SYNTHEMA + ERN-EuroBloodNet

Joint Training Programme on Synthetic Data Generation in SCD and AML

Funded by the European Union

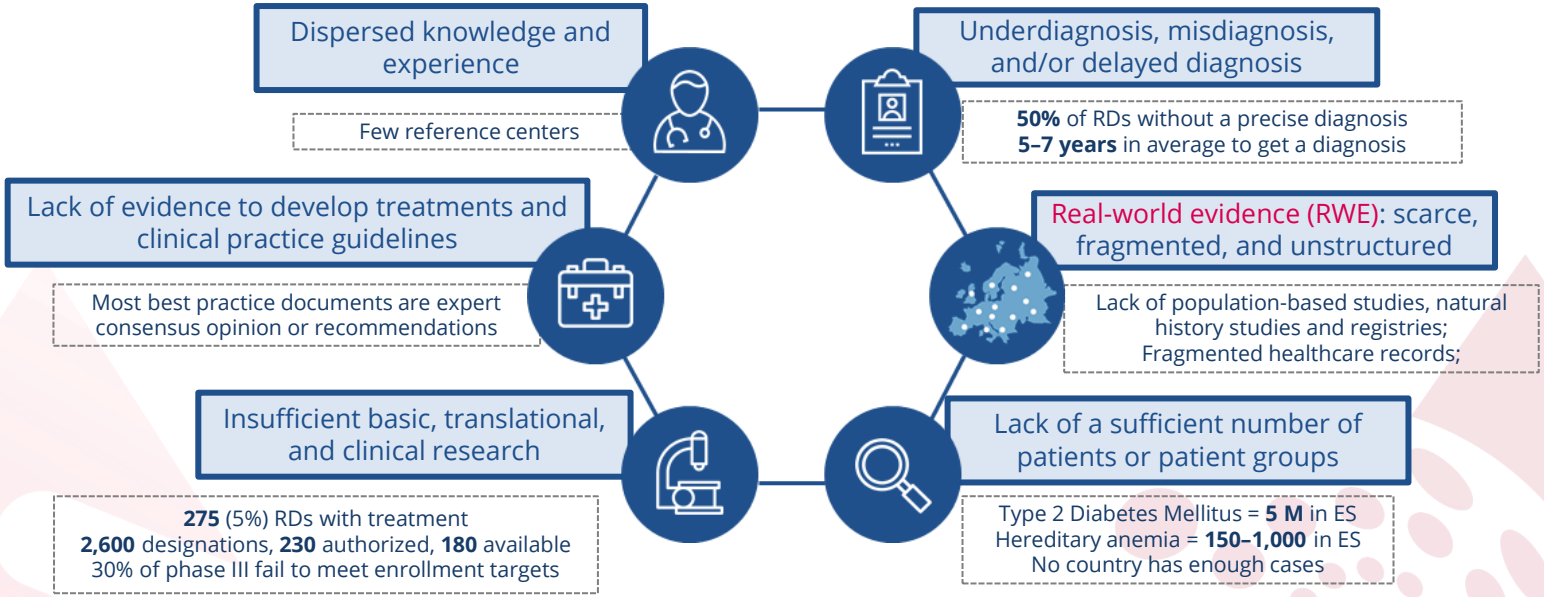
SYNTHEMA - Infrastructure for SDG in ERN-EuroBloodNet

María del Mar Mañú Pereira - ERN-EuroBloodNet Scientific Coordinator

8th May 2026

#1

ERN's perspective: the unmet need in RDs: +7000 RDs, 30Mi PLWRD in Europe



<15% of paediatric rare-disease patients can account for up to 50% of hospital care processes and costs

ERN's perspective: main challenges

Rare disease data across the European Union remains fragmented across multiple registries and national systems, limiting its effective use for research, regulatory decision-making, and health system planning.

1 Data repositories and legal framework

- RDs health data are scattered across registries, biobanks, hospitals, research centres and national databases.
- Divergent legal, ethical, and governance frameworks between Member States.

2 Limited interoperability and standards adoption

- Uneven implementation of FAIR principles.
- Limited implementation of RD codification (e.g. Orphacodes), inconsistent adoption of standards, ontologies and common data models (e.g. OMOP)

3 Limited translation from data to clinical value

- Limited quality metrics and assessment of coverage, bias and representativeness
- Limited validation and uptake of AI and data driven tools.

4 Limited integration of RWE

- Lack of systematic approach to integrate real-world data:
- Research Data: x-OMICs, functional data
- Patient-reported outcomes (PROs)

5 Translation into Regulatory

- Registries based RWE is not incorporated into drug regulatory frameworks (PPPs, EMA, HTA).
- Inconsistent inclusion of patient advocacy groups in design, governance and validation of data and RWE

Synthetic data generation in Rare Hematological Diseases

- **Synthetic cohorts for better stratification**

Create synthetic longitudinal datasets combining EHR, genomic, and omics data to improve AML and SCD sub groups classification, identify rare patient subgroups, and enable AI-driven phenotype–genotype matching.

- **Synthetic data for biomarker and endpoint discovery**

Integrate synthetic clinical and omics data to identify prognostic biomarkers, predict disease progression, and develop better clinical endpoints for precision hematology.

- **Synthetic external control arms for faster clinical trials**

Use synthetic patient populations generated based on real-world data from registries as external control arms to optimize recruitment, simulate trials, and accelerate evaluation of new AML and SCD therapies.

- **Synthetic data augmentation to improve quality and reduce bias**

Apply generative AI to address missing or imbalanced data, improve dataset representativeness, and support patient-centered and regulatory-grade real-world evidence analyses.

