UNIVERSITAT POLITÈCNICA DE VALÈNCIA

VRAIN
Valencian Research Institute for Artificial Intelligence

PROS
Centro de Investigación en Métodos de Producción de Software

# Software Engineering for Life Engineering: Deciphering the Language of Life

Almati, Kazajistan - December 6, 2023

*Oscar Pastor*
*opastor@pros.upv.es*

PROS
Centro de Investigación en Métodos de Producción de Software

# Acknowledgements

What it is presented here is the result of years of work with an amazing set of brilliant researchers now, formerly PhD students that I was happy and honored to supervise. Nothing would have occurred without them. My full gratitude and recognition to all of them.

Ignacio Panach, Sergio España, Paco Valverde, Nelly C. Fernandez, Nathalie Aquino, Beatriz Marín, Giovanni Giachetti, José Luis de la Vara, Marcela Ruiz, María José Villanueva, Verónica Burriel, José Reyes, Ana León, Alberto García, Mireia Costa.

VRAIN
Valencian Research Institute for Artificial Intelligence

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

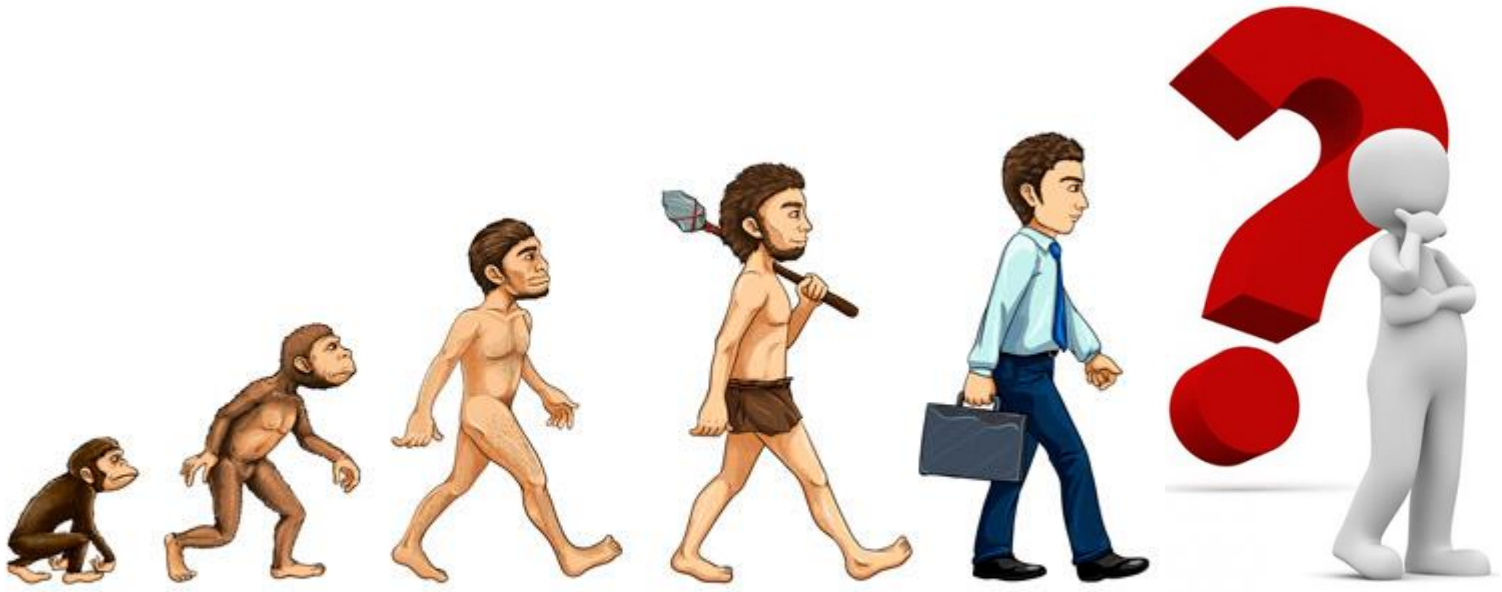
# AGENDA

1. Introduction
2. Problem under Investigation
   - Understanding the Genome; From Genome to Reality; The SE/ISE perspective/ The Genome Data Chaos
3. Treatment Design
   - The SILE Method; the DELFOS platform
4. Validation
   - Running Projects and Practical Experiences
5. Conclusions

