



# SYNTHEMA

DECEMBER 1<sup>ST</sup> 2022

**PRESS RELEASE**

## **A cross-border hub to develop and validate Artificial Intelligence techniques for anonymization and synthetic data generation in rare hematological diseases**

The consortium of **SYNTHEMA** (Synthetic generation of hematological data over federated computing frameworks) is pleased to announce the launch of an ambitious joint initiative selected and granted almost €7M by the European Commission as part of the Horizon Europe Programme.

SYNTHEMA aims to establish a privacy-preserving, cross-border hub to develop and validate innovative Artificial Intelligence (AI) models for clinical data anonymisation and synthetic data generation in rare hematological diseases.

Haematological diseases are a large group of disorders resulting from abnormalities in blood cells, lymphoid organs and coagulation factors. They are generally split into two categories: oncological –haematological malignancies, i.e. lymphomas, myelomas, leukaemias–, and non-oncological –i.e., hemoglobinopathies, haemolytic anaemias, coagulopathies. Over 70% of hematological diseases are considered rare, and despite the existence of several collaborative research groups at national and European level, current clinical approaches are often ineffective due to the relatively low number of patients and the prevalence of data silos in unconnected clinical sites and registries.

Additionally, we find that rare diseases are not actually rare when seen under a global lens. Their impact might seem small on paper but in reality, the numbers are staggering: roughly 30 million people are living with a rare disease in the EU, and conservative estimations speak

of a total of 300 million worldwide. This is exactly why precision medicine is key: when we shift our focus to the individual, all diseases become unique.

Rare hematological diseases inherently suffer from data scarcity and fragmentation, but there is also the dilemma of data privacy and protection. Can we talk about full anonymity when the risk of re-identification is so high? How can we generate meaningful synthetic data to circumvent the lack of quality data to train AI models on?

The overarching ambition of SYNTHEMA is to increase the number of existing samples in this disease space (with a focus on two highly representative use cases: Sickle Cell Disease (SCD) and Acute Myeloid Leukemia (AML)), thus fighting off the critical issues of data scarcity and fragmentation and pushing the boundaries of patient-centric, GDPR-compliant research.

Ultimately, SYNTHEMA intends to generate reliable, high-quality synthetic data that can shape new “virtual patients” to further enhance diagnostic capacity, assess treatment options and predict outcomes in rare hematological diseases. To achieve this goal, SYNTHEMA will develop a novel Federated Learning infrastructure, equipped with secure multiparty computation and differential privacy protocols to effectively connect clinical sites with computing centres, academia and SMEs across Europe.

SYNTHEMA enjoys the support, resources and active participation of [ERN-EuroBloodNet](#), as the [European Reference Network on rare haematological diseases \(RHDs\)](#) concentrating 103 highly specialised multidisciplinary healthcare teams in 24 Member States. Moreover, the [European Rare Blood Disorders Platform \(ENROL\)](#), conceived in the core of [ERN-EuroBloodNet](#) in line with the EC strategy for Rare Diseases as an umbrella for new and existing RHDs registries, directly contributes to SYNTHEMA on the promotion of interoperability standards of the EU RD platform to tackle the scarcity and fragmentation of data and widen the basis for GDPR-compliant research in RHDs. All in all, [ERN-EuroBloodNet](#) and [ENROL](#) constitute the perfect environment for SYNTHEMA and the creation of a cross-border health data hub for RHDs, as the space to develop and validate innovative AI-based techniques for clinical data anonymisation and synthetic data generation.

For the next 4 years, 16 partners from 10 countries (Spain, Italy, Austria, United Kingdom, Belgium, Netherlands, France, Germany, Portugal and Luxembourg) will join forces to create standardised, interoperable and multimodal pipelines and datasets that can be validated for their clinical value, statistical utility and residual privacy risks.

At SYNTHEMA, we are devoted to expanding the landscape of personalized medicine in rare hematological diseases: a new paradigm built on robust ethics to bring community members together through explainable, trustworthy AI.

## Key contacts

- **Project Coordinator:** Federico Álvarez (UPM) – [federico.alvarez@upm.es](mailto:federico.alvarez@upm.es)
- **Management Office:** Gustavo Hernández (UPM, Deputy Coordinator) – [gustavo.hernandez.penalosa@upm.es](mailto:gustavo.hernandez.penalosa@upm.es) and Anna Rizzo (Project Manager, Datawizard) – [a.rizzo@datawizard.it](mailto:a.rizzo@datawizard.it)
- **Communication:** Arantxa Echarte (AUSTRALO) – [arantxa@australo.org](mailto:arantxa@australo.org)

## FOLLOW US!



@synthema\_eu



/synthema