Delfos Platform: Information System for the Management of Genomic Variations

Alberto García S.[1,*,†], José Fabián Reyes Román[1,†], Mireia Costa[1,†], Ana León[1,†] and Oscar Pastor[1,†]

[1]*Universitat Politècnica de València, Camí de Vera S/N, Valencia, 46022, Spain*

Abstract

The practical application of the vast amount of genomic data publicly available for clinical practice is a bottleneck for advancing Precision Medicine. Unresolved issues in the genomics domain include dispersion, heterogeneity, discrepancies, a lack of standardization, and data quality issues. In this project, we create the Delfos platform, a conceptual model-based solution developed with a rigorous methodological and ontological foundation, with the main goal of minimizing the impact of these issues when transferring research findings to clinical practice. To validate this platform, we collaborate with geneticists and doctors to help them improve patient diagnosis and treatment in three health areas: cardiology, oncology, and inherited retinal diseases.

Keywords

Precision Medicine, Conceptual Modeling, Big Data,

1. Introduction

Recent advances in high-throughput sequencing technologies [1] have resulted in terabytes of heterogeneous omics data of genomes, DNA variants, population frequencies, association studies, and other topics. These data enable researchers and clinicians to understand the relationship between the genotype and various phenotypic manifestations in order to improve human disease diagnosis, treatment, and prevention (i.e., interpreting the clinical significance of DNA variants). Geneticists strive to efficiently collect and integrate all of this knowledge to interpret, prioritize, and summarize significant information in the context of published literature, clinical trials, and various knowledge databases and repositories [2]. After this process, clinical experts use the generated knowledge to provide personalized patient care in clinical practice.

One of the most prominent problems to manage genomics data and interpreting the clinical significance of DNA variants is associated with the vast amount of existing data repositories. These repositories are classified into three types:

RPE@CAiSE'23: Research Projects Exhibition at the International Conference on Advanced Information Systems Engineering, June 12–16, 2023, Zaragoza, Spain

*Corresponding author.

†These authors contributed equally.

algarsi3@pros.upv.es (A. García S.); jreyes@pros.upv.es (J. F. Reyes Román); miscossan@pros.upv.es (M. Costa); aleon@vrain.upv.es (A. Leon); opastor@dsic.upv.es (O. Pastor)

0000-0001-5910-4363 (A. García S.); 0000-0002-9598-1301 (J. F. Reyes Román); 0000-0002-8614-0914 (M. Costa); 0000-0002-0877-7063 (A. Leon); 0000-0002-1320-8471 (O. Pastor)

1. Raw Data: Constituted by raw omics data generated by sequencing technologies. These data can be described in terms of DNA sequences, scientific articles, identification of DNA variants, proteins, etc.
2. Data Collection: Databases and repositories that collect data on a particular aspect of genomics. These repositories process and distribute "raw data", making it available to the community. Examples of such a repositories include PubMed [3], RefSeq [4], or dbSNP [5].
3. Integration and Interpretation: Platforms that provide new insights based on repositories classified into the "data collected" type. Their primary goal is to determine whether a DNA variant is linked to disease expression, drug or treatment response, etc. ClinVar [6], Varsome, [7] and Mastermind [8] are examples of such repositories.

Although these third-layer platforms represent a significant advancement in interpreting the clinical significance of DNA variants, they are only a partial solution and have their own limitations. Conflicting results are expected due to differences in how clinical laboratories interpret empirical evidence. Furthermore, the complexity and subjectivity of the interpretation process make automation difficult [9]. For instance, Varsome requires the user's manual intervention to complete the interpretation. Going further with another example, while MasterMind assists in finding relevant literature, their findings must be re-evaluated whenever new literature becomes available.

All of these issues together form a bottleneck that prevents precision medicine from becoming a common and standardized practice. This includes i) efficiently translating research findings into clinical practice and ii) assessing the risk-benefit of any genetic-based drug or treatment for patients. In many cases, simple questions like "Is this variant the cause of the symptoms my patient is experiencing?" cannot be answered with confidence.