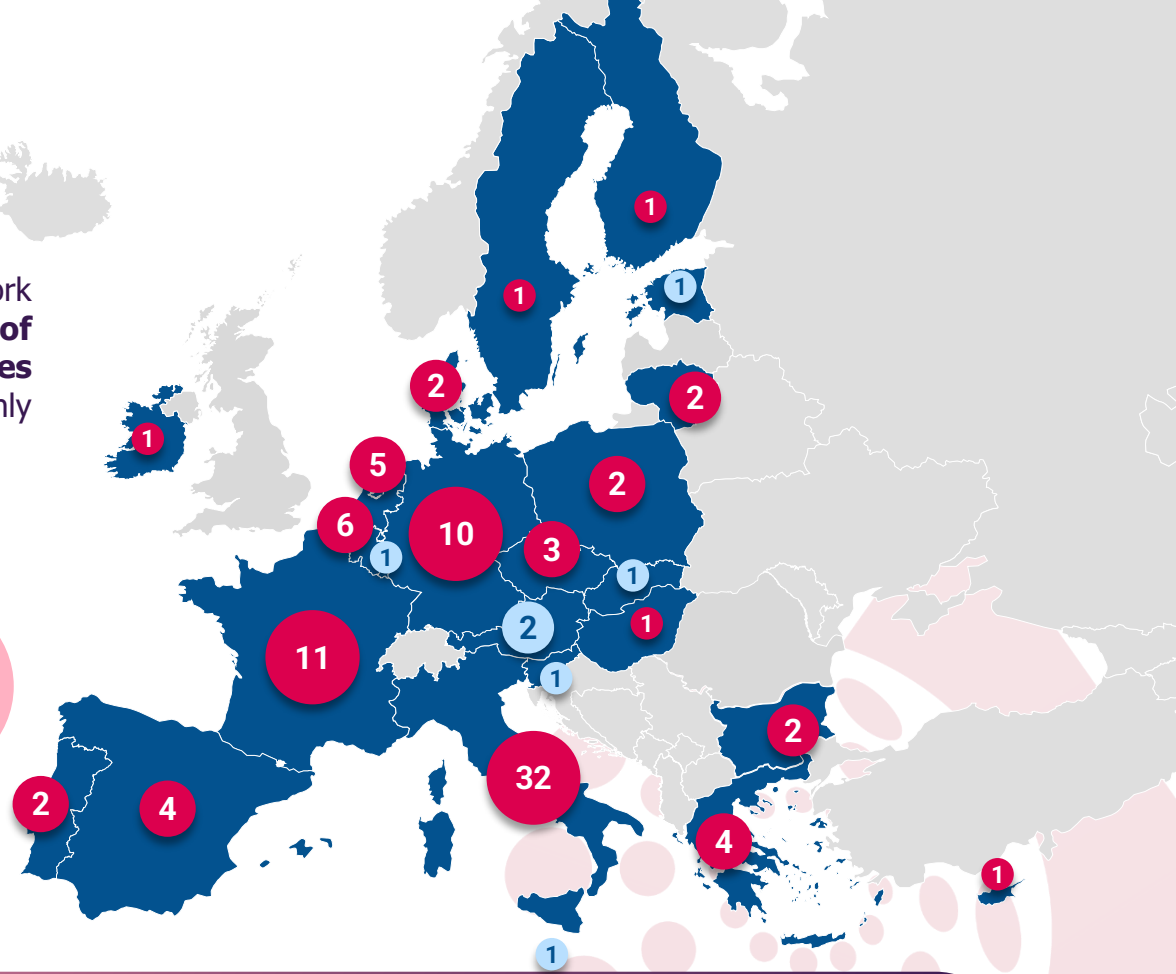
ERN-EuroBloodNet

ERN-EuroBloodNet is a collaborative network aiming to **improve the healthcare services of complex or rare hematological diseases** (RHD) and conditions that require highly specialized treatment in Europe.

89
Members in 18 MS

7
Affiliated Partners in 6 MS

10
ePAGs



Members

Affiliated Partners

ePAG: European Patient Advocacy Group

ebn ERN | EuroBloodNet SYNTHEMA

European Rare Blood Disorders Platform – ENROL Registry

Endorsed by the **European Hematology Association (EHA), ENROL is the central registry of the ERN-EuroBloodNet, covering** both new and already existing registries on rare hematological disorders (RHD).

Aims to **avoid fragmentation** of data by promoting the standards for patients registries' interoperability in line with the **EU-RD-Platform.**

ENROL's principle is to **maximize public benefit from data on RHD** with the only restriction needed to guarantee patient rights and confidentiality, in agreement with EU regulations for cross-border sharing of personal data.

