Reyes Román, JF.; León-Palacio, A.; García-Simón, A.; Cabrera Beyrouti, R.; Pastor López, O. (2023). Integration of clinical and genomic data to enhance precision medicine: a case of study applied to the retina-macula. Software & Systems Modeling. 22(1):159-174. https://doi.org/10.1007/s10270-022-01039-4

Noname manuscript No.
(will be inserted by the editor)

Integration of Clinical and Genomic Data to Enhance Precision Medicine

A Case-of-Study Applied to the Retina-Macula

José Fabián Reyes Román · Ana León Palacio · Alberto García S. · Rubén Cabrera Beyrouti · Oscar Pastor

Received: date / Accepted: date

Abstract Age-Related Macular Degeneration (AMD) is a complex, multifactorial, and neurodegenerative disease that is the third cause of blindness after cataracts and glaucoma. To date, there are no effective remedies available for treating the disease. Therefore, the main goal of the scientific community is to uncover the underlying role that both genetics and environmental factors play in the development of the disease. Nevertheless, the complexity of the domain, the heterogeneity of the information, and the massive amounts of existing data hinder the daily work of clinical experts to provide an accurate diagnosis and treatment. In this work, we present how clinicians can benefit from the development of ontologically well-grounded Information Systems to support the management of both clinical and genomic data. First, we summarize the results obtained in a previous work that cover the clinical perspective using an information system called G-MAC, that has been specially developed for the management of clinical data. Then, we present the results of an exhaustive

José Fabián Reyes Román
VRAIN Research Institute, Universitat Politècnica de València, Spain
E-mail: jreyes@vrain.upv.es

Ana León Palacio
VRAIN Research Institute, Universitat Politècnica de València, Spain
E-mail: aleon@vrain.upv.es

Alberto García S.
VRAIN Research Institute, Universitat Politècnica de València, Spain
E-mail: algarsi3@vrain.upv.es

Rubeń Cabrera Beyrouti
Ophthalmology Service, Hospital Vega Baja de Orihuela, Spain
E-mail: rubeyrouthy@hotmail.com

Oscar Pastor
VRAIN Research Institute, Universitat Politècnica de València, Spain
E-mail: opastor@vrain.upv.es

study of the genetic factors of Age-Related Macular Degeneration by using an information system that was developed with the aim of enhancing the management of complex genomic data. Finally, we state how the connection of both perspectives through the use of Conceptual Models can benefit clinicians and patients through a more accurate Medicine of Precision.

Keywords Conceptual Modeling · Information Systems · Precision Medicine · Age-Related Macular Degeneration · Genomics

1 Introduction

Age-Related Macular Degeneration (AMD) is a complex, multifactorial, and neurodegenerative disease [23,26] that affects the pigmented area that is located inside the retina (macula) and which plays a key role in the process of vision. AMD has a devastating impact on both individuals and society, being the third cause of blindness after cataracts and glaucoma [24,25].

To date, there are no effective remedies available for treating AMD. Although some drugs can reduce the risk of progression of the disease in some patients (temporarily), there is no statistical evidence regarding their effectiveness (depending on the patient and duration), and they are very costly. Since the main risk factor of AMD is age [22], more and more resources are being designated to this disease as the global population ages. Therefore, the main goal of the scientific community is to uncover the underlying molecular mechanisms and the role of both genetics and environmental factors in order to:

1. identify patients in an early stage, improving the diagnosis,
2. identify patients whose condition is more likely to progress at a faster pace,
3. reduce the high costs of treatment, and
4. find new treatments and increase the effectiveness of the existing ones to slow down the rate of progression. [27,28].

Nevertheless, the huge amount of clinical and genetic data that is being generated is overwhelming [22]. Hence, the professionals that work in the Ophthalmology domain have to confront several issues that are well known to the Information Systems (IS) community. These issues include the inherent complexity of the domain, the heterogeneity of the information to be managed, the dispersion of the data sources, and the management of massive amounts of data. The development of ontologically well-grounded information systems that are specially developed for the management of clinical and genomic data, can help to deal with or at least minimize such problems. It can also reduce the cost of the treatments by optimizing management and resources [22] and can increase the amount of money that can be spent in other medical areas.

One of the main problems when managing genomic and clinical data is the lack of standards to represent the main concepts of the domain, which hinders the data integration and its further analysis. Therefore, in order to join the two

perspectives, we must harmonize the concepts that are used to ensure data interoperability. To such aim, conceptual schemas are the most appropriate tool for obtaining a shared understanding of the domain [6]. Conceptual schemas are graphical representations of ontologies that allow the characterization of the main concepts of a specific domain, as well as the relationships among them in an intuitive, direct, and natural way.

In a previous work, we worked together with domain experts to tackle the problems associated to the management of clinical data by developing an information system called G-MAC, which specializes in the management of such complex data [43]. This IS is based on the Conceptual Schema of the Macular Pathology (CSMP) and helps ophthalmologists to track the evolution of their patients. In this work, we face the genomic perspective by determining the genetic variants that are associated to the development of AMD, using an information system whose main aim is the efficient management of genomic information [47]. This second IS is based on the Conceptual Schema of the Human Genome (CSHG) [14,31]. Both models are interoperable, which helps to connect the two perspectives and provides a complete view of the disease context. The main contribution of our work is to show how two different perspectives can be connected by their underlying ontological structure in order to provide more complete and robust management of complex data.

To such aim, the remainder of this work is organized as follows: Section 2 describes related works on Conceptual Modeling in the Health domain. Section 3 presents the Research Methodology applied in this work. Section 4 describes the development of the Conceptual Schema of the Macular Pathology (CSMP) as the basis of the G-MAC information system. Section 5 introduces the Conceptual Schema of the Human Genome (CSHG) and the SILE method as the basis of the DELFOS Oracle (the information system used to determine the genetic causes of the disease). Finally, Section 6 presents our conclusions and outlines future work.

2 Related Works

A wide variety of research works aim to provide solutions to existing problems in the context of health, in order to have a direct impact on the quality of care. The advantages of using Conceptual Modeling (CM) techniques in clinical practice have been demonstrated several times [7], and software engineers use them to add value when defining their development processes. In the context of Information Systems Engineering, CM is used to obtain a (general) description of how a system is organized and operated [6]. However, it is also used to clearly describe a specific domain by defining the main entities involved and their relationships with each other [8]. This allows different system users to share a common understanding of the domain, thus increasing its suitability for solving the analytical requirements.

The application of CM techniques in Bioinformatics with the aim of improving the management of clinical and genomic data has given rise to a wide

variety of research works that seek to demonstrate their usefulness and benefit for this community.

One of the first proposals was presented by Paton and Bornberg-Bauer (2000, 2002) in [10] [11]. In these works, the authors describe the structure of the genome from different perspectives: the description of the genome of the eukaryotic cell, the interaction between proteins, and the transcriptome, etc. This was a preliminary work that served as the basis of future research works. In 2004, a proposal by Ram et. al. [12] focused on the specialized application of conceptual modeling principles to the description of proteins in which large amounts of data with a fairly complex structure were required. The aim of this work was to facilitate the development of user-friendly tools to search and compare proteins by their structure.

In following years, research efforts focused on the development of widely used ontologies to precisely describe the main elements of DNA sequencies. Examples of these ontologies are Sequence Ontology [48], Variation ontology [49], and Gene Ontology [50].

In 2018, with the increasing use of Next Generation Sequencing technologies, Reyes proposed the Conceptual Schema of the Human Genome (CSHG) [14] as a basis to build solutions that are able to deal with the increasing complexity of genomic data (heterogeneity, dispersion, isolation). Using the CSHG as the basis, in 2020, Pastor et. al. present the application of CM techniques for improvement in the management of genomic data [9] [18]. These works provide a solution that helps researchers to organize, store and process information by focusing on the data that is relevant, and minimizing the impact of information overload in clinical and research contexts.