# AGENDA

1. Introduction
2. Problem under Investigation
   - Understanding the Genome; From Genome to Reality; The SE/ISE perspective/ The Genome Data Chaos
3. Treatment Design
   - The SILE Method; the DELFOS platform
4. Validation
   - Running Projects and Practical Experiences
5. Conclusions

# Introduction



Australopithecus | Homo habilis | Homo erectus | Homo sapiens neanderthalensis | Homo sapiens sapiens | **¿Homo genius?**

# Introduction

## Problem Statement

Why "from Homo Sapiens to Homo Genius"?

✓ Capability of understanding and manipulating the Genome

# Introduction

## Problem Statement

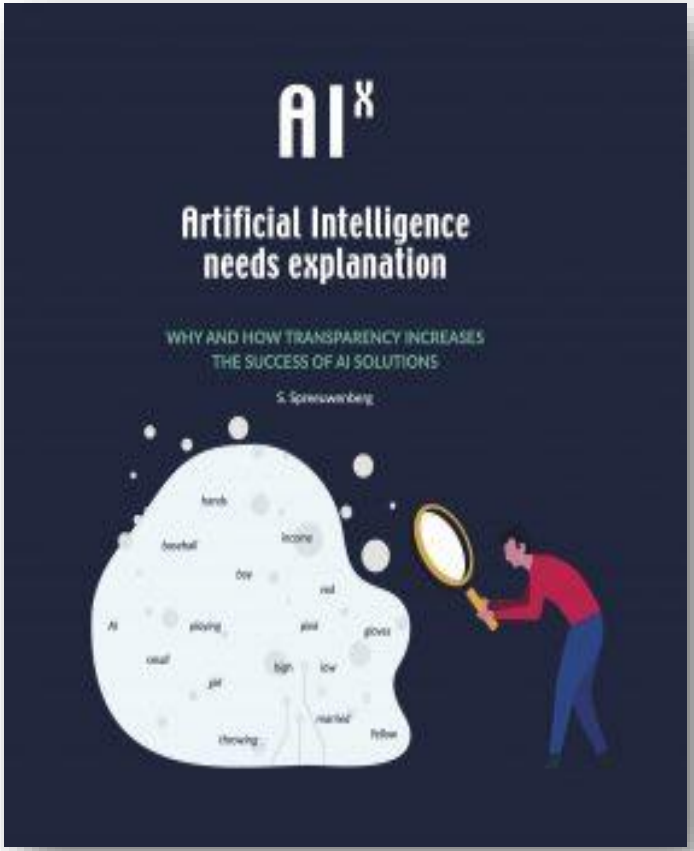

# Introduction
## Deep Learning

This book is about a solution to (...) intuitive problems (...). This solution is to allow computers to learn from experience and understand the world in terms of a hierarchy of concepts, with each concept defined through its relation to simpler concepts...

The hierarchy of concepts enables the computer to learn complicated concepts by building them out of simpler ones...