- **Identify unmet needs in RHDs**
- **Share and pool data, reach critical numbers**
- **Analyse KPIs, perform clinical trials & research projects**
- **Knowledge and Evidence generation**
- **better healthcare for RHDs patients**

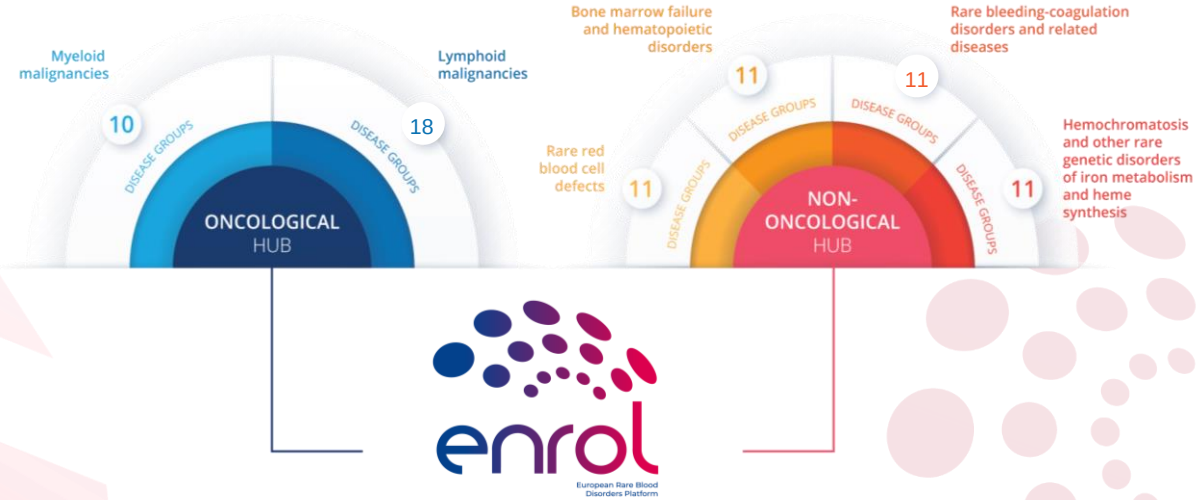
European Rare Blood Disorders Platform - ENROL