Specific works that focus on the use of CM for the management of health data related to specific diseases were proposed in 2017 by Burriel with the development of an information system to manage clinical and genomic data associated with breast cancer [13] and in 2019 by Arevshatyan et. al. with the description of a conceptual model for the understanding of Neuroblastoma [16]. Once again, the use of CM techniques became crucial to efficiently manage and analyze the complexity of the domain.

The above-mentioned works constituted a valuable step forward in the development of specific solutions intended to increase the understanding of the mechanisms that lead to human disease. Nevertheless, the complexity and heterogeneity of each disease requires the specialization of these approaches and the development of new models.

In this work, the need to incorporate Software Engineering techniques that help ophthalmology specialists to attack a large number of problems that arise when managing the information associated with the treatments and follow-ups of their patients has been detected. This is due to the fact that there is a high dispersion of the information due to the use of files and documents in different formats (e.g., *.doc, *.xls, *.csv, etc.) and with different structures. These problems increase when trying to provide a holistic perspective of the disease that includes both clinical data and genomic data. Currently, there is only one

program worldwide, called FRB or Project "*Fight Retinal Blindness!*"[1] that allows specialists to follow the progress of the patient (individual treatment for each affected eye) and generate graphs that help to obtain a better diagnosis. Nevertheless, the license per hospital is quite expensive, has limited access, and does not include the genomic perspective.

The lack of specific solutions to these problems is the motivation for this work as a way of bringing together all of the existing knowledge of the "retina-macula" domain and allowing specialists to address the problems they face with greater clarity and precision.

3 Research Methodology

In our research, we apply the Design Science (DS) Methodology [17] which is defined as *"the design and investigation of artifacts in context"*. Our ultimate goal is to provide a software platform (the artifact) that is capable of facilitating the management of clinical and genomic data associated with AMD. This work is part of the clinical domain, specifically the ophthalmology service (the context), it focuses on the application of conceptual modeling techniques and Software Engineering to guarantee a solution that enhances the knowledge acquired in hospital centers. Figure 1 presents the steps of the engineering cycle for research, which we apply according to Design Science.

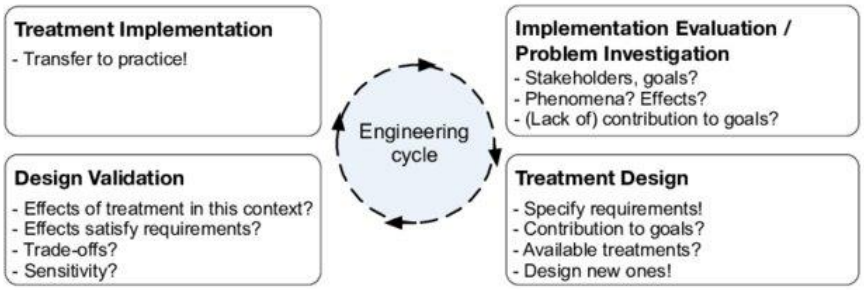

Fig. 1 Design Science as a regulatory cycle [17].

1. **Implementation Evaluation / Problem Investigation:** This task presents the motivation for the work and the benefits of applying conceptual modeling techniques in the domain under study (Section 1). The research methodology used during the study is also included (Section 3). The *"Implementation Evaluation"* phase is outside the scope of this work.

[1] `https://www.savesightinstitute.org.au/research-units/save-sight-registries/fight-retinal-blindness/`

2. **Solution Specification:** This task presents the state of the art on the use of conceptual modeling in the health domain as well as in the field of ophthalmology and AMD (Section 2). The proposal of the Conceptual Schema of the Macular Pathology is also presented, this serves as ontological support for the development of the G-MAC platform, which will allow the management of the clinical data of the work (Section 4).
3. **Specification Validation:** In this task, two software solutions that are focused on promoting the two perspectives of the work are presented: 1) G-MAC, as a prototype for the treatment of clinical data (Section 4); and 2) DELFOS Oracle, as a prototype for the study and analysis of variations associated with AMD (Section 5).
4. **Specification Implementation:** This task presents the results obtained after completing the study from the clinical (Section **4.3**) and genomic (Section **5.5**) perspectives, as well as the contributions of the work to the scientific community (Section 6).

Through the application of DS applied to the engineering cycle, the tasks associated with the design cycle have been carried out and the last step validates the solution provided by applying it in a real case study.

4 Clinical Perspective

The Clinical Perspective describes the steps of our proposal in order to allow ophthalmologists to effectively track the evolution of their patients with a high level of detail. First, we present the Conceptual Schema of the Macular Pathology (CSMP), the model that we have used to obtain a shared understanding of the domain and ensure data interoperability. Second, we describe the G-MAC platform, which gives ophthalmologists a tool to track the evolution of patients and improve the macula treatments. Third, we validate the G-MAC platform with the help of a group of experts (doctors, medical specialists, nurses, and research staff) from the Ophthalmology Services of three hospitals in the Valencian community.

4.1 The Conceptual Schema of the Macular Pathology (CSMP)

The use of conceptual schemas facilitates the understanding of the relevant concepts of complex and continuously growing contexts. A Conceptual Schema makes the thoughts and knowledge of domain experts explicit, reinforcing this so that the experts can communicate using a shared set of terms. It helps to improve data management in dimensions like integration or semantic meaning. Also, conceptual schemas are the foundation on which IS that use a Model-Driven Development (MDD) approach are built.

Before starting the modeling tasks, a series of brainstorming and discussion meetings were held with the group of experts in ophthalmology to study all of the relevant elements and dimensions in this domain. The entire medical

process of the Ophthalmology Services has been taken into account during the definition of the CSMP. We considered the entire process: starting from when a patient arrives at the hospital and has the first consultation with the doctor, and finishing when the medical treatment ends. This includes both the clinical management carried out by professionals in ophthalmology (e.g., hospitals, doctors, drugs, etc.) as well as the treatment of patient information.

In ophthalmology departments, in an initial phase, there is a first consultation with the patient. If the doctor considers it appropriate to start the treatment, they proceed to take the patient's data. This personal information contains the following attributes: name, date of birth, address, etc. These attributes are represented in the conceptual model through the *"Patient"* class.

It is also convenient to know if the patient has a medical history, for example, if the patient has suffered some type of heart attack, is or has been a smoker, has diabetes, cholesterol, obesity, etc., or if the patient has any type of disease that can affect the eye: *Age-related Maculopathy*, *Diabetic Macular Edema*, or *Central Retinal Vein Occlusion*. All of this information can be represented in the model through the *"Anteced_Medical"* and *"Disease"* classes respectively.

In addition, it is relevant to keep track of the patient's anthropometric data, such as weight, abdominal girth, and height, since these can significantly influence the evolution of the disease. Through the "Anthropometric_data " class, we can obtain a more efficient history, thus enhancing the diagnosis.

Once all of the information related to the patient has been collected, a series of *medical reviews* or *checkups* (*revisions*) is carried out where the evolution of the patient is observed. The *"Revision"* class allows knowing the date on which the checkups took place, the reason for them, the observations made by the specialist, and the total cost (it is important to highlight that the concept of *"Revision"* is very key in this work since it presents the combination of three factors: i) *new consultation with the patient*, ii) *a new diagnosis*, and iii) *the treatment provided to the patient by the specialist*). The total cost will allow the expenses to be determined in the event that treatments have been scheduled for the patient. This value is a derived attribute obtained that is from the sum of the prices of each of the drugs used in the treatments.

The first step in generating the G-MAC Information System is to create the *Conceptual Schema of the Macular Pathology (CSMP)* since we have envisioned G-MAC to be an MDD platform. The CSMP is depicted in Fig. 2 [43]. It is composed of three different views, which encapsulate the set of entities that are conceptually related in some way (See details in [1]):

- **The Organizational View**: This view groups those concepts that represent information associated with the Spanish public health system. Here, we focused on the management, human resources, and organizational infrastructure (healthcare organization).
- **The Patient View**: This view groups those concepts necessary to identify patients and get contextual information regarding their health conditions.