(Goodfellow et al., 2020)

# Introduction
## Explainable AI




1. Get a shared understanding of the domain
2. Understand the task and select the right scope
3. Collect the right data and improve its quality
4. Select AI techniques that deliver results
5. Generate good explanations
6. Evolve the system over time



Explainable AI is feasible using a six-step approach
(S. Spreeuwenberg)

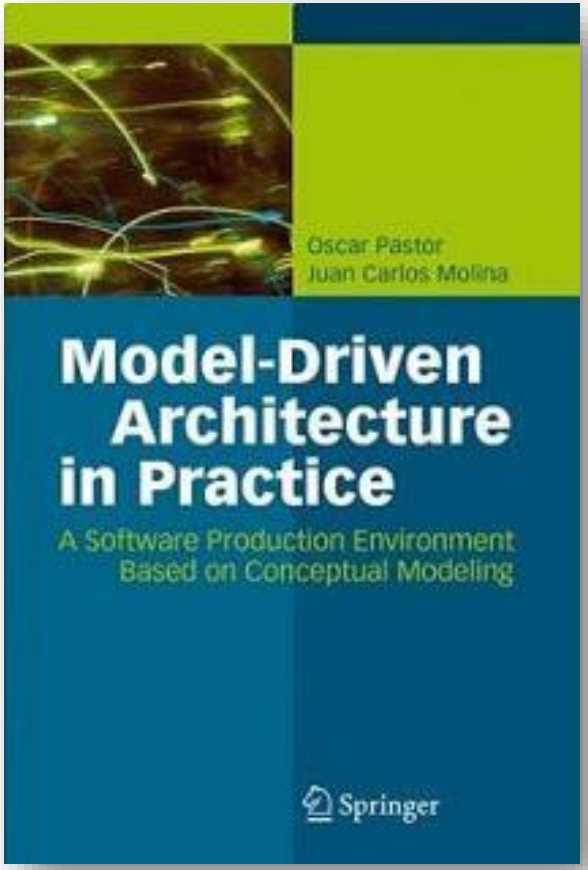
# Thinking Fast and Slow in Software Engineering

Giancarlo Guizzardi, Oscar Pastor, and Veda C. Storey

**THINK OF A** child learning how to catch a ball repeatedly thrown to her by her father. As the child practices or continues with this activity, she becomes better at it. Through a process of trial and error and across several attempts, the child, in essence, is gathering more data on what works well and what does not work and, in this manner, mapping what she learns to the outcome (ball catching).

# Introduction

## Problem Statement

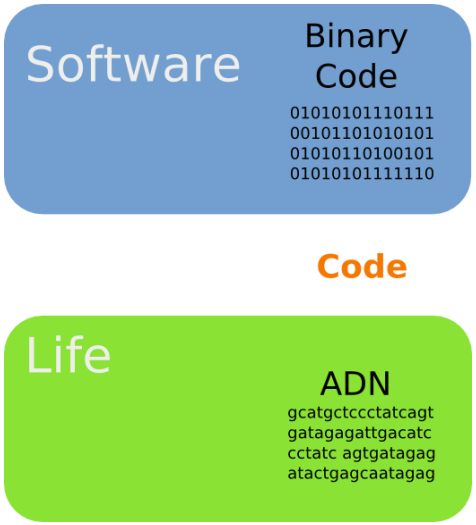



- We have been building
  - Traditional Information Systems
  - Web-based Information Systems
  - SOA-based systems
  - Pervasive Systems

  - ... but, what is next?