Facilitate epidemiological surveillance

Enhance health planning

Enable the identification of patients' cohorts

Promote research & innovative therapies

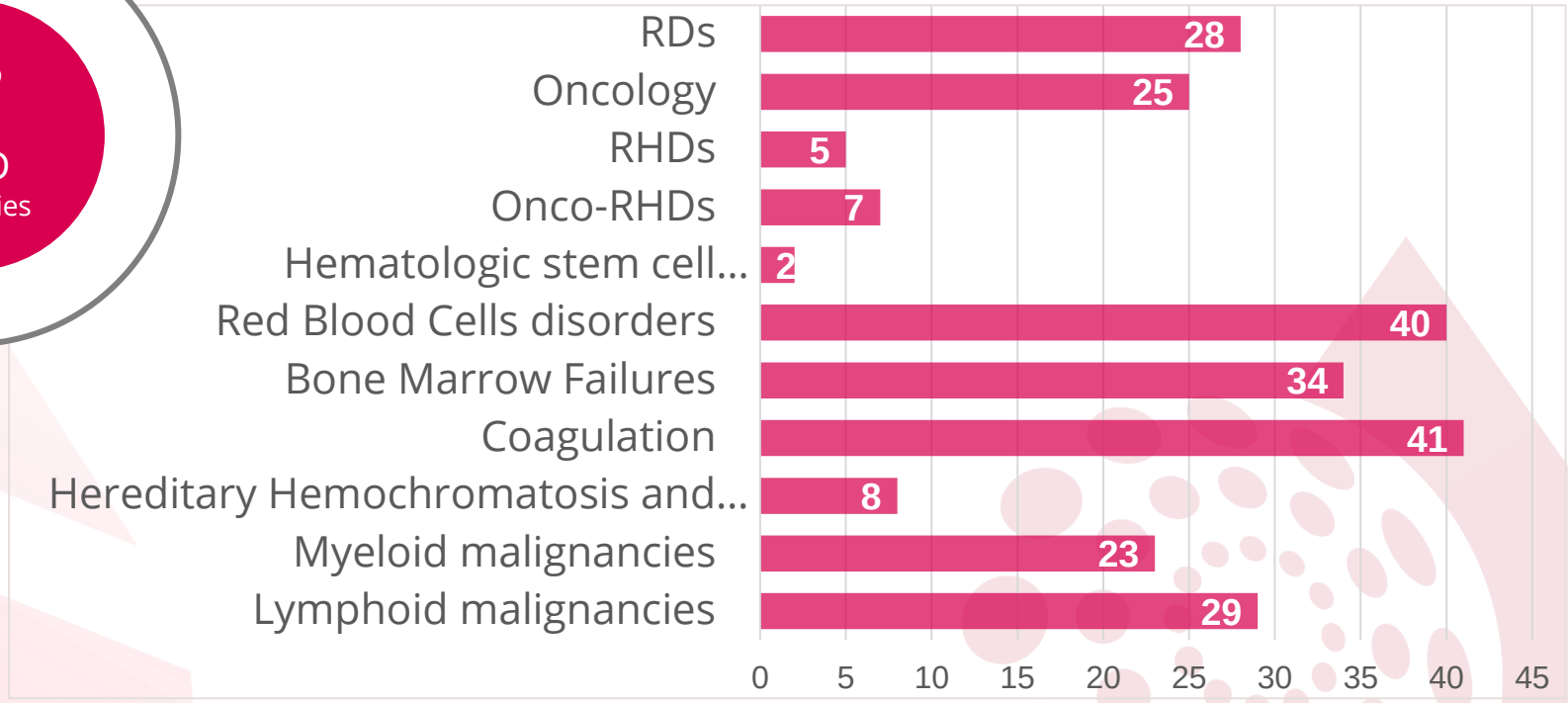


ern | EuroBloodNet

SYNTHEMA

ENROL Directory of Registries in RHD

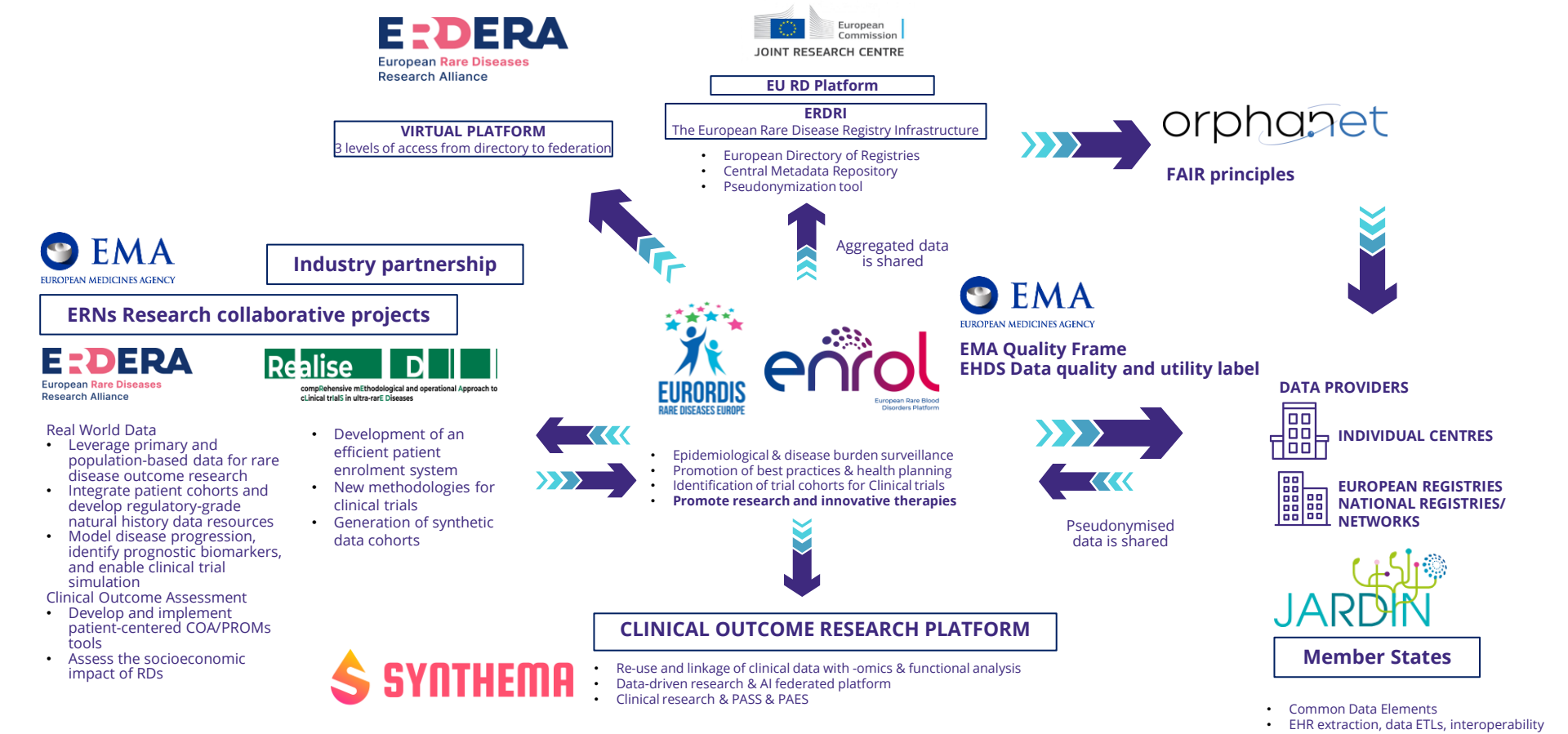
235
EU
RHD
Registries



ERN | EuroBloodNet

SYNTHEMA

ERN's perspective: opportunities to accelarate research through RWE



Synthema Use cases & objectives

Sickle cell disease (SCD)

The disease

SCD is an inherited chronic disorder caused by the presence of abnormal adult haemoglobin.

Red blood cells become rigid and sickle-shaped, breaking down or blocking normal blood circulation, resulting in acute and chronic pain and progressive organ-specific clinical complications.

Kidney disease

Cohort

1,000+ patients with genetic diagnosis for SCD disorder, including paediatric (1+ year-olds) and adult patients.

Acute myeloid leukaemia (AML)

The disease

AML is a type of blood cancer that starts in immature myeloid cells (blasts) in the bone marrow.

These abnormal cells grow and divide too quickly, interfering with normal blood cell production.

Overall survival analysis

Cohort

2,500+ patients with "de novo" AML (2016 WHO classification criteria).