There is a need for additional tools and platforms to assist genome analysts in managing genomics information and help clinical experts rely more on their decisions.

2. Project Objectives & Tangible Outputs

The main goal of this project is to contribute to make precision medicine a common and standardized practice. This is subdivided as follows.

2.1. Delfos

The first step is developing a Genome Information System (GeIS), called Delfos, that allows for:

1. Increase the automation of the interpretation of the clinical significance of DNA variants.
2. Increase the level of confidence of clinical experts when interpreting the clinical significance of DNA variants.

Delfos is a conceptual model-based platform that can be classified in a novel type, which we called clinical actionability [10], that is placed on top of the "Integration and Interpretation" type. (see Fig. 1). Delfos implements the four stages of the SILE method [11], specially defined for managing genomics data.

The outputs of this step are the following:

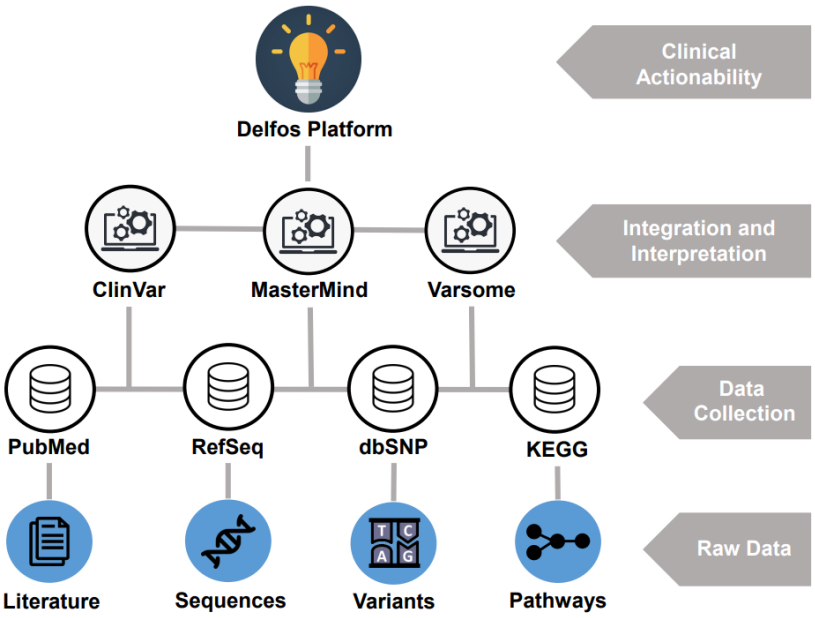

Figure 1: Types of datasources according to their role in the genetic diagnosis process.

1. **The Conceptual Schema of Genomics**: The ontologically well-grounded basis for developing the tool. This schema has been published in:
   - *García S., A., Costa, M., Leon, A. et al.* ***The challenge of managing the evolution of genomics data over time: a conceptual model-based approach.*** *BMC Bioinformatics 23 (Suppl 11), 472 (2022). https://doi.org/10.1186/s12859-022-04944-z.*
   - *García S, A., Guizzardi, G., Pastor, O., Storey, V.C., Bernasconi, A. (2022).* ***An Ontological Characterization of a Conceptual Model of the Human Genome.*** *In: De Weerdt, J., Polyvyanyy, A. (eds) Intelligent Information Systems. CAiSE 2022. Lecture Notes in Business Information Processing, vol 452. Springer, Cham. https://-doi.org/10.1007/978-3-031-07481-3_4*
   - *García S., A., Bernasconi, A., Guizzardi, G., Pastor, O., Storey, V.C., Costa, M. (2022).* ***An Initial Empirical Assessment of an Ontological Model of the Human Genome.*** *In: Guizzardi, R., Neumayr, B. (eds) Advances in Conceptual Modeling. ER 2022. Lecture Notes in Computer Science, vol 13650. Springer, Cham. https://doi.org/10.1007/978-3-031-22036-4_6*
2. **The Delfos Platform**: A conceptual model-based platform to assist genome analysts in managing genomics information and help clinical experts rely more on their decisions. This platform has been published in:
   - *Bernasconi, A., García S., A., Ceri, S., Pastor, O. (2022).* ***A Comprehensive Approach for the Conceptual Modeling of Genomic Data.*** *In: Ralyté, J., Chakravarthy, S., Mohania, M., Jeusfeld, M.A., Karlapalem, K. (eds) Conceptual Modeling. ER 2022. Lecture Notes in Computer Science, vol 13607. Springer, Cham. https://doi.org/10.1007/978-3-031-17995-2_14*

