

Algorithms to transform Knowledge into better Medicine Outcomes

The HARMONY Big Data Platform: What is it and how does it work?

Despite data being a key driver for modern evidence-based medicine, we have been struggling with the same challenge for years: how can we have access to the scientific data that has been kept isolated and convert it into practical applications in order to make real progress in human sciences? The HARMONY Big Data Platform is a significant step towards breaking down these silos and exploiting the potential of Big Data to improve the treatments and quality of care for patients with hematologic malignancies.

From volume to value

53 Partners and 32 Associated Members, totaling 80 public-private organizations from 22 European countries, are working together within the HARMONY Alliance – the European Network of Excellence for Big Data in Hematology, establishing a data platform to unlock knowledge about blood cancers, known as Hematologic Malignancies (HMs). Machine learning techniques and advanced algorithms will be used to transform the data into valuable evidence-based outcomes to improve patient care and their quality of life.

To make progress in the study of hematology there is an urgent need to have access to a massive amount of high-quality, deep data. The HARMONY Big Data Platform is the key to enabling this: by collecting and engaging multiple individual data sources featuring high-quality data, we will allow the performance of meaningful analysis. This will result in the establishment of best practices and generate findings that will help address challenges such as improving the diagnostics and treatment of patients with blood cancer.

Researchers can focus on finding answers to four questions: how can we diagnose patients faster and with greater precision; what are the best practices that could help doctors in making better treatment decisions; how can we tackle the unmet needs of patients; and what is the best way to make progress in new drug development?



The HARMONY Big Data Platform: What is it and how does it work?

Michel Van Speybroeck, JPNV-Janssen, HARMONY Partner and leader HARMONY Work

Package 4: Data Platform: *"We already have an operational platform, which means that the data management infrastructure has been established," says. "We've been able to go through the entire flow of data uploaded to our honest broker – who manages the final steps in the anonymization process – to transfer the data to the HARMONY big data platform, after which the data is transformed into our common data model structure."*

How will this benefit the patient? HARMONY is oriented toward blood diseases, 7 indications within hematology. However, these 7 Hematological Malignancies break down into a plethora of sub-variations, where patients from one sub-variation may sometimes respond very differently than patients in another cluster.

Michel Van Speybroeck : *"The research that we're performing is the basis we need to create better-targeted therapies for the different subgroups so that there is not only an improvement in the 'average' response but also in the actual outcomes for the different subgroups".*

Interoperable and complete data for reliable results

The HARMONY Big Data Platform hosts multiple types of data, including clinical data collected from symptom diagnoses, biochemistry and physical examinations, and other information gathered on treatment, survival, omics, quality of life, and resource utilization. To ensure that the descriptive, comparative, and predictive information generated by the analyses performed on the data platform is reliable, one of our priorities is to determine the quality of the data that will enter our Big Data Platform. Thus, before algorithms are used to mine the available data, the input information must be checked precisely, to ensure it is standardized, anonymized, complete, and correct.

The data transferred to HARMONY is received in the same format as the source, which implies disparate formats, organizations, and coding. Therefore, a deep understanding of the information provided is paramount in order to guarantee data completeness and harmonization across the platform.

First, the completeness is evaluated prior to intake, which guarantees that only high-quality data enters the platform. Another essential step is to convert all the data to a common data model. This part of the project has the highest priority, as it determines the usability and value of the output data. It does not affect the meaning or the clinical value of the data, but it does allow information that was initially incomparable and not interoperable to be processed in a standardized way.



The HARMONY Big Data Platform: What is it and how does it work?

Ruben Villoria Medina, GMV, HARMONY Partner and leader HARMONY Work Package 4: Data Platform: *"Once in the platform, the process of standardization and harmonization in HARMONY is achieved by converting the data to the Observational Medical Outcomes Partnership (OMOP) common data model, which uses international standard terminologies such as SNOMED or LOINC to refer to the information contained in the source. Eventually, all sources become structurally and semantically homogeneous."*

Data discovery, business intelligence, and analytic tools have all been deployed on the platform so that HARMONY data can be input into cutting-edge statistical algorithms. Zeppelin is a tool designated to allow web-based, data-driven, collaborative research in different programming languages. Although it also permits data visualization, this feature will be enhanced by the development of interactive workbooks and dashboards on Tableau, which will eventually help the users to uncover insights into disease and visualize the patterns hidden within the data pool.

Data security first

Although all the information is stored in a single database, only subsets of the complete data lake are available for analysis. The fractions of data retrieved correspond with the essential data to be used as input for the analysis designed for a specific research question. Access to the subsets will be granted to a limited group of users during a specific time span. Besides this, the fact that only anonymous data enters the platform and that it is ISO27001 certified adds further reassurance concerning data privacy and security.

www.harmony-alliance.eu

[Follow our LinkedIn Page](#) | [Follow us on Twitter](#) | [Subscribe to our Youtube channel](#)

About the HARMONY Alliance: a public-private European Network of Excellence for Big Data in Hematology, established in January 2017. Our mission is to unlock and spread valuable knowledge on hematologic malignancies (blood cancers) among a large number of stakeholders, with the goal to harness and mine Big Data to speed up the development of improved treatments for patients and more effective treatment strategies. HARMONY currently has 53 Partners and 32 Associated Members from 22 countries. HARMONY funded through the Innovative Medicines Initiative (IMI) (LINK), Europe's largest public-private initiative aiming to speed up the development of better and safer medicines for patients. HARMONY has received funding from IMI 2 Joint Undertaking and is listed under grant agreement No. 116026. This Joint Undertaking receives support from the European Union's Horizon 2020 Research and Innovation Programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA) (LINK). IMI supports collaborative research projects and builds networks of industrial and academic experts in order to boost pharmaceutical innovation in Europe.

