDPR-compliant data linkage.

Discussion

We demonstrated the value of a stepwise, CQ-driven approach for the retrospective FAIRification of RHD platforms, allowing for iterative improvements over accessible transformations. Future work includes regarding data ingestion for all competency questions and FAIRifying Patient Reported Outcome Measures. We will improve accessibility via a dashboard for visualizing SPARQL results and develop a secure, FAIR-compliant application for programmatic access and federated querying across research networks.

References

[1] P. Kerloot, C. Stephanou et al., "The international hemoglobinopathy research network (IHARN): An international initiative to study the rare genetic hemoglobinopathies. Am J Hematol 95(12):1454-1460 (2024).
 [2] C. Bernabé, L. Valsé, et al., "A patient-centred method for FAIRification of patient registries. Int J Med Inform 170:102510 (2024).
 [3] M. A. Kountouris, P. B. Kountouris et al., "Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc 29(10):1702-1710 (2022).
 [4] P. Kountouris, M. B. Kountouris, "The CARE-OM of your patient data. Clinical and Regulatory Science (CARE-OM) Research: An Introduction to the 100 International Collaborative on Rare Diseases Web Application and Data for Health Care and Public Health. OMOP-CDSM 2024, 10:1000-1001 (2024).
 [5] P. A. Haines, E. Taylor et al., "Research data point (RDP) in multi-site data integration: A framework for providing transparent research information rapidly. J Biomed Inform 121:103818 (2023)."

Methods

Our FAIRification strategy began with the Goal-Oriented Planning method [2], identifying six central competency questions (CQs) to align technical steps with clinical needs. Using a stepwise, CQ-driven workflow, we modeled the first CQ (see Box 1) in Protégé [6].

- Box 1. Competency Questions**
1. What is the prevalence of specific thalassemia genotypes (overall, per centre/region, country, ethnic group) and its association with the transfusion status?
 2. How do the annual average of blood transfusions and pre-transfusion hemoglobin levels correlate with yearly quality of life assessments?
 3. What is the incidence of new iron-related (side) complications?
 4. What is the relationship between annual serum ferritin levels and liver/fat/diabetic/indocrine complications?
 5. What is the long-term survival rate of thalassemia patients?
 6. What are the key criteria for initiating regular blood transfusions in beta thalassemia patients, and how do these criteria impact long-term disease outcomes and quality of life assessments?
- Simultaneously, to evaluate existing models against our research needs, we implemented parallel registry-aware ETL workflows for both the OMOP Common Data Model (CDM) [4] and the CARE Semantic Model (CARE-SM) [5].
- The OMOP pipeline (see Figure 1) extracts (metadata from REDCap [7]), mapping them to OMOP concepts for ingestion into the CDM. Ontop then materializes these data into RDF triples for storage in a GraphDB repository. For CARE-SM, a similar process extracts (metadata for the rare Disease CDM Data Elements), to generate class-specific CSVs validated by the CARE-SM toolkit. These are converted into JSON using a MARSHAL Editor and YAML templates before being imported to a GraphDB repository.



WHAT'S NEXT FOR HEMAFAIR?

ON THE HORIZON



CONTINUED STAFF EXCHANGES
Strengthening collaboration and knowledge exchange across partners.

CAPACITY BUILDING
Sharing knowledge and building essential skills together.

FAIR DATA EXCELLENCE
Advancing findability, interoperability and reusability of health data.

Together, we continue to build a FAIR and interoperable future for rare hematological disease research.
 www.hemafairproject.eu/

MORE