- *García S., A.; Costa, M.; León, A.; Reyes, J. and Pastor, O. (2023).* ***Human-Centered Design for the Efficient Management of Smart Genomic Information****. In Proceedings of the 18th International Conference on Evaluation of Novel Approaches to Software Engineering, ISBN 978-989-758-647-7, ISSN 2184-4895, pages 15-26. DOI: 10.5220/0011635800003464*
- *A. León, A. García S., J. F. Reyes Román, M. Costa, and O. Pastor,* ***"The Delfos Platform: A Conceptual Model-Based Solution for the Enhancement of Precision Medicine"*** *[Manuscript submitted for publication]*

2.2. Validation

The second step is to validate Delfos in a real-world clinical context. The validation considers three use cases. All three use cases aim to improve the interpretation DNA variants to deliver a better diagnostic and treatment to patients in the following areas: cardiology, oncology, and inherited retinal diseases (IDR).

In first use case, we collaborated with two hospitals, namely, the Health Research Institute Hospital La Fe (IIS La Fe) and the Alicante Institute of Health and Biomedical Research (ISABIAL). The DNA sequences of 84 patients with some form of cardiomyopathy were analyzed, and the findings were discussed with domain experts.

In second use case, we collaborated with two hospitals, namely, the IIS La Fe and the Valencian Biomedical Research Institute (INCLIVA). The DNA sequences of 84 cancer patients were analyzed, and the findings were discussed with domain experts.

In the third use case, we are collaborating with a research group of the IIS La Fe specialized in rare diseases. First, the DNA sequences of 20 patients with some form IDR were analyzed. Like with the previous use cases, the findings were be discussed with domain experts. We are preparing an additional round of analysis at this moment with additional patients.

After finishing these three use cases, we will be able to assess whether the tool can improve patient diagnosis and treatment as well as if it is ready to be used in clinical practice on a regular basis.

In addition to these, we also validated the platform in a non-clinical context. We collaborated with the Valencian Institute of Agricultural Research (IVIA) in the agri-food field. The tool was used in a food improvement program to make citrus varieties more palatable and resistant to extreme weather conditions. Nearly 90 citrus varieties' DNA sequences were analyzed, and the results were discussed with domain experts.

The outputs of this step are the following:

1. **Validation of Delfos for cardiology**: Doctors agreed that the results were correct and reported very positive feedback. A portion of this work has been published in:
    - *Costa, M., García S., A., Pastor, O. (2022).* ***Conceptual Modeling-Based Cardiopathies Data Management.*** *In: Guizzardi, R., Neumayr, B. (eds) Advances in Conceptual Modeling. ER 2022. Lecture Notes in Computer Science, vol 13650. Springer, Cham. https://doi.org/10.1007/978-3-031-22036-4_2*
2. **Validation of Delfos for oncology**: Doctors agreed that the results were correct and reported very positive feedback. A portion of this work has been published in:

- *Costa, M., García S., A., Pastor, O. (2022).* ***A Comparative Analysis of the Completeness and Concordance of Data Sources with Cancer-Associated Information****. In: Guizzardi, R., Neumayr, B. (eds) Advances in Conceptual Modeling. ER 2022. Lecture Notes in Computer Science, vol 13650. Springer, Cham. https://doi.org/10.1007/978-3-031-22036-4_4*