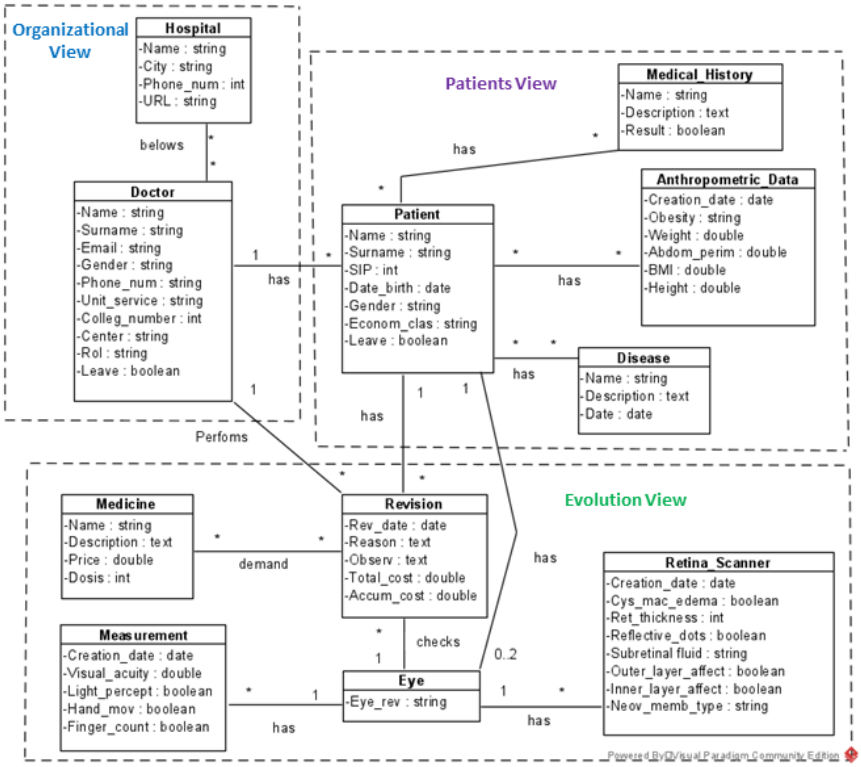

Fig. 2 Conceptual Schema of the Macular Pathology (CSMP) [43].

The core concept of this view is the PATIENT class, which is used to identify patients that suffer from AMD.

– **The Evolution View**: This view is responsible for grouping those concepts that are necessary to monitor the evolution of disease in either or both eyes of the patient. This means that if the doctor checks both eyes, two revisions will be created on the same date. The reason for this is that each eye can be at a different stage of the disease. Here, we defined the REVISION that doctors carry out, including both economic (total cost, accumulated cost) and health (observations, date, reasons) data. Finally, the *"Retina _Scanner"* and *"Measurement"* classes contain the data related to the studies that the specialist performs on each eye to see the degree of involvement of the disease (AMD).

After having defined the conceptual model (CSMP) that gathers the knowledge of the Retina-Macula domain, we can talk about the design and implementation of the resulting IS (Section 4.2), which has served as the base/support for its construction.

4.2 The G-MAC Information System

This section presents the solution[2] generated from the conceptual model (CSMP) defined in **Section 4.1**. This conceptual model has served as ontological support in the creation of the software solution, which has been developed with innovative technologies that have great support within the Software Engineering (SE) community. This is all for the purpose of providing an IS that responds to the needs of clinical experts in ophthalmology services, and that also has the appropriate degree of maturity for its application in medical practice.

The first challenge is to provide the services and the functionality of the application through the publication of a REST API. The second challenge is to develop a user interface, that allows communication with the server (*backend*) and consumes the services published by this API (See the solution architecture in [43], [1].

The design and development of the G-MAC information system was carried out from three (3) different perspectives, which are defined below (See detail in [43]):

- *Data Design (data layer design):* This task includes the creation of the application data model, extracted from the CSMP. This requires a database server configuration process (DBMS) and the creation of entities or models that allow storing all of the information required by the IS.
- *Functional Design:* This task focuses on the construction of an API (Application Programming Interface) in order to provide a service that allows two software technologies to be connected together, facilitating the exchange of messages and information in a specific format.
- *Interface Design:* This task focuses on the design of the application's user interface. For this, the *Angular* framework and its *Angular-CLI*[3] have been used, which facilitate a faster and more agile workflow.

It is important to highlight that the design and development of the software solution for this work is based on an MVC architecture[4] pattern. To see the rest of the components of this architecture in detail (e.g., *database layer*, see [1], **Chapter 4**, pp. 50-57). This proposal increases the quality and integrity of the solution architecture. This work is part of an ongoing research project (TRL **4**). Figure 3 shows an example of the web interface on which the integration of all the improvements detected by clinical experts is maintained.

4.3 G-MAC Validation

in [17], Wieringa defines the validation phase as the evaluation of the use of an artifact in a context in order to predict the impact of the artifact in the real world by stakeholders. In this section, we validate our proposed solution with

[2] **G-MAC**, `https://genomics-hub.pros.dsic.upv.es:4000/#/login`

[3] Angular-CLI (command line interpreter), `https://angular.io/cli`

[4] `https://www.w3schools.in/mvc-architecture/`

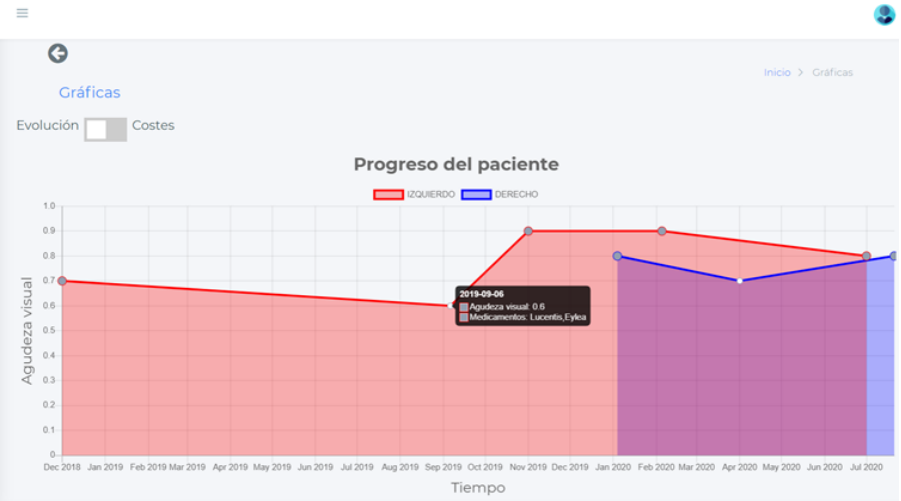

Fig. 3 G-MAC user interface: Evolution of a patient's visual acuity [43].

experts in retina-macula of the Valencian system of health. Our validation process included two main exercises (E) [43]:

- **(E1) Prototype Validation**: conducting several meetings with software developers and domain experts.
- **(E2) Expert Opinion**: reviewing the artifact under laboratory conditions together with domain experts to obtain their feedback.

In *prototype validation* (E1), two types of meetings were performed. In the first type of meetings, we met with software developers from the PROS Research Center[5] in order to validate the architecture and the developing process of the generated artifact. These meetings also allowed us to detect future improvements at the technological level. In the second type of meetings, we met with domain experts in order to validate the that the gathered requirements were appropriately implemented.

In *expert opinion* (E2), we validated the functionality of the generated artifact. We used the *User Interview* technique, which allowed us to gather the domain experts' feedback in an agile way [19]. The expert opinion exercise was divided into two steps: i) *user observation*, and ii) *user interrogation*.