Synthema Clinical partners

SCD

Vall d'Hebron
Institut de Recerca
VHIR

UMC Utrecht

Università degli Studi di Padova

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

AML

HUMANITAS
RESEARCH HOSPITAL

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

HOSPITAL DA LUZ LEARNING HEALTH
TRAINING, RESEARCH & INNOVATION CENTER

UMC Utrecht

Università degli Studi di Padova

AML - Strategy for data collection

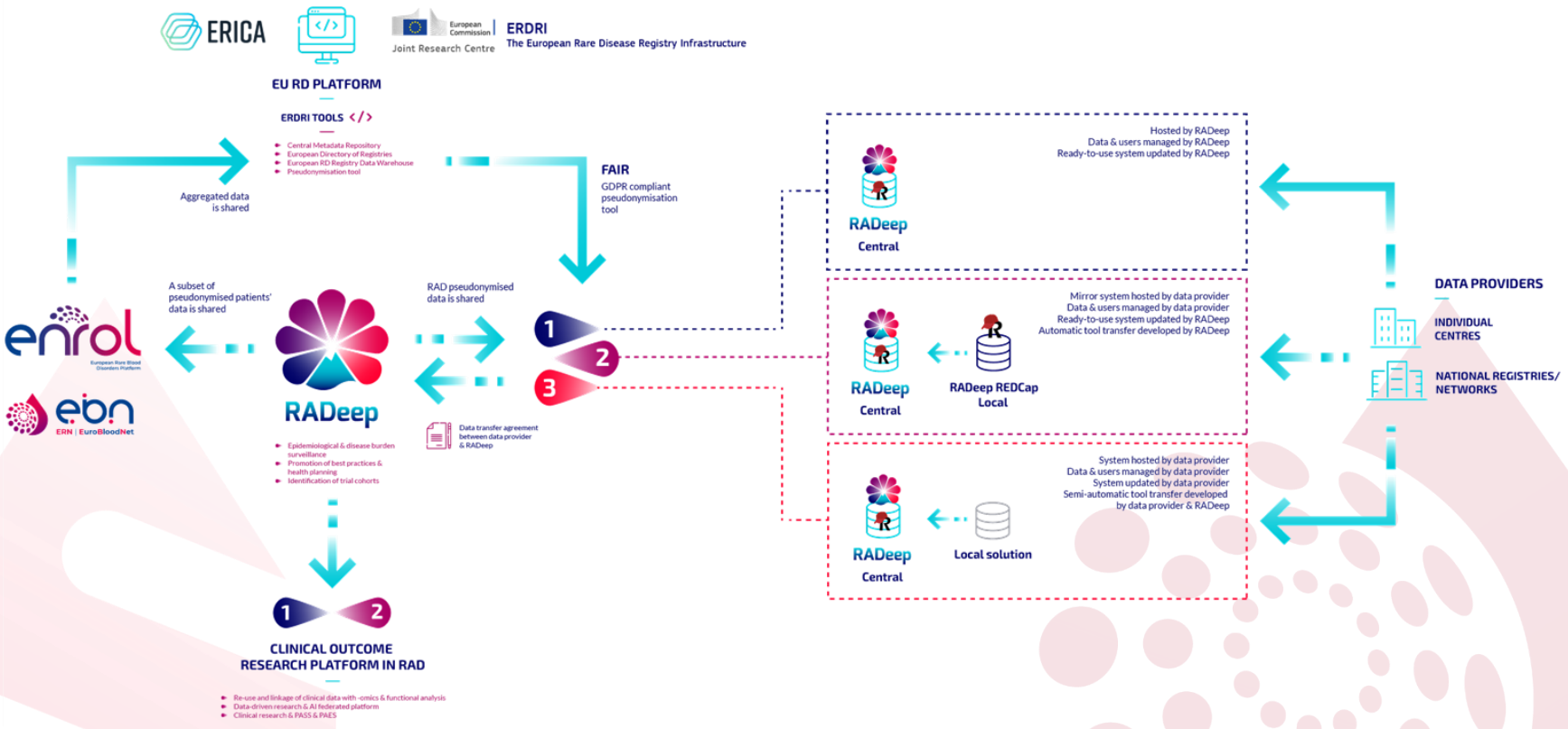
1. Utilizing a public AML dataset comprising clinical, genomic and treatment data (CHA)

1. Release of dataset including 1,500 patients from CHA to the consortium
2. Data cleaning and curation including treatment information (ICH)
3. Clinical interpretation of the variables for technical teams

2. Retreiving AML data from clinical partners (ICH, UMCU, UNIPD, Hospital Da Luz, CHA)

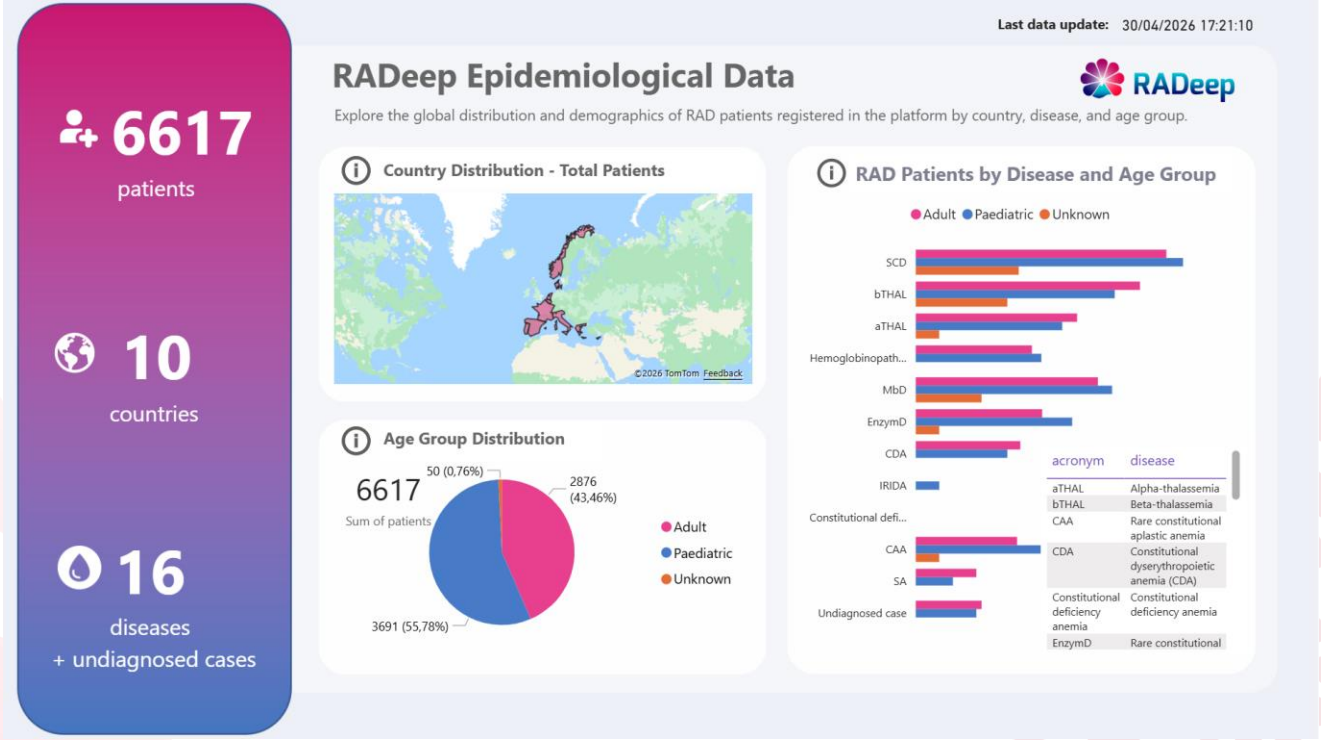
1. The minimum dataset has been implemented into ENROL platform (ERN EuroBloodNet registry)
2. Clinical meetings to define the dataset to be filled by each partner with local data
3. Images availability (types of stainings and format)
4. OMOP data mapping