3. **Validation of Delfos for IRD**: A first validation round of the tool in the context of IRD has been carried out, and we are preparing the second one. A portion of this work has been published in:
    - *Reyes Román, J.F., León Palacio, A., García S., A. et al.* ***Integration of clinical and genomic data to enhance precision medicine: a case of study applied to the retina-macula****. Softw Syst Model 22, 159–174 (2023). https://doi.org/10.1007/s10270-022-01039-4*
4. **Validation of Delfos for a non-clinical domain**: We assessed that Delfos can also be used in non-clinical contexts that are relevant. Genomics has the potential to revolutionize food production and improve food quality and safety, making it an important tool for food improvement programs. Domain experts reported very positive feedback towards the use of Delfos to achieve their goals. A portion of this work has been published in:
    - *García S., A. et al. (2023).* ***CitrusGenome: A Bioinformatics Tool to Characterize, Visualize, and Explore Large Citrus Variant Datasets****. In: , et al. Current Trends in Web Engineering. ICWE 2022. Communications in Computer and Information Science, vol 1668. Springer, Cham. https://doi.org/10.1007/978-3-031-25380-5_13*

2.3. Adoption

The third step is to connect with non-clinical companies that offer genetic counseling and could benefit from using the platform. The following actions have been carried out to improve the adoption of the tool:

1. We are in the process of creating a spin-off. We have already started to receive funding in order to define a business plan. We are collaborating with TRL+, an Spanish initiative committed to highlighting research projects in order to turn them into commercially and economically viable triple-impact spin-offs.
2. We are also working with Viromii, a consulting firm focused on innovation and tech-transfer, that supports the dissemination of the project results among the relevant institutions for the transfer of results.
3. The IP of the project has given talks in relevant national forums to disseminate the results of our project.

3. Relevance for CAiSE

This project is highly relevant to the topic of Cyber-Human Systems. This topic refers to the interface between humans and technology, and our project explicitly addresses this. Cyber-Human Systems aim to advance our ability to effectively work in complex environments, which

is also a key objective of our tool. This tool (i.e., Delfos) deals with the practical application of genomic data in clinical practice, which involves the interaction between humans and computer systems. It aims to reduce the impact of various genomics challenges when translating research findings into clinical practice. These challenges include dispersion, heterogeneity, discrepancies, lack of standardization, and data quality issues, all of which require a are relevant when considering the interaction between humans and technology.

Furthermore, we contribute to another a relevant topic to the CAiSE community, namely, conceptual modeling. Conceptual modeling is an essential aspect of Information Systems Engineering, as it provides a way to represent and reason about complex systems in a structured way. Using conceptual models allows for a shared understanding among stakeholders, which is essential for successfully designing and implementing information systems. Model-based solutions have gained significant attention in the field of Information Systems Engineering due to their ability to capture complex relationships and interactions between various elements of a system, which can be particularly useful in the healthcare domain. This submission contributes to the field of conceptual modeling by proposing a model-based platform developed with a rigorous methodological and ontological foundation for the practical application of genomic data in clinical practice.

In addition, this project highlights the application of information systems engineering in the healthcare domain, a significant area of research and development; this aligns particularly well with the conference's emphasis on innovative and rigorous research. As mentioned above, the main goal of the work herein presented is to minimize the impact of unresolved genomics issues when transferring research findings to clinical practice. This is a significant and timely challenge, as the correct management of the vast amount of genomic data generated by current sequencing technologies is a bottleneck for advancing Precision Medicine.

Finally, this project has the potential to stimulate further discussion in the CAiSE community about the use of Advanced Information Systems Engineering in healthcare and Precision Medicine, particularly in the context of model-based solutions.

4. Current Project Status

Based on the three steps defined in Section 2, we indicate the current progress of the project (see Table 1). The design and implementation phases of the model-based tool have been completed, and it operates as expected. Two of the three use cases have concluded with promising results; the third is in the planning phase and will begin in a few months. To date, we have contacted four potential customers for our tool: the three hospitals with which we collaborated on the use cases, as well as BIONOS, the company that performed the patient sequencing process.

Acknowledgments

We wish to thank the IIS La Fe, INCLIVA, ISABIAL, and BIONOS for their huge contribution to this project and their willing to help. This work was supported by the Spanish State Research Agency through the PDC2021-121243-I00 grant (MICIN/AEI/10.13039/501 100011033) and co-financed with ERDF and the European Union Next Generation EU/PRTR.