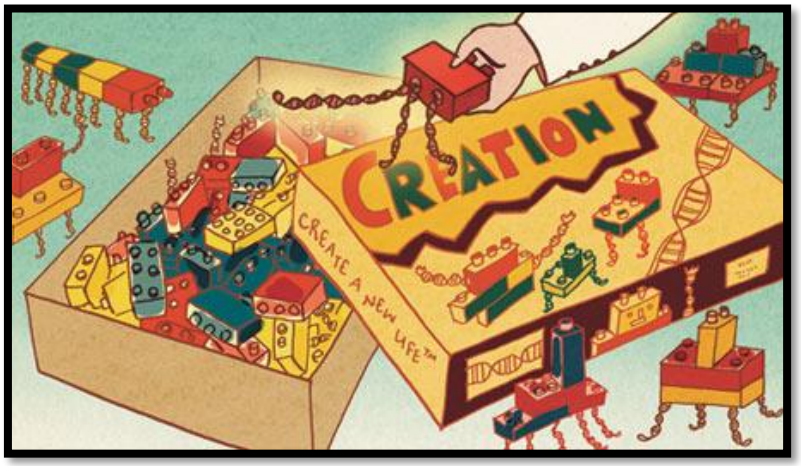
# Introduction
## A parallelism…

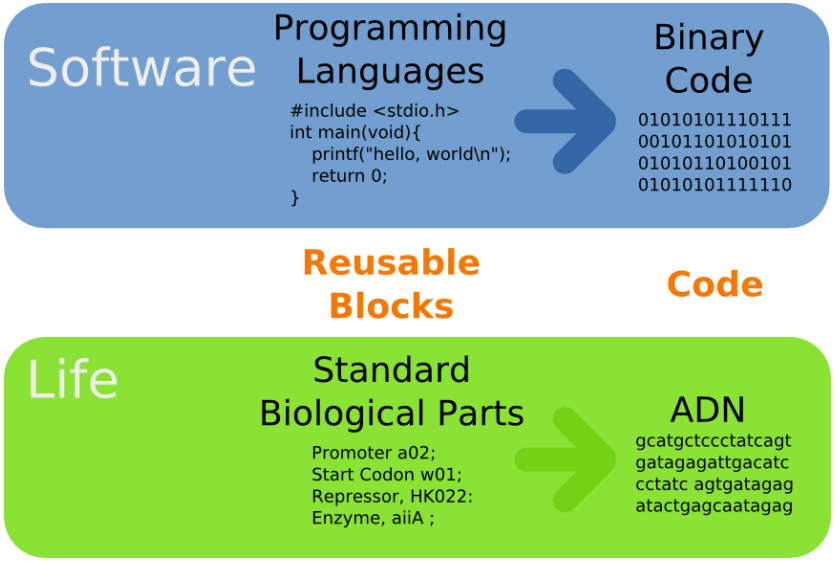
"A living organism is a **computer** or machine made up of genetic circuits in which DNA is the **software** that can be hacked." — *Drew Endy, MIT*