SCD - Strategy for data collection: RADeep



Data quality management

Public dashboard: Epidemiological data | Radeep



ERN | EuroBloodNet

SYNTHEMA

European Medicines Agency (EMA) Data Quality Framework (DQF) for RWE generation

Four maturity levels describing the natural evolution of how an organization manages, controls, and improves the quality of its data:

Level 1 — Documented

- This constitutes the **minimum baseline**. Organizations collect essential documentation, including data dictionaries, operational processes, variable definitions, data capture flows, and data transformations.

Level 2 — Formalized / Standardized

- Standardized procedures
- Harmonized validation rules
- Common data models
- Clearly defined roles

This level creates **structural consistency** within the system, enabling reproducibility and comparability across centers or countries.

Level 3 — Automated

Once processes are standardized, they can be **automated**:

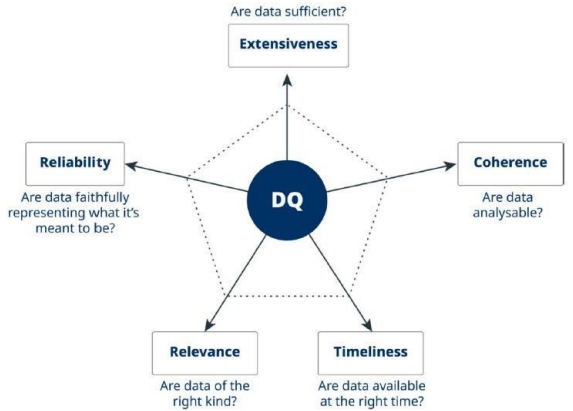
- Automated data validations
- Real-time detection of inconsistencies
- Automatic updating of dashboards
- Replicable ETL workflows
- Automatically generated quality alerts

Automation is **only possible** if the previous steps exist and are robust. This level enables scalability, ensures **continuous quality, and reduces human-driven variability**.

Level 4 — Continuous Feedback

- Continuous improvement systems
- Intelligent alerts
- System learning (identifying recurring errors and points of failure across centers)
- Cyclical quality review
- Indicator-based governance
- Recurring audit mechanisms

While Level 3 automates, Level **4 learns, evaluates, and corrects**. It builds on the automated ecosystem of Level 3 and closes the data quality loop.



EHDS Article 78: data quality and utility label

1. Datasets made available through health data access bodies may obtain from health data holders a Union label relating to data quality and utility.
2. Datasets containing **electronic health data collected and processed with the support of national or Union public funding shall have a data quality and utility label** covering the elements referred to in paragraph 3.
3. The data quality and utility label shall cover the following elements, where applicable:

a) **for data documentation:** metadata, supporting documentation, the data dictionary, the format and standards used, the source of the data and, where applicable, the data model;

b) **for technical data quality assessment:** completeness, uniqueness, accuracy, validity, timeliness and consistency of the data;

c) **for data quality management processes:** the level of maturity of data quality management processes, including review and audit processes and bias assessment;

d) **for coverage assessment**: the time period, population coverage and, where applicable, the representativeness of the population included in the sample, and the average time frame during which a natural person appears in a dataset;

e) **for access and provision information:** the time elapsed between the collection of electronic health data and their inclusion in the dataset, and the time limit for providing the electronic health data following the issuance of a data permit or the approval of a data access request;

f) **for information on data modifications:** the combination and integration of data into an existing dataset, including links with other datasets.