1. In user observation, we asked the users to perform a set of predefined tasks using the G-MAC information system. These tasks were defined based on the user requirements and feedback obtained in the previous exercise (i.e., prototype validation). The experts that performed the tasks included medical staff (Ophthalmic specialists and research team) from three hospitals: Vega Baja (Orihuela), Lluís Alnayís (Xátiva), and Virgen de los Lirios (Alcoy). Since this was the first validation of the functionality of G-MAC, this

[5] `http://www.pros.webs.upv.es/`

exercise was conducted in a lab context so that we had greater control over this process.

For instance, the defined tasks included:

- *Register a patient*: The user has to register a new patient, show the list of existing patients, and search the new patient using two criteria: *"SIP number"* and *"full name"*.
- *Register a set of revisions*: The user has to create a set of new revisions for the newly added patient. Also, the user has to list the revisions of this patient and consult the most recent one.
- *Consult the evolution of a patient*: The user has to consult the evolution of i) visual acuity and ii) the treatment cost of the patient that was registered in the first task.

2. In user interrogation, we interviewed the users to know their opinion regarding the use of the application after performing the tasks defined above. The gathered feedback can be summarized in four (4) relevant points:
   - *The tool improves their daily work*: They stated that the use of the prototype allowed them to manage the information associated with the patients of the Ophthalmology Service efficiently. It also allowed them to achieve greater knowledge regarding the evolution of patients and to provide an improved and more accurate service to patients.
   - *The tool is easy to use*: The tool is easy to use and intuitive. They were satisfied with the workflow of the tool and the information displayed. In line with that, we observed that the users did not struggle to complete the defined tasks.
   - *It achieves stakeholders' goals*: The stakeholders expressed that they were satisfied with use of the prototype. They indicated that the data that is integrated, managed, and displayed will improve the management of the costs that are associated with the AMD treatment. In addition, they expect to be able to identify treatment patterns in order to increase the effectiveness of drug treatments so that patient recovery is improved.
   - *They identified potential improvements to make the tool more useful*: They provided us with minor visual improvements, which shows their interest in using the tool. The three improvements that they identified are the following:
     (a) The *"Result"* section of the *"Medical Background"* management can be eliminated since this requirement is already complete with the *"name"* and *"description"* sections.
     (b) In the case of *"Anthropometric Data"*, the system must allow the data to be stored, even if it is incomplete. This is because many times in the initial revision this data is not obtained, and the possibility of adding it to the system in the future should be considered.
     (c) Provide an interface where you can view all of the information associated with the checkup of each eye (i.e., one for the right eye and one for the left eye).

In general, the stakeholders' reported a positive use of the prototype. The reason is that the provided tool allowed them to achieve their goals and improve AMD treatment in an easy and intuitive way. They also mentioned that adopting the tool takes a small amount of time. This is crucial in their working context (i.e., *hospitals*).

In conclusion, it is an undeniable fact that integrating the data related to AMD treatment and automating patient monitoring produces greater satisfaction and benefits than performing this manually. The validation has provided encouraging findings, but the results should be studied further. After obtaining new funding for the project, a study based on TAR (*Technical Action Research* [17]) should be performed in order to solve the problems detected in the previous activities (**E1** and **E2**).

5 The Precision Medicine Perspective

There are studies that indicate that genetics complements environmental factors. These studies also report that genetics has a strong and consistent association with the risk of progression to advanced AMD (together with age, smoking, and previous cataract surgery) [22].

The study of genetics in patients improves the ability to predict progression to visual impairment [33]. To date, the existing genomic studies have been particularly successful in the case of AMD, generating a huge body of knowledge [26]. For example, even though one of the major causes of suffering AMD is **genetics**, a single variant is insufficient to cause AMD and more complex interactions between several variants seem to have an impact [34].

Currently, twenty chromosomal regions (i.e., protein-coding genes) associated with AMD[34] have been identified. These regions can be grouped into five clusters based on their biological relevance [52]:

1. **Complement system and immune response**. This cluster is composed of the following genes: CFH, C2, CFI, C3, and C9.
2. **Lipid transport**. This cluster is composed of the following four genes: APOE, LIPC, CETP, and BAIAP2L2.
3. **Extracellular matrix (ECM) remodeling** [35]. This cluster is composed of the following seven genes: COL8A1, COL10A1, TIMP3, ADAMTS9, TGFBR1, HTRA1, and B3GALTL.
4. **Angiogenesis** [36]. This cluster is composed of the following three genes: VEGFA, TGFBR1, and ADAMTS9.
5. **Cell survival**. This cluster is composed of the following three genes: ARMS2, RAD51B, and TNFRSF10A.

Identifying whether patients present genetic alterations significantly alters how they are treated [26]. The effect of genetics on the treatment response can be summarized in three points:

1. **Dietary supplementation**: One of the main means of reducing the progression of AMD is dietary supplementation with zinc and vitamins since

they act against one of the drivers of AMD (i.e., oxidative stress [38]. Some studies indicate that a good response to dietary supplementation is dependent on specific genotypes of the CFH gene [39, 40].

2. **Anti-VEGF treatment**: Intravitreal injections of VEGF antibodies have shown significant improvements in the visual acuity of patients with neovascular AMD. However, not all patients benefit from this treatment; it is estimated that 20% of patients treated with anti-VEGF drugs show no improvement in visual acuity [41]. It has been suggested that genetics is the cause of this situation [42].
3. **Complement inhibitors**: Several therapies are being developed to treat dry AMD. A subset of them target different components of the complement system (i.e., the use of complement inhibitors to depress the complement system). To date, the results of the clinical trials where complement inhibitors are used are not satisfying. One explanation for these results is that complement inhibitors are more effective in patients whose complement system is more affected [44]. Genetic variants have been associated with different complement levels, and their identification could improve the selection of patients who are eligible for complement inhibitor-based treatments [44].

As we have described, genetics is one of the main factors associated with AMD. In addition, patients' responses to existing treatments depend on their specific genetic code. These two facts justify the need for having a precision medicine perspective in order to achieve the two grand challenges mentioned in **Section 1**, namely, reducing the costs of treating AMD and finding drugs to treat AMD effectively.

5.1 The Conceptual Schema of the Human Genome (CSHG)

Modeling the human genome is a significant challenge whose accomplishment can have a potentially large impact on precision medicine by improving how genomics data is managed and shared. The huge amount of relevant concepts that play important roles, and their variability over time [45], implies that this domain has a high degree of complexity. We, therefore, use decomposition to ease the representation of this complexity, this results in the division of the CSHG into different conceptual views, each of which focuses on a specific dimension.

The initial work focused on representing the most relevant concepts when studying genomics, such as chromosomes or variations. This model was then expanded to include the concept of phenotype and its relations with other genomics components in the second version [46]. The second version drastically changed how the DNA sequence is represented: from a gene-centric vision to a chromosome-centric vision [14]. This version included the *"chromosome element"* class, which led to an increased generalization of the elements In the previous versions, only genes could be characterized; in this version, any sequence with specific functionality can be characterized (e.g., enchancers,

promoters, etc.). The third and most recent version is the biggest update of the model to date [31]. Since genomics is an ever-changing domain, our approach is to evolve the model continuously by adding new entities and updating the existing ones based on the most recent scientific discoveries as well as the experience from domain experts. This version expanded the representation of the transcription process, reevaluated the characterization of variants, included the changes caused by variations at the DNA, RNA, and amino acid level, and increased the generality of multiple concepts of the model.

