PUBLIC SELECTION BASED ON QUALIFICATIONS AND INTERVIEW FOR THE AWARDING OF NO. 1 GRANT LASTING 12 MONTHS FOR CONDUCTING RESEARCH IN ACCORDANCE WITH ART. 22 OF LAW OF 30.12.2010 NO. 240 AT THE DEPARTMENT MANAGEMENT, INFORMATION AND PRODUCTION ENGINEERING OF THE UNIVERSITY OF BERGAMO (ACADEMIC RECRUITMENT FIELD 09/G1 – SYSTEMS AND CONTROL ENGINEERING) ACADEMIC DISCIPLINE ING-INF/04 – SYSTEMS AND CONTROL ENGINEERING (FUND: DIGIPFUTURMAN)

announced with decree of the Rector Rep. no. 722/2017 of 11.12.2017 and posted on the official registry of the University on 11.12.2017

RESEARCH PROJECT

The genome analysis process consists of several stages. Some require the preparation, manipulation and analysis of physical samples, so they have durations not compressible. Others concern only the treatment of information with numerous different activities (external DB access, local mass storage access, access to the cloud, computing, etc ...) to "heavy" files (a standard complete annotated file, at the end of the process, occupies approximately 20 GB). In addition, the genetic diagnostic process requires access to many DB containing information on diseases associated with genetic mutations. The information retrieved from queries to DBs are then associated with mutations in the genome, generating the so-called "annotated File." Finally, once the diagnosis is obtained, the clinical case is closed by associating the symptoms of the underlying diagnostic question to one mutation for which the diagnosis was made. It can occur that other medical research centers have clinical cases with symptoms similar to cases already seen but with different questions or diagnostic hypotheses. Currently, search automatically and quickly among the already stored clinical cases based on different diagnostic hypotheses is a complex and not standardized process.

The objectives of the research are:

a) Carry out a functional analysis of the various activities of the process in order to identify those that are more time-consuming and identify the technical reasons.

b) To propose a new architecture of "local" information system (only for the Medical Genetics Service) clearly highlighting the benefits in terms of performance and methods of interaction with the "global" hospital system, with particular attention to ensuring the safety and data integrity.

c) To make the annotated file generation fully automatic, for example by developing a script Ruby on Rails / Python trying to automate the query process and download the results from the DB Web site, either by changing the Bash program existing interpreting the results returned by the query and generates the annotated files completed.

d) the design of a new genetic data storage system that allows users to make queries to create new associations between the data contained in the files noted and symptoms. The candidate must first define the database technology best suited to manage the genetic data and then develop implementation. Example of RDBMS technologies: Oracle DB, MySQL, IBM DB2, PostrgreSQL. Example of non-relational technologies: Cassandra, HBase.