RADeep - RADiANT

ERN- EuroBloodNet
Rare Anemia Disorders registry

GDPR / EHDS

Federated platform registry

GDPR / EHDS / Drug and MD Regulations / AI Act

Federated platform: 4 nodes in 4 countries

11 Ongoing collaboration agreements involving
180 (117 active) HCPs in 10 EU countries:
12 Member States

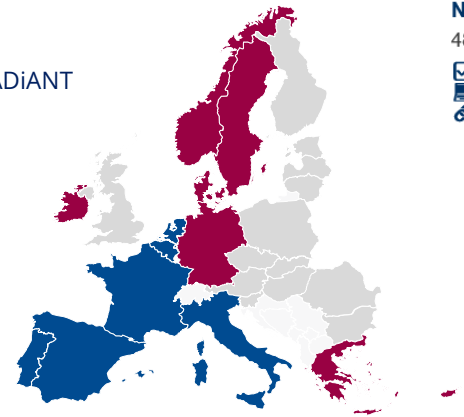
- Belgium
- Cyprus
- Denmark
- France (2)
- Greece
- Italy
- Portugal
- Spain
- The Netherlands
- + Norway

4 in negotiation
- *France*
- *Ireland*
- *Sweden*
- *Switzerland*

Countries onboard in RADiANT
4 EU countries

- France
- Italy
- Spain
- The Netherlands

2 in negotiation
- *Belgium*
- *Portugal*



1,289 enrolled Sickle Cell Disease patients across Europe

Netherlands (SCORE, 4 sites)
487 patients (45% PED)
Deployment
Virtual Machine
Connected to Central Node

Spain (INTEGRA, 10 sites)
334 patients (76% PED)
Deployment
Virtual Machine
Connected to Central Node

RADiANT
Central Node
Federated AI
Data Harmonisation
Secure Computation

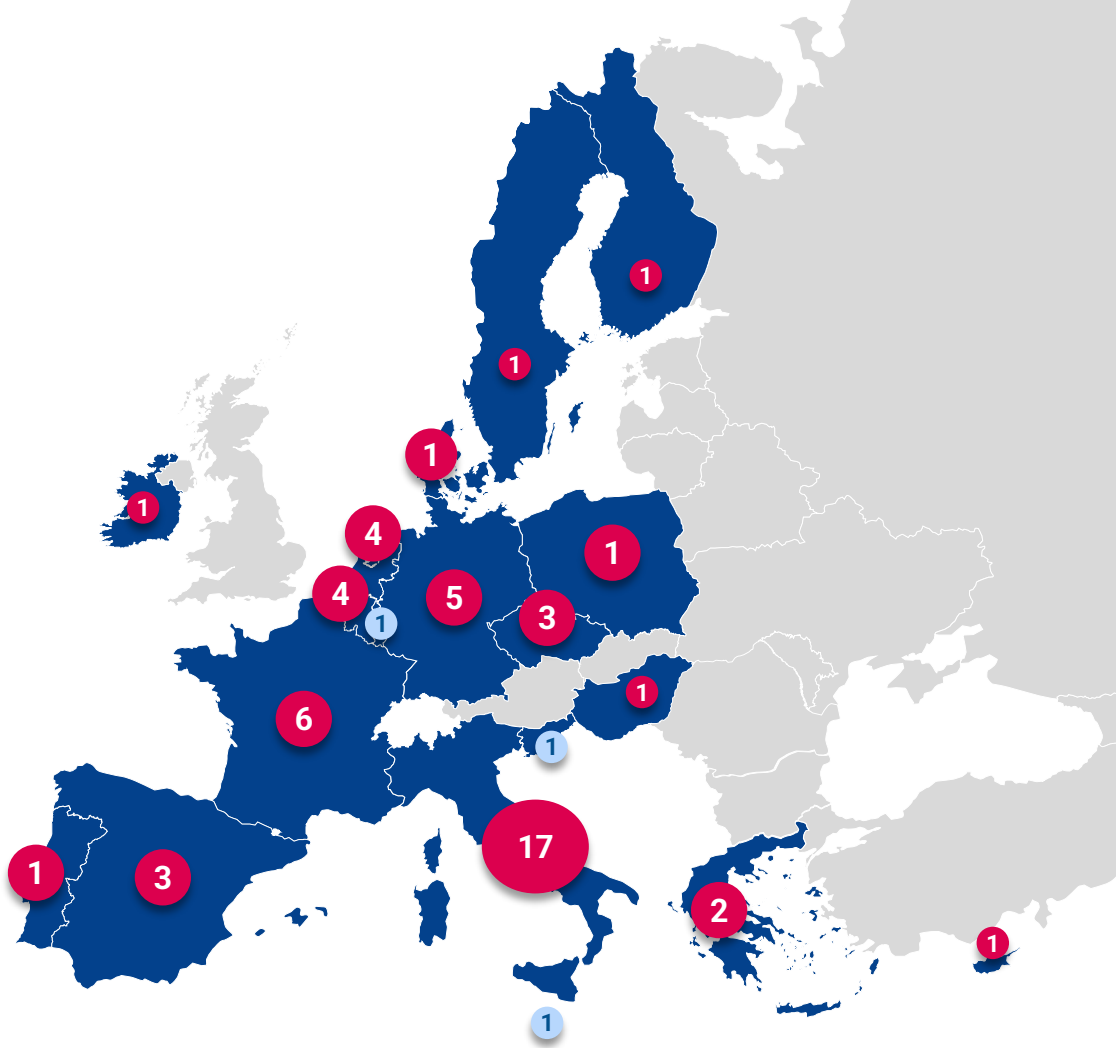
Italy (1 site)
130 patients (75% PED)
Deployment
Virtual Machine
Connected to Central Node

France (2 sites)
494 patients (81.6% PED)
Deployment
Virtual Machine
Connected to Central Node