The process of creating a holistic Conceptual Schema of the Human Genome (CSHG) [51] requires the integration of a set of different conceptual components that cover all of the relevant data that connect the genome structure (*genotype*) with its expression in the form of real-world behavior (*phenotype*). The CSHG created at the PROS Research Center of the Universitat Politècnica de València (UPV) has incrementally incorporated six views into one holistic conceptual model. These views are the following [51]:

1. The **Structural view** describes the structure of the human genome. This view focuses on the different chromosome elements that exist in our genome (transcribable elements –genes, exons. . . -, regulatory elements, conserved regions. . .) and how they are composed.
2. The **Variations view** characterizes changes in the sequence of the human genome regarding a reference sequence. This view allows us to identify what type of changes may occur in the genome.
3. The **Transcription view** models protein synthesis. This view deals with the many particularities of transcribing from DNA to RNA and translating from RNA to protein.
4. The **Pathways view** represents human metabolic pathways. This view describes the chemical reactions that explain the different molecular processes occurring in our body.
5. The **Proteome view** characterizes the structure and properties of proteins, i.e., macromolecules that play a fundamental role within every cell and metabolic reaction.
6. The **Bibliography and data sources view** details the information and sources related to the elements of the conceptual schema. This view focuses on identifying relevant information that is related to sources of valid information (publications, genome data sources, etc).

Conceptual models are intuitive, direct, and natural representations of a given domain. It allows us to answer fundamental questions by identifying relevant concepts and their relationships with each other and helps establish a common ontological framework. This framework facilitates both communication and knowledge evolution in complex domains, such as *genomics*. The use of a conceptual model provided us with a solution to deal with the particularities of one of the most complex domains (i.e., genomics).

5.2 The SILE Method

One of the main problems when performing genetic testing is the appropriate management of the vast amount of available data. Genetic data is characterized by its heterogeneity, dispersion, and variable levels of quality [20]. These constitute a bottleneck in the process of effectively determining the genetic causes of human disease. This has a deep impact on the quality of the results derived from the analysis of the genetic data as well as on the diagnosis and treatment of patients.

To solve, or at least minimize, the problems derived from the management of such complex data, we must develop Information Systems that are capable of providing solutions to the main tasks that are required to determine the most relevant variants that cause the phenotype under study (in our case study, the predisposition of having Retina-Macula degeneration). These tasks are: search and selection of the most suitable genomic data repositories, identification of the relevant data that complies with the minimum quality requirements, storage of the data in the appropriate data storage, and exploitation of the results in terms of identifying if a patient has any of the sequence variants identified as disease-causing. Each of these tasks presents well-known challenges to the Information Systems community such as the collection of dispersed data, the integration of heterogeneous data formats, the lack of ontological agreements in the representation of the concepts, the appropriate measure of data quality, etc.

The SILE method has been developed with the aim of providing a general and standardized solution to overcome these challenges in the specific and complex context of genetic data. SILE is the acronym derived from each of the tasks mentioned that are required to manage genomic data (*Selection, Identification, Load and Exploitation*). The method comprised of four consecutive stages, each of which provides solutions to resolve the specific problems that can arise. These include how to select the required genomic repositories to store the required data, how to determine the quality of the repositories and how to define the quality requirements that the data must fulfill, how to select the most suitable storage technology, and how to exploit the data according to the analytical requirements. The entire process is supported by the ontological structure provided by the CSHG, which ensures data consistency during the collection and integration process. More details about each of the states of the SILE method can be found in [20] and [21]. The method has been validated by experts in genetic diagnosis for the identification of the genetic causes of diseases such as migraine, Crohn's disease, and epilepsy.

5.3 The DELFOS Oracle

The SILE method has been developed as a guideline to build Information Systems that are able to efficiently manage genomic information. Therefore, its implementation requires technological support to automate the different

stages as much as possible. The DELFOS Oracle has been developed to do this. Figure 4 shows the connection between the SILE method, the DELFOS platform, and the CSHG.

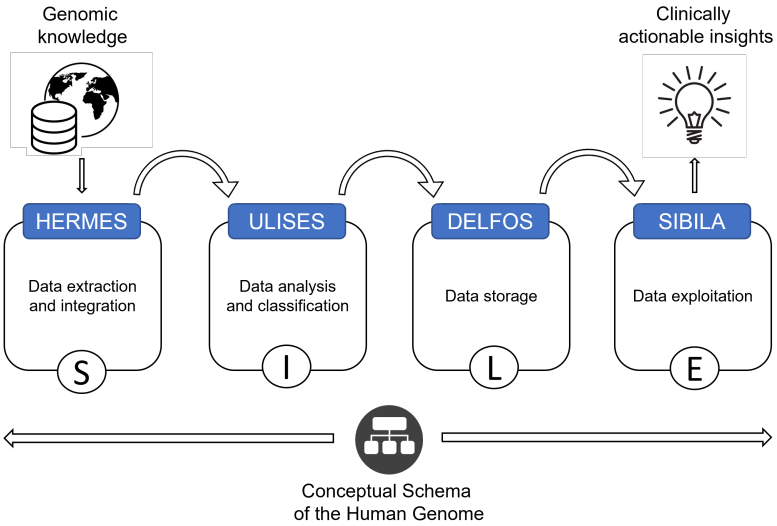

Fig. 4 The DELFOS platform as a technological support for the implementation of the SILE method.

DELFOS is composed by four modules (Hermes, Ulises, Delfos, and Sibila) which provide automation for each of the tasks required to succeed at each stage of the SILE method:

– **Hermes** is the module that provides support to the *Search* stage. This module is in charge of automating the connection, extraction, transformation, and integration of genomic data coming from different well-known genomic repositories. Hermes is implemented as a set of R libraries that allow the user to build personalized pipelines to integrate the data from different repositories. Once the user has selected the repositories that must be used, Hermes automates the extraction of the raw data and transforms the different formats according to the structure defined by the Conceptual Schema of the Human Genome (CSHG). This harmonization allows the identification and integration of common genomic entities (e.g., variants, genes, proteins) independently from the structure used by the source repository. The collection, transformation, and integration of the genomic data is automated using a set of mapping and transformation rules whose structure is defined by the SILE method. This means that the user does not need to have a deep knowledge of the underlying structure of each repository. This reduces the time required to collect and integrate the data, reducing the redundancy and the errors that commonly occur when performing these tasks manually.

- **Ulises** is the module that provides support to the *Identification* stage. This module is in charge of implementing the algorithms that determine the quality and relevance of the collected data for application in the clinical context. In our case study, the genetic variants are related to the development of Retina-Macula degeneration. The algorithm that determines the relevance of the variants is based on the recommendations provided by the SILE method for the identification of relevant genomic variants. These recommendations take into account different aspects of the variant context such as the clinical significance provided by experts, the genomic structure affected by the variant, and the statistical significance derived from case-control studies. Ulises is implemented as a web platform. Its input is the data obtained from Hermes. The output is the classification of the variants: *accepted* (variants with strong, limited, or moderate evidence); *to follow up* (variants whose relevance is near to being accepted); and *rejected* (variants that have not relevance). For each classification, Ulises also provides explanations about the reasons why the decision has been made, increasing trust in the results and allowing the user to customize the algorithm to specific disease contexts.
- **Delfos** is the module that supports the *Load* of the data in a database that is specifically designed to ensure the persistence of the genomic data and its further analysis. Since the main analytical requirement of this case study is to determine the presence of relevant variants in a patient's genome, the technology selected to store the relevant variants and make comparisons against a genome is an SQL database. The Delfos module automates the storage of the variants classified by Ulises and evaluates their quality according to the constraints defined by the SQL database in order to ensure consistency.
- **Sibila** is the module that provides all of the analytical and interaction tools that are required to allow the *Exploitation* of the data in the clinical context. The tools developed in Sibila allow the user to explore the data stored in Delfos in a comprehensive way, facilitating the decision-making process, and helping the inference of new knowledge from the stored data. Through the web interface, the user can explore the variants identified by Ulises as relevant, get access to different statistical analyses, and determine if their presence in a patient's genome explains the symptoms or could lead to the development of the disease.

