Haplotypes Treatment: An Extension to the Conceptual Scheme of Human Genome to develop GeIS

José Fabián Reyes Román

Doctoral Program in Computer Science

Software Engineering, Formal Methods and Information Systems

Dr. Óscar Pastor López

Research Centre on Software Production Methods (PROS)

Universitat Politècnica de València, Spain

INTRODUCTION

Motivation

Our motivation is based on problems presented in the current genomic chaos.

* DISPERSION AND INCONSISTENCY OF DATA: a set of data that do not coincide with each other; which don´t have a common ground for better management.

- Improving the quality and accuracy in determining data: use of statistical models to select, process, transform, interpret and evaluate data to yield reliable and accurate Genetic Diagnostics.

Contribution

This poster presents the application of Conceptual Models to the genomic environment. Conceptual modeling allows us to integrate all knowledge through well-defined and orderly arrangements that enhance the data management.

Goal

Extending the CSHG with the integration of haplotypes and Statistical Models for the creation of a powerful and reliable Genomic Repository for the exploitation of software applications for the Genetic Diagnosis.

OBJECTIVES

General:

The overall objective is to address the dispersion and inconsistency of genomic data. As a solution we propose to extend the Conceptual Scheme of Human Genome with a criteria untreated up to date: determine the data quality of genomic information.
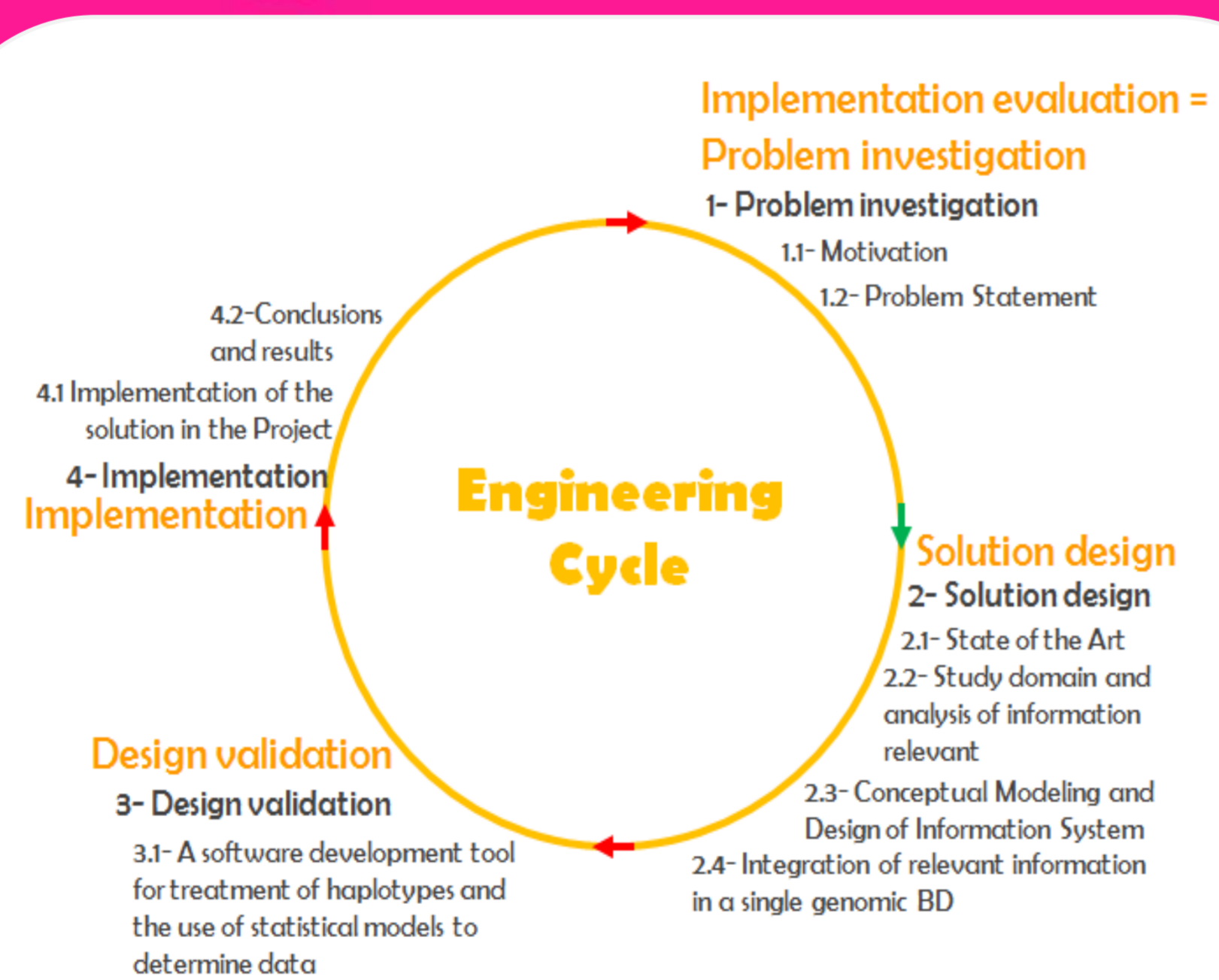
Specifics:

- Integrating Treatment of haplotypes in CSHG
- Application of Statistical Models in CSHG

RESEARCH-PHASES

To develop this research we apply the Engineering Cycle, proposed by Roel Wieringa, since it adapts positively to the work we are doing.