ERN | EuroBloodNet
SYNTHEMA

ERN | EuroBloodNet

RBC subnetwork



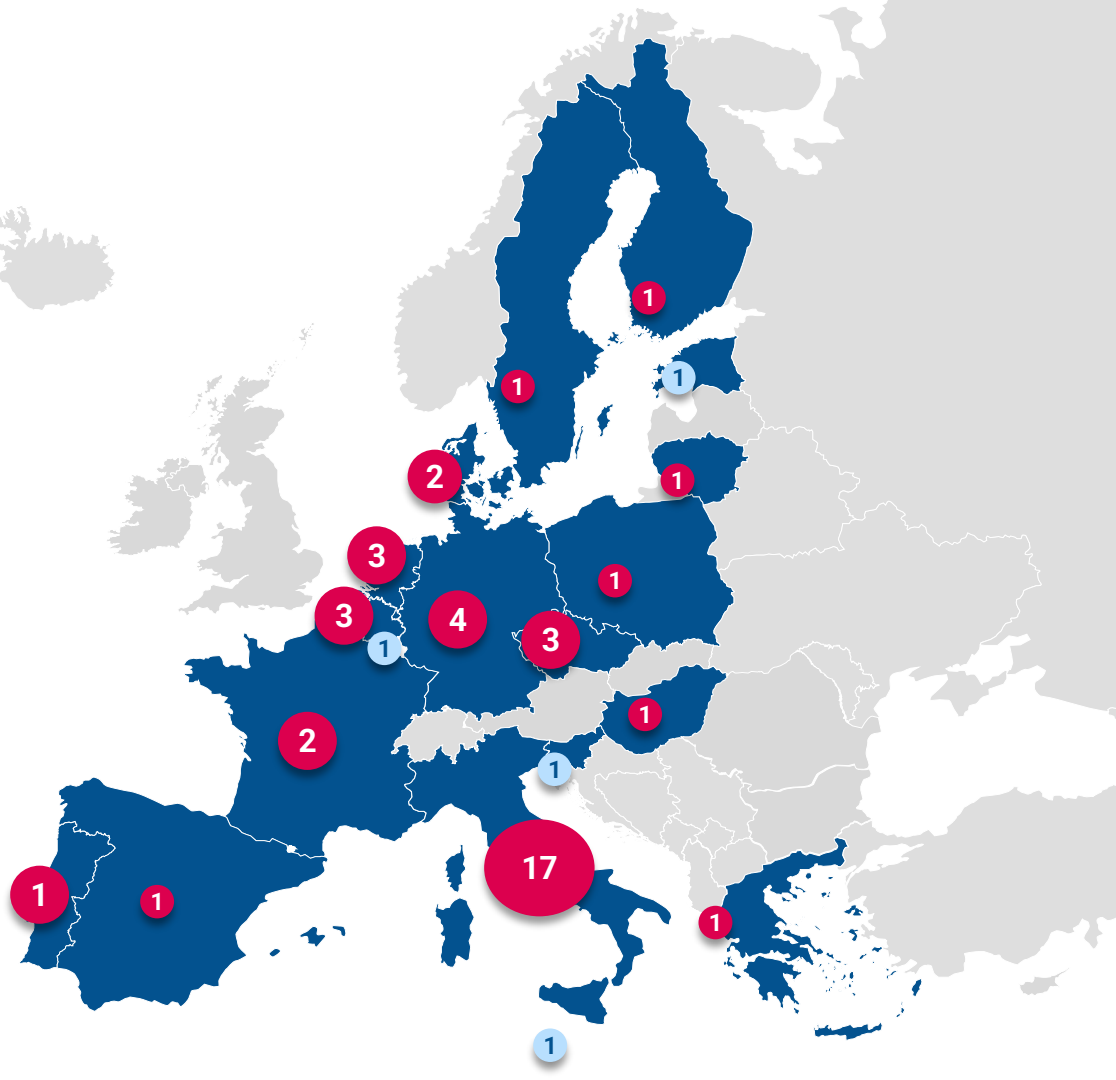
Total Members & APs: 55

Members (52)

Affiliated Partners (3)

ERN | EuroBloodNet

Myeloid subnetwork



Total Members & APs: 46

Members (42)

Affiliated Partners (4)

Next webinars

- **Session 2: Regulatory on Medical Devices, AI Act, Privacy-Preserving and Ethical Considerations**
  Date: 15/05/2026 12:30
  Speakers: Nathan Lea, Rudolf Mayer& Sofia Tsekeridou.

- **Session 3: Data Quality Standardisation, synthetic data generation, anonymization and interoperability**
  Date: 22/05/2026 12:30
  Speakers: Sara Reidel, Giulio Spinozzi, Sofia Tsekeridou, Imanol Isasa, Rafael Redondo, Inês Sousa, Jan Ramon & Max Salmi.

- **Session 4: Opportunities of Synthetic Data Generation in Rare Hematological Diseases**
  Date: 29/05/2026 12:30
  Speakers: María del Mar Mañú Pereira, Eleonora Iascone & Raffaella Colombatti.

Thanks!

Any questions?

Keep in touch!

eurobloodnet.eu /ERNEuroBloodNet @ERNEuroBloodNet @erneurobloodnet.bsky.social

synthema.eu /synthema @SYNTHEMA_EU @synthemaeu.bsky.social

Funded by
the European Union

Acknowledgements

European Reference Network
for rare or low prevalence complex diseases
Network Hematological Diseases (ERN EuroBloodNet)

Funded by the European Union

This project is supported by the European Reference Network on Rare Haematological Diseases (ERN-EuroBloodNet)-Project ID No 101085717. ERN-EuroBloodNet is partly co-funded by the European Union within the framework of the Fourth EU Health Programme.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

SYNTHEMA

Funded by the European Union

SYNTHEMA is an initiative funded by the European Union's Horizon Europe Research and Innovation programme under grant agreement No. 101095530.