# Introduction
## First step: Assembling

- First abstraction step
  - Standard Biological Parts

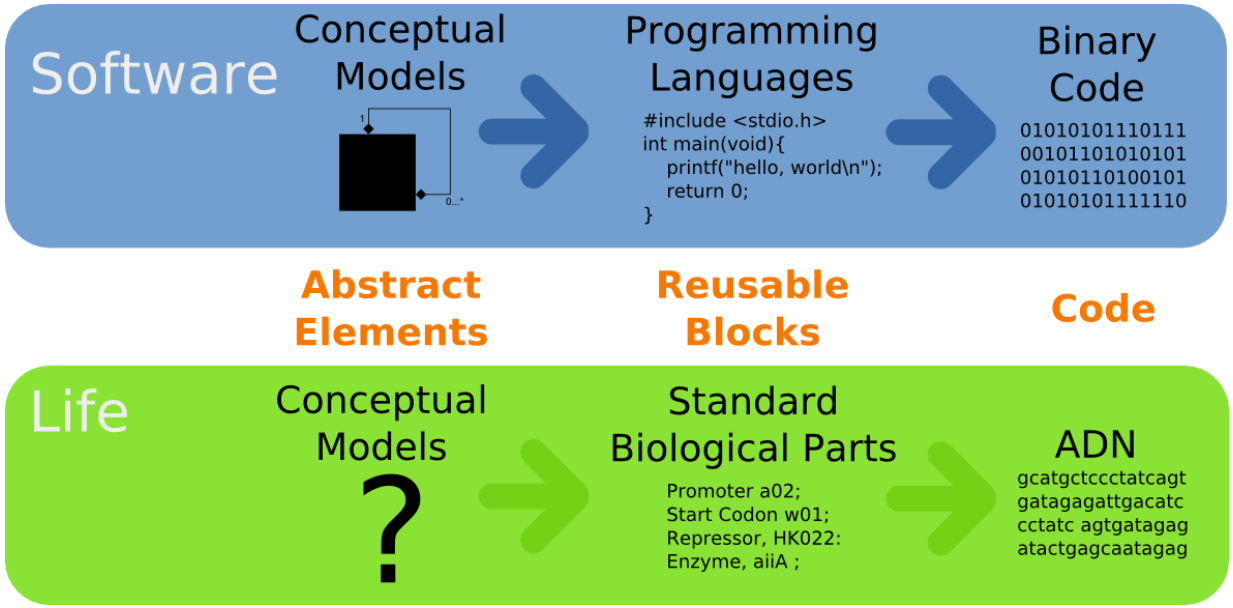


# Introduction
## Going further…

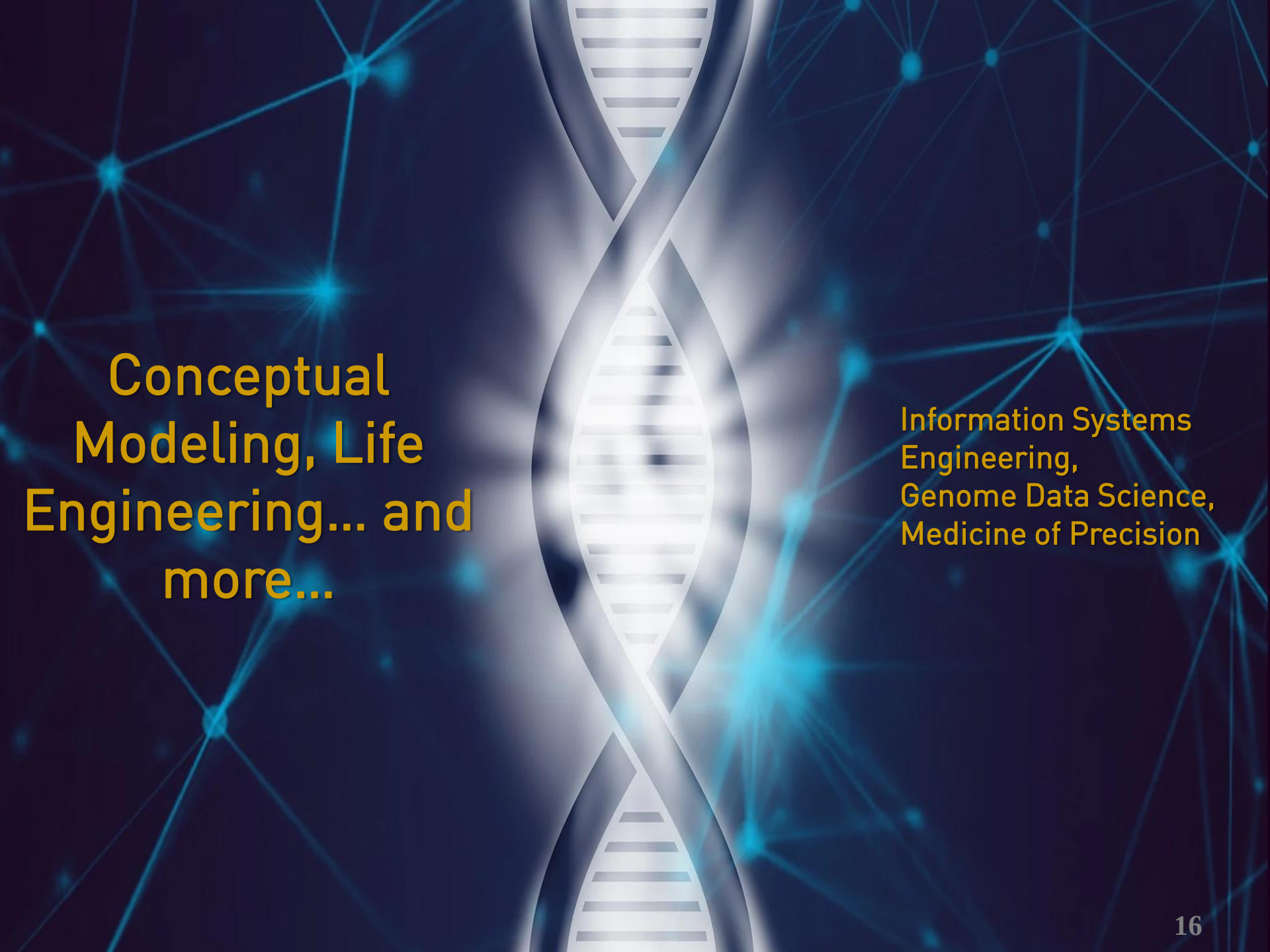
- Next step: Modeling

  Conceptual models are needed for a systematic development of biological systems



# AGENDA

1. Introduction
2. **Problem under Investigation**
   - Understanding the Genome; From Genome to Reality; The SE/ISE perspective/ The Genome Data Chaos
3. Treatment Design
   - The SILE Method; the DELFOS platform
4. Validation
   - Running Projects and Practical Experiences
5. Conclusions


Conceptual Modeling, Life Engineering… and more…
Information Systems Engineering, Genome Data Science, Medicine of Precision

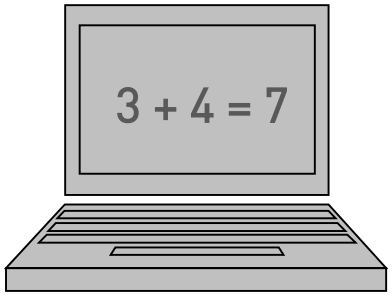
# From Genome To Reality…

| 00010011 | 00000111 | 00000011 | 00001000 |
|---|---|---|---|

*Physical Level*

ADD $7 $3 $8 *Instruction Level*

*Semantics: Add the values from the processor registers '3' and store the result in the register '8'*




*Representation Level*

# From Genome To Reality…

| AUG | GAA | CAC | GAC | GAG | UAA |
|---|---|---|---|---|---|

*Physical Level*

START Glu His Asp Glu STOP

*Instruction Level*

*Semantics: Process a protein with the four selected aminoacids*

However, ¿Why?

