

# KAMPAL

## Pilot Description

Private section:

- [Kampal \(Private\)](#)

|  |   |
|--|---|
| <b>Title</b>                           | <b>Kampal: Artificial Intelligence for rare disease diagnosis</b>   |
| <b>Subtitle</b>                        | Assessing the probability of development of further diseases in Gaucher disease patients  |
| <b>Description</b>                     | <a href="#">The Spanish Foundation for the Study and Treatment of Gaucher Disease and other Lysosomal Diseases</a> (FEETEG) promotes the scientific research of Gaucher disease and its treatment methods. The Foundation is interested in predicting the probability of development of diseases such as neoplasms or Parkinson's disease in patients of Gaucher disease (correlations between diseases). For this purpose, <a href="#">Kampal Data Solutions</a> was contacted by FEETEG to develop an advanced analytical model based on Artificial Intelligence with the information available in the Gaucher Spanish Disease Registry.  |
| <b>Challenge</b>                       | <p>Due to the fact that Gaucher disease is a rare disease with few national registries, the computational power of a local computer for the study of correlations with other diseases was enough to analyse the data collected.</p> <p>The challenge now is to generate a new model able to <b>predict if a person has the probability of developing</b> Gaucher disease. In this case, the AI model must include not only data from current Gaucher disease patients but also data from healthy patients. Opening our sample universe also to healthy patients exponentially increases the sample size (from hundreds to millions) and potentially the model's complexity. This implies the need of advanced computational resources such as the cloud platform provided by EOSC.</p> <p>Although this proof of concept is focused in Gaucher disease, the developed solution could be adapted in the future to other diseases data bases. The obtained general-purpose solution will be exploited by Kampal Data Solutions in the mid-term.</p> |
| <b>Work Plan</b>                       | <p>In the context of the EOSC-hub project, Kampal Data Solutions will develop the following tasks:</p> <ul style="list-style-type: none"><li>• Import the healthy and ill patients' registries to a data base on EOSC infrastructure.</li><li>• Statistically analyse the data and develop a classifying model based on machine learning techniques.</li><li>• Optimize the machine learning algorithm for a cloud based environment.</li><li>• Validate the model performance and produce plots /charts of diverse KPIs.</li></ul>   |
| <b>Achievements</b>                    | Kampal Data Solutions has developed a machine learning model able to cope with big data samples by using the cloud infrastructure provided by EOSC DIH. The case study was based on a medical data set provided by the FEETEG containing information of patients with Gaucher's disease. In addition, extra data was generated following what the current literature considers normal values. This way allowed to obtain a big data sample that loosely resembles the natural proportion of patients with Gaucher Disease. To be able to handle the problem size increment the parallelization of the code was required, benefiting from cloud computing.   |
| <b>How they used EOSC-hub services</b> | The pilot required extra computational resources to cope with the problem size (1 million samples). For that, Kampal Data Solutions got benefit from the EOSC DIH cloud infrastructure where 16 VCPUs with 32GB of RAM were used. To speed up the process and benefit from all the cores, the code was parallelized. This way, different operations can be done simultaneously on each core using only a fraction on the sequential computational time. The parallelization of the code was greatly simplified by using the R packages parallel and dplyr.  |
| <b>The value proposal of the pilot</b> | Although the obtained results do not have medical value, this proof of concept shows that the chosen model is scalable and could be efficiently applied to other conditions or illnesses where more data is available. The challenge now will be identify the business opportunities to exploit the model.  |
| <b>How EOSC-hub helped</b>             | <p>EOSC-hub has provided Kampal Data Solutions with powerful cloud infrastructure to support the scaled up analytics required for validating the proof of concept. Using the computing power of the EOSC-hub services, Kampal Data Solutions could experiment and test its new models for the disease prediction.</p> <p>The technical support provided from the EOSC DIH team helped Kampal to access and manage the Cloud and provided a better understanding of the EOSC computing infrastructure, meanwhile the visibility service enhance the exposure of the pilot through different European communities.</p>  |

|       |  |
|-------|--|
| Logos |  |
|-------|--|