The DELFOS platform has been used to explore existing knowledge about the genetic causes of Retina-Macula degeneration, to extract the relevant variants that comply with the quality criteria required to be used in a clinical context, and to create a database of knowledge to determine the implications of the presence of these variants in a patient's genome. The databases used to explore the relevant variants fulfill the quality requirements established by the SILE method in terms of currency, reliability, and accessibility. The stored data represents different genetic contexts: genotype-phenotype associations (ClinVar[6],

[6] `https://www.ncbi.nlm.nih.gov/clinvar/`

Ensembl[7], and LOVD[8]), polymorphism information (dbSNP[9]), published evidence (PubMed[10]), and case-control studies (GWAS Catalog[11]). All of this information is used by Ulises to determine the relevance of each variant and is finally stored in the Delfos database to be accessed by the user.

5.4 Analysis of Variants (Results)

Using the functionality provided by Hermes, we were able to explore the selected genomic repositories and integrate all of the evidence. About **2,996** variants were initially considered as being associated with the development of AMD. Table 1 summarizes the results obtained by each repository.

Repository	Variants
ClinVar	1,074
Ensembl	2,592
GWAS Catalog	287
LOVD	0
Total	**2,996**

Table 1 The variants obtained from each genomic repository.

As Table 1 shows, the LOVD database does not provide any results since none of the variants stored in this database are initially related to the disease under study. Nevertheless, since this database is frequently updated it is important to repeat the search periodically in case new relevant variants could be added. This can be easily done thanks to the automation that the DELFOS platform provides.

According to the summary shown in Fig. 5, the variants are mainly located in Chromosome 1 and correspond to single nucleotide variants whose role in the development of the disease was not determined. This supports the assertion regarding the lack of knowledge of the development of the disease and the complexity of precisely determining the specific causes of Retina-Macula degeneration.

Once the data has been collected from the origin data sources, transformed into a common format, and integrated, Ulises analyzed the evidence to determine the relevance of the variants for clinical practice. According to the results shown in Fig. 6, only 5% of the variants are considered to be relevant for use in a genetic diagnosis of the disease (1% are accepted with limited evidence and 4% are accepted with moderate evidence). Some variants (1%) are considered as being interesting to follow up in case new evidence appears that supports the role of the variant in the development of the disease.

[7] `http://www.ensembl.org/index.html`
[8] `https://www.lovd.nl/`
[9] `https://www.ncbi.nlm.nih.gov/snp/`
[10] `https://pubmed.ncbi.nlm.nih.gov/`
[11] `https://www.ebi.ac.uk/gwas/`

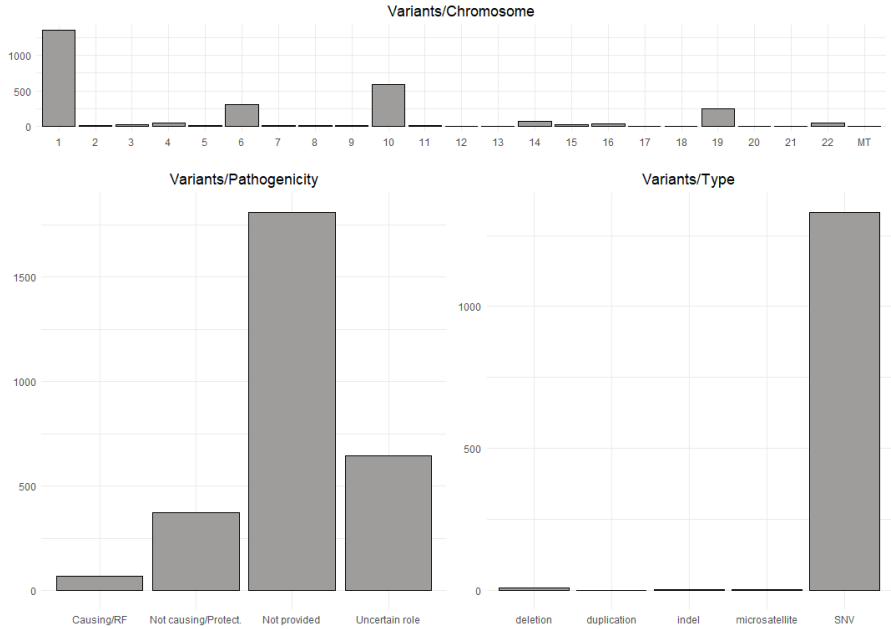

Fig. 5 Distribution of the variants by chromosome, pathogenicity, and type.

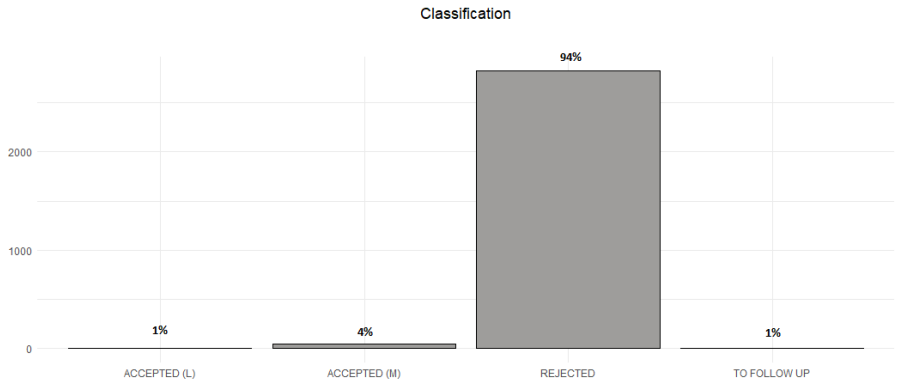

Fig. 6 Distribution of the variants by chromosome, pathogenicity, and type.

The reasons for rejecting the rest of the variants are the following:

- Even though the variants affect a gene of interest and are predicted to have a deleterious effect, no relevant studies support the association with the phenotype (0.12%).
- Even though the variants affect a gene of interest, they are not predicted to have a deleterious effect (5.65%).
- The variants are not clinically actionable (20.84%).

Variant	Gene	Phenotype	Type
Chr1:g.94098874-94098874:A>T(GRCh38)	ABCA4	AMD 2	SNV
Chr1:g.94111561-94111561:G>A(GRCh38)	ABCA4	AMD 2	SNV
Chr4:g.109760514-109760532:longchange(GRCh38)	CFI	AMD 13	Indel
Chr14:g.91870369-91870370:CT>?(GRCh38)	FBLN5	AMD 3	Deletion
Chr10:g.49472472-49472472:T>C(GRCh38)	ERCC6	AMD 5	SNV

Table 2 Example of variants accepted with moderate evidence.

- The clinical significance of the variants has not been interpreted and they do not affect a gene of interest (72.68%).
- There are conflicts in the interpretation of the clinical significance of the variants (0.2%).

Once the variants have been classified according to their relevance for the genetic diagnosis, the relevant ones are uploaded to the Delfos database and the details of each variant are available for exploration. As an example, a summary of some of the variants accepted with moderate evidence are shown in Table 2.

5.5 Validation by Clinical Experts

During the execution of the tasks related to the genomic study for AMD (**Section 5**), there has been a team of experts who have collaborated in each of the phases described above. This has made it possible to refine the search and identification processes of the relevant variations associated with AMD.

After completing an initial phase of research on the genetic implications in AMD with the aim of knowing and understanding the domain, the phases of the SILE method were carried out using the DELFOS Oracle. Thanks to this, a significant number of relevant variations for AMD were obtained from the different genomic repositories, which were studied and analyzed (**Section 5.4**) with the support of the clinical experts participating in the project.

The clinical experts from the different Ophthalmology Services carried out various tasks to validate the results obtained. These include review of the literature, review of the electronic medical records of the patients, and the study of preliminary cases shared in different hospital centers in the Valencian Community. After concluding the review phase of the results, the experts indicated that this work is an interesting proposal for transfer to the clinic, since nowadays genomic studies are not carried out routinely in Ophthalmology Services due to the high costs. However, work is currently being done to assess genetic predisposition in the short term, in order to provide personalized precision medicine in the different Ophthalmology Services for the benefit of patients (*treatments* and *prevention*).