*Representation Level*

PROS
Centro de Investigación en Métodos de Producción de Software

# The Genome Project

Genetic Sample

Next Generation Sequencing

TECHNOLOGIES

- SOLiD (Life Tech)
- 454 (Roche)
- Illumina

BIOINFORMATICS

Alignment

BWA, BowTie, Blat

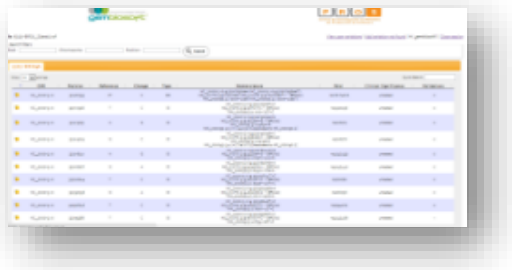
FASTQ

SAM

Design of Genomic Information Systems

ECGH

genes.me
www.geneslove.me



VRAIN
Valencian Research Institute for Artificial Intelligence

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

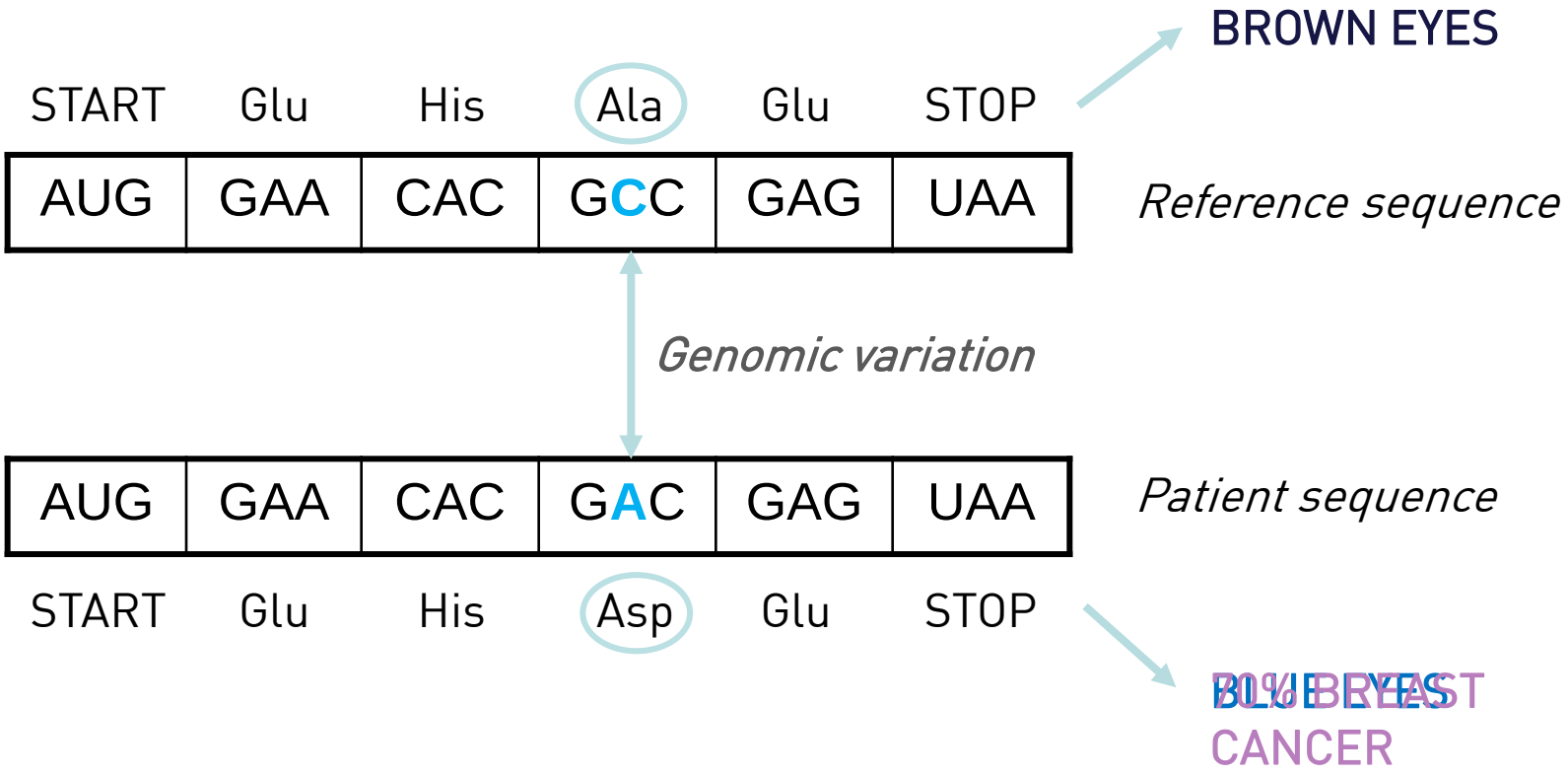
# The Genome Project

## Our Vision...

## One Platform to Rule Them All

# IS and Bioinformatics

## How to understand genomic code?



Centro de Investigación en Métodos de Producción de Software

# IS and Bioinformatics

## Manual Data Analysis Methods

**Tedious and repetitive**

**No explicit methods**


**Human error**

**Navigating through hyperlinks**

# Conceptual Modeling and Bioinformatics

## Genome Data Chaos!

# The Genomic Data Chaos

Lack of a clear ontological agreement to define the key concepts of the field.