Table 1
Current status of the project

Step	Status	Comments
Design and development of the tool	Completed (100%)	A first version of the tool has been created.
Validation of the tool	In progress (66%)	The cardiology and oncology use cases have been validated. The IDR use case is on going.
Adoption of the tool	In progress (50%)	We have started the creation of the spin-off and are currently contacting potential partners.

References

[1] R. Pereira, J. Oliveira, M. Sousa, Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics, Journal of Clinical Medicine 9 (2020) 132. doi:10.3390/jcm9010132.

[2] B. M. Good, B. J. Ainscough, J. F. McMichael, A. I. Su, O. L. Griffith, Organizing knowledge to enable personalization of medicine in cancer, Genome Biology 15 (2014) 438. doi:10.1186/s13059-014-0438-7.

[3] J. White, Pubmed 2.0, Medical Reference Services Quarterly 39 (2020) 382–387. doi:10.1080/02763869.2020.1826228, pMID: 33085945.

[4] N. O'Leary, M. Wright, J. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, K. Pruitt, Reference sequence (refseq) database at ncbi: current status, taxonomic expansion, and functional annotation, Nucleic Acids Research 44 (2015). doi:10.1093/nar/gkv1189.

[5] S. Sherry, M. Ward, K. Sirotkin, dbsnp-database for single nucleotide polymorphisms and other classes of minor genetic variation, Genome research 9 (1999) 677—679. URL: http://www.genome.org/cgi/content/full/9/8/677.

[6] M. J. Landrum, J. M. Lee, M. Benson, G. R. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, W. Jang, K. Karapetyan, K. Katz, C. Liu, Z. Maddipatla, A. Malheiro, K. McDaniel, M. Ovetsky, G. Riley, G. Zhou, J. B. Holmes, B. L. Kattman, D. R. Maglott, ClinVar: improving access to variant interpretations and supporting evidence, Nucleic Acids Research 46 (2018) D1062–D1067. doi:10.1093/nar/gkx1153.

[7] C. Kopanos, V. Tsiolkas, A. Kouris, C. E. Chapple, M. Albarca Aguilera, R. Meyer, A. Massouras, VarSome: the human genomic variant search engine, Bioinformatics 35 (2019) 1978–1980. doi:10.1093/bioinformatics/bty897.

[8] L. M. Chunn, D. C. Nefcy, R. W. Scouten, R. P. Tarpey, G. Chauhan, M. S. Lim, K. S. Elenitoba-Johnson, S. A. Schwartz, M. J. Kiel, Mastermind: A Comprehensive Genomic Association Search Engine for Empirical Evidence Curation and Genetic Variant Interpretation, Frontiers in Genetics 11 (2020). doi:10.3389/fgene.2020.577152.

[9] T. Brandt, L. M. Sack, D. Arjona, D. Tan, H. Mei, H. Cui, H. Gao, L. J. Bean, A. Ankala, D. Del Gaudio, A. Knight Johnson, L. M. Vincent, C. Reavey, A. Lai, G. Richard, J. M. Meck, Adapting ACMG/AMP sequence variant classification guidelines for single-gene copy number

variants, Genetics in Medicine 22 (2020) 336–344. doi:10.1038/s41436-019-0655-2.

[10] A. León, A. García S., M. Costa, A. Vañó Ribelles, O. Pastor, Evolution of an adaptive information system for precision medicine, in: S. Nurcan, A. Korthaus (Eds.), Intelligent Information Systems, Lecture Notes in Business Information Processing, Springer International Publishing, 2021, pp. 3–10. doi:10.1007/978-3-030-79108-7_1.

[11] A. León, J. Reyes, V. Burriel, F. Valverde, Data Quality Problems When Integrating Genomic Information, in: S. Link, J. C. Trujillo (Eds.), ER 2016 Workshops, LNCS, Springer International Publishing, Gifu, Japan, 2016, pp. 173–182. doi:10.1007/978-3-319-47717-6_15.