6 Conclusions and Future Directions

The objective of this work is to enhance the results obtained from the clinical perspective by integrating the genomic perspective, which contributes directly to the treatment and prevention of diseases, allowing patients to receive personalized precision medicine.

This research project has focused on providing an Information System (G-MAC) that ensures efficient and effective management of data associated with the Retina-Macula. As an extension of [43], this work addresses the inclusion of the genetic implications regarding AMD. The DELFOS Oracle, which consists of an innovative solution developed by the PROS Research Center (as a member of VRAIN institute), has been applied to the treatment of genomic data in order to obtain valuable information for clinical experts.

After developing all of the previous work from the clinical perspective (*CSMP*, *prototype of the software tool*, *validation tasks*, etc.) [43], the need to incorporate valuable information at the genomic level was detected, allowing us to compare some preliminary studies in this direction carried out in different hospital centers.

To achieve the proposed objective, a series of tasks were carried out. These include:

- the study and analysis of the genetic implications in AMD,
- the application of a methodological approach (SILE Method) for the identification of relevant variations in AMD, having obtained a total of **2.996** variants in the four repositories used,
- the processing and curation of the data obtained with the support of the DELFOS Oracle,
- an analysis of the results obtained (classification of the variations according to their relevance for genetic diagnosis),
- a validation by clinical experts of the genomic data obtained by applying the DELFOS Oracle.

Thanks to this work, a connection has been created between the currently managed clinical data and the existing knowledge in the genomic domain. This is to advance *Precision Medicine*, which seeks to improve the quality of care (e.g., treatments) of patients suffering from this disease.

The following aspects are planned as future lines of work: i) the extension of the CSMP that supports the G-MAC information system (integrating new knowledge/tests, methodologies, etc.); ii) the development of a software platform that allows the integration and exploitation of clinical and genomic data from patients and achieving a higher degree of maturity (TRL) of the G-MAC prototype; iii) the continuation of the G-MAC validation phase by executing the TAR technique; and iv) the study and analysis of the requirements necessary to obtain certification as a medical product to facilitate its transfer to the different hospitals that request it. In the future, each center is expected to have its own license to use the G-MAC information system.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the PROS Research Center Genome group for the fruitful discussions regarding the application of CM in the medical field. This work was supported by the Valencian Innovation Agency and Innovation through the OGMIOS Project (INNEST/2021/57), the Preparatory Action - UPV-FISABIO (A36-G-MAC, 2019), and the Spanish State Research Agency through the DELFOS project (MICIN/AEI/10.13039/501100011033) and co-financed with ERDF and the European Union Next Generation EU/PRTR.

References

1. Marco Palomares, A. (2020). Desarrollo Dirigido por Modelos para el Análisis y Gestión de los Datos asociados a la Patología Macular. Universitat Politècnica de Valencia. http://hdl.handle.net/10251/151886
2. Plazzotta, F., Luna, D., and González Bernaldo de Quirós, F. (2015). Sistemas de información en salud: integrando datos clínicos en diferentes escenarios y usuarios. Revista Peruana de Medicina Experimental y Salud Pública, 32(2), 343-351.
3. Instituto de Microcirugía Ocular. (2018). DMAE: Síntomas y tratamientos — IMO. https://www.imo.es/es/dmae
4. Pardo, I. C., Varona, D. G., and de Miranda Remedios, D. I. (2014). Degeneración macular relacionada con la edad. Archivo Médico Camagüey, 12(2).
5. Ruiz-Moreno, J. M., Arias-Barquet, L., Armadá-Maresca, F., Boixadera-Espax, A., García-Layana, A., Gómez-Ulla-de-Irazazábal, F., ... and Suárez-de-Figueroa, M. (2009). Guías de práctica clínica de la SERV: Tratamiento de la degeneración macular asociada a la edad (DMAE) exudativa. Archivos de la Sociedad Española de Oftalmología, 84(7), 333-344.
6. Olivé, A. (2007). Conceptual modeling of information systems. Springer Science and Business Media.
7. Aguilera, D., Gómez, C., and Olivé, A. (2013). Enforcement of conceptual schema quality issues in current integrated development environments. In International Conference on Advanced Information Systems Engineering (pp. 626-640). Springer, Berlin, Heidelberg.
8. Delcambre, L. M., Liddle, S. W., Pastor, O., and Storey, V. C. (2018). A reference framework for conceptual modeling. In International Conference on Conceptual Modeling (pp. 27-42). Springer, Cham.
9. Pastor, Ó., León, A., Reyes Román, J. F., García S., A., and Casamayor, J. C. (2021). Using conceptual modeling to improve genome data management. Briefings in Bioinformatics, 22(1), 45-54. DOI: 10.1093/bib/bbaa100
10. Paton, N. W., Khan, S. A., Hayes, A., Moussouni, F., Brass, A., Eilbeck, K., ... and Oliver, S. G. (2000). Conceptual modelling of genomic information. Bioinformatics, 16(6), 548-557.
11. Bornberg-Bauer, E., and Paton, N. W. (2002). Conceptual data modelling for bioinformatics. Briefings in bioinformatics, 3(2), 166-180.
12. Ram, S., and Wei, W. (2004, November). Modeling the semantics of 3D protein structures. In International Conference on Conceptual Modeling (pp. 696-708). Springer, Berlin, Heidelberg.
13. Burriel Coll, V. (2017). Diseño y Desarrollo de un Sistema de Información para la Gestión de Información sobre Cáncer de Mama (Doctoral dissertation, Universitat Politècnica de València). https://doi.org/10.4995/Thesis/10251/86158
14. Reyes Román, J. F. (2018). Diseño y Desarrollo de un Sistema de Información Genómica Basado en un Modelo Conceptual Holístico del Genoma Humano (Doctoral dissertation, Universitat Politècnica de València). https://doi.org/10.4995/Thesis/10251/99565