Using this methodology we seek to obtain specific and satisfactory results that bring significant value to the scientific environment.

RESULTS

Phase I. Current status of our genomic repository, from which we start with CSHG this conceptual definition and carry the load of data from different data sources (biomedical) to our HGDB (Human Genome Database).
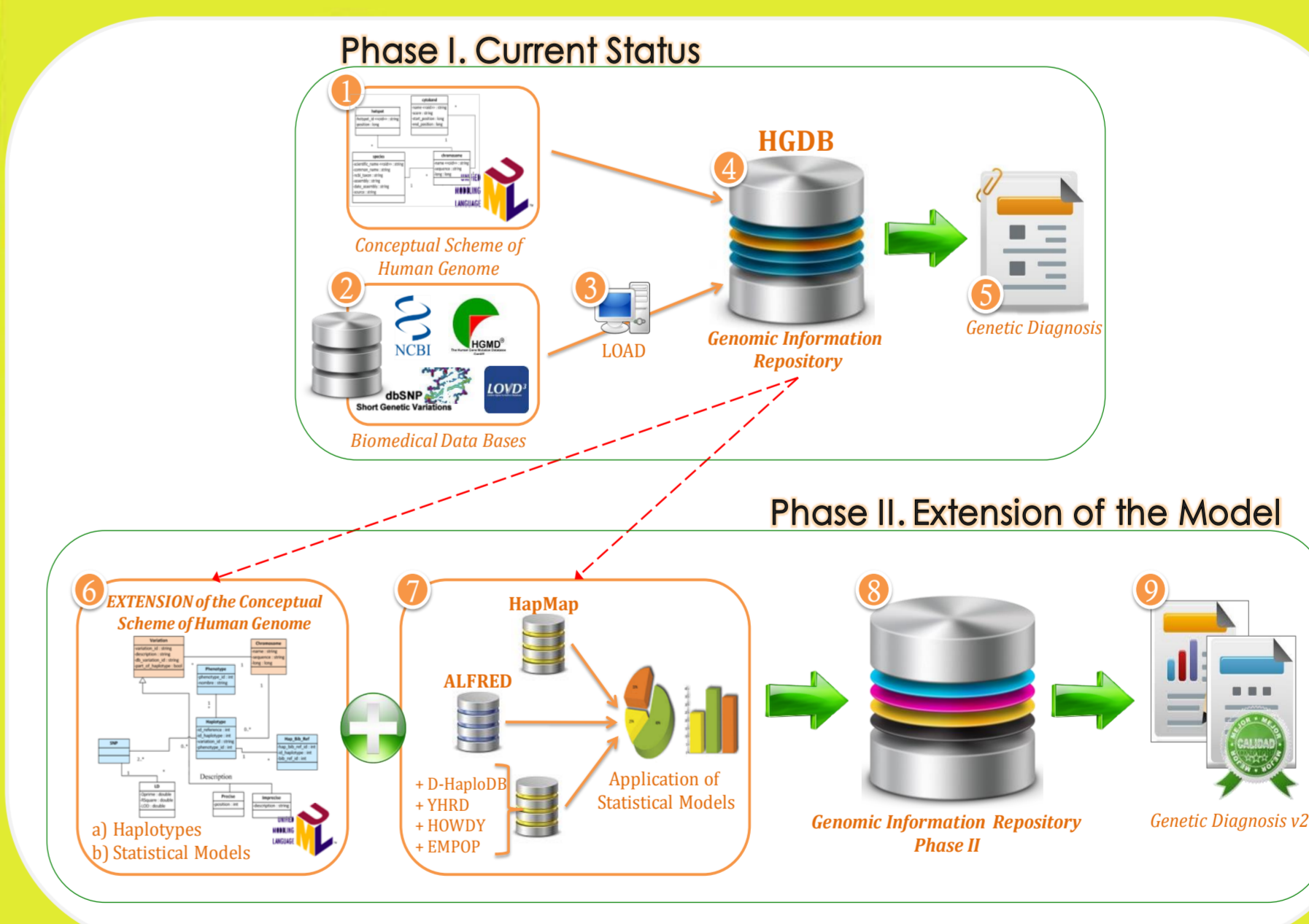
Phase II. Extension of the Model: enrichment with new knowledge, this generate a new version of CSHG, where we integrate the definition of haplotypes and statistical factors. Implementation of ETL process with new data repositories, like eg .: HapMap, ALFRED, YHRD, etc.

Conclusions:

The application of Model Driven Development (MDD) in the bioinformatic field has been useful in the growth of Genomic Information Systems (GeIS), which have evolved to facilitate data management. Large volumes of current genomic data are accomplished deal productively through the advantages and facilities of Genome Information Systems (GeIS).

Email: jreyes@pros.upv.es