15. García, A., Reyes Román, J. F., Casamayor, J. C., and Pastor, O. (2019). Towards an Effective and Efficient Management of Genome Data: An Information Systems Engineering Perspective. In International Conference on Advanced Information Systems Engineering (pp. 99-110). Springer, Cham.
16. Arevshatyan, S., Reyes Román, J. F., Burriel, V., Cañete, A., Castel, V., and Pastor, Ó. (2019). Integration and Analysis of Clinical and Genomic Data of Neuroblastoma applying Conceptual Modeling.
17. Wieringa, R. (2014). Design Science Methodology for Information Systems and Software Engineering. Springer-Verlag Berlin Heidelberg, 1-332. DOI: https://doi.org/10.1007/978-3-662-43839-8
18. García, A., León Palacio, A., Reyes Román, J. F., Casamayor, J. C., and Pastor, O. (2020). Towards the understanding of the human genome: a holistic conceptual modeling approach. IEEE Access, 8, 197111-197123. DOI: 10.1109/ACCESS.2020.3034793
19. Jenny Preece, H. S., and Rogers, Y. (2015). Interaction design: Beyond human-computer interaction.
20. León Palacio, A., Pastor, Ó., and Casamayor R., J. C. (2018). A method to identify relevant genome data: conceptual modeling for the medicine of precision. In International Conference on Conceptual Modeling (pp. 597-609). Springer, Cham.
21. León Palacio, A., and Pastor, Ó. (2018). Smart data for genomic information systems: the SILE method. Complex Systems Informatics and Modeling Quarterly, (17), 1-23. DOI: 10.7250/csimq.2018-17.01
22. Velez-Montoya, R., Oliver, S. C., Olson, J. L., Fine, S. L., Quiroz-Mercado, H., and Mandava, N. (2014). Current knowledge and trends in age-related macular degeneration: genetics, epidemiology, and prevention. Retina, 34(3), 423-441. DOI: 10.1097/IAE.0000000000000036
23. National Eye Institute (2021). Age-Related Macular Degeneration, https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/age-related-macular-degeneration
24. Katta, S., Kaur, I., and Chakrabarti, S. (2009). The molecular genetic basis of age-related macular degeneration: an overview. Journal of genetics, 88(4), 425-449. DOI: 10.1007/s12041-009-0064-4
25. Resnikoff, S., Pascolini, D., Etya'Ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G. P., and Mariotti, S. P. (2004). Global data on visual impairment in the year 2002. Bulletin of the world health organization, 82, 844-851. DOI: /S0042-96862004001100009
26. de Jong, E. K., Geerlings, M. J., and den Hollander, A. I. (2020). Age-related macular degeneration. Genetics and genomics of eye disease, 155-180. DOI: 10.1016/B978-0-12-816222-4.00010-1
27. Loewenstein, A. (2007). The significance of early detection of age-related macular degeneration: Richard and Hinda Rosenthal Foundation lecture, The Macula Society 29th annual meeting. Retina, 27(7), 873-878. DOI: 10.1097/IAE.0b013e318050d2ec
28. Boyer, D. S., Antoszyk, A. N., Awh, C. C., Bhisitkul, R. B., Shapiro, H., Acharya, N. R., and MARINA Study Group. (2007). Subgroup analysis of the MARINA study of ranibizumab in neovascular age-related macular degeneration. Ophthalmology, 114(2), 246-252. DOI: 10.1016/j.ophtha.2006.10.045
29. Baird, P. N., Hageman, G. S., and Guymer, R. H. (2009). New era for personalized medicine: the diagnosis and management of age-related macular degeneration. Clinical and experimental ophthalmology, 37(8), 814-821. DOI: 10.1111/j.1442-9071.2009.02136.x
30. Senra, H., Macedo, A. F., Nunes, N., Balaskas, K., Aslam, T., and Costa, E. (2019). Psychological and psychosocial interventions for depression and anxiety in patients with age-related macular degeneration: a systematic review. The American Journal of Geriatric Psychiatry, 27(8), 755-773. DOI: 10.1016/j.jagp.2019.03.001
31. García, A., León Palacio, A., Reyes Román, J. F., Casamayor, J. C., and Pastor, O. (2020). Towards the understanding of the human genome: a holistic conceptual modeling approach. IEEE Access, 8, 197111-197123. DOI: 10.1109/ACCESS.2020.3034793
32. Palacio, A. L., and López, Ó. P. (2018). Smart data for genomic information systems: the SILE method. Complex Systems Informatics and Modeling Quarterly, (17), 1-23. DOI: 10.7250/csimq.2018-17.01

33. Seddon, J. M., Reynolds, R., Maller, J., Fagerness, J. A., Daly, M. J., and Rosner, B. (2009). Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables. Investigative ophthalmology and visual science, 50(5), 2044-2053. DOI: 10.1167/iovs.08-3064
34. Fritsche, L. G., Fariss, R. N., Stambolian, D., Abecasis, G. R., Curcio, C. A., and Swaroop, A. (2014). Age-related macular degeneration: genetics and biology coming together. Annual review of genomics and human genetics, 15, 151-171. DOI: https://doi.org/10.1146/annurev-genom-090413-025610
35. Bonnans, C., Chou, J., and Werb, Z. (2014). Remodelling the extracellular matrix in development and disease. Nature reviews Molecular cell biology, 15(12), 786-801. DOI: 10.1038/nrm3904
36. Folkman, J. (2006). Angiogenesis. Annu. Rev. Med., 57, 1-18. DOI: https://doi.org/10.1146/annurev.med.57.121304.131306
37. Mehta, S. (2015). Age-Related Macular Degeneration. Primary Care: Clinics in Office Practice. DOI: 10.1016/j.pop.2015.05.009
38. Beatty, S., Koh, H. H., Phil, M., Henson, D., and Boulton, M. (2000). The role of oxidative stress in the pathogenesis of age-related macular degeneration. Survey of ophthalmology, 45(2), 115-134. DOI: 10.1016/S0039-6257(00)00140-5
39. Vavvas, D. G., Small, K. W., Awh, C. C., Zanke, B. W., Tibshirani, R. J., and Kustra, R. (2018). CFH and ARMS2 genetic risk determines progression to neovascular age-related macular degeneration after antioxidant and zinc supplementation. Proceedings of the National Academy of Sciences, 115(4), E696-E704. DOI: 10.1073/pnas.1718059115
40. Seddon, J. M., Silver, R. E., and Rosner, B. (2016). Response to AREDS supplements according to genetic factors: survival analysis approach using the eye as the unit of analysis. British Journal of Ophthalmology, 100(12), 1731-1737. DOI: 10.1136/bjophthalmol-2016-308624
41. Brown, D. M., Michels, M., Kaiser, P. K., Heier, J. S., Sy, J. P., and Ianchulev, T. (2009). Ranibizumab versus verteporfin photodynamic therapy for neovascular age-related macular degeneration: two-year results of the ANCHOR study. Ophthalmology, 116(1), 57-65. DOI: 10.1016/j.ophtha.2008.10.018
42. Tsilimbaris, M. K., López-Gálvez, M. I., Gallego-Pinazo, R., Margaron, P., and Lambrou, G. N. (2016). Epidemiological and clinical baseline characteristics as predictive biomarkers of response to anti-VEGF treatment in patients with neovascular AMD. Journal of ophthalmology, 2016. DOI: 10.1155/2016/4367631
43. Reyes Román, J. F., Marco-Palomares, A., García S., A., and Pastor, O. (2021). A Model-Based Application for the Effective and Efficient Management of Data Associated with Retina-Macula Pathology. In Enterprise, Business-Process and Information Systems Modeling (pp. 366-379). Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-79186-5_24
44. Lorés-Motta, L., de Jong, E. K., and den Hollander, A. I. (2018). Exploring the use of molecular biomarkers for precision medicine in age-related macular degeneration. Molecular diagnosis and therapy, 22(3), 315-343. DOI: 10.1007/s40291-018-0332-1
45. García Simón, A., Costa Sánchez, M., and Pastor, O. (2021). Characterization and Treatment of the Temporal Dimension of Genomic Variations: A Conceptual Model-Based Approach. In International Conference on Conceptual Modeling (pp. 104-113). Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-88358-4_9
46. Reyes Román, J. F., Pastor, O., Casamayor, J. C., and Valverde, F. (2016). Applying conceptual modeling to better understand the human genome. In International Conference on Conceptual Modeling (pp. 404-412). Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-46397-1_31
47. León, A., and Pastor, Ó. (2021). Enhancing Precision Medicine: A Big Data-Driven Approach for the Management of Genomic Data. Big Data Research, 100253. DOI: https://doi.org/10.1016/j.bdr.2021.100253
48. Eilbeck, K., Lewis, S.E., Mungall, C.J. et al (2005). The Sequence Ontology: a tool for the unification of genome annotations. Genome Biol 6, R44. https://doi.org/10.1186/gb-2005-6-5-r44
49. Vihinen, M. (2014). Variation Ontology for annotation of variation effects and mechanisms. Genome research, 24(2), 356-364.

50. Ashburner, M., Ball, C., Blake, J. et al. (2000). Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29. https://doi.org/10.1038/75556
51. García, A., Palacio, A. L., Román, J. F. R., Casamayor, J. C., and Pastor, O. (2021). A Conceptual Model-Based Approach to Improve the Representation and Management of Omics Data in Precision Medicine. IEEE Access, 9, 154071-154085.
52. Nurk, S., Koren, S., Rhie, A., Rautiainen, M. et al. (2022). The complete sequence of a human genome. Science, 376(6588), 44-53. DOI: 10.1126/science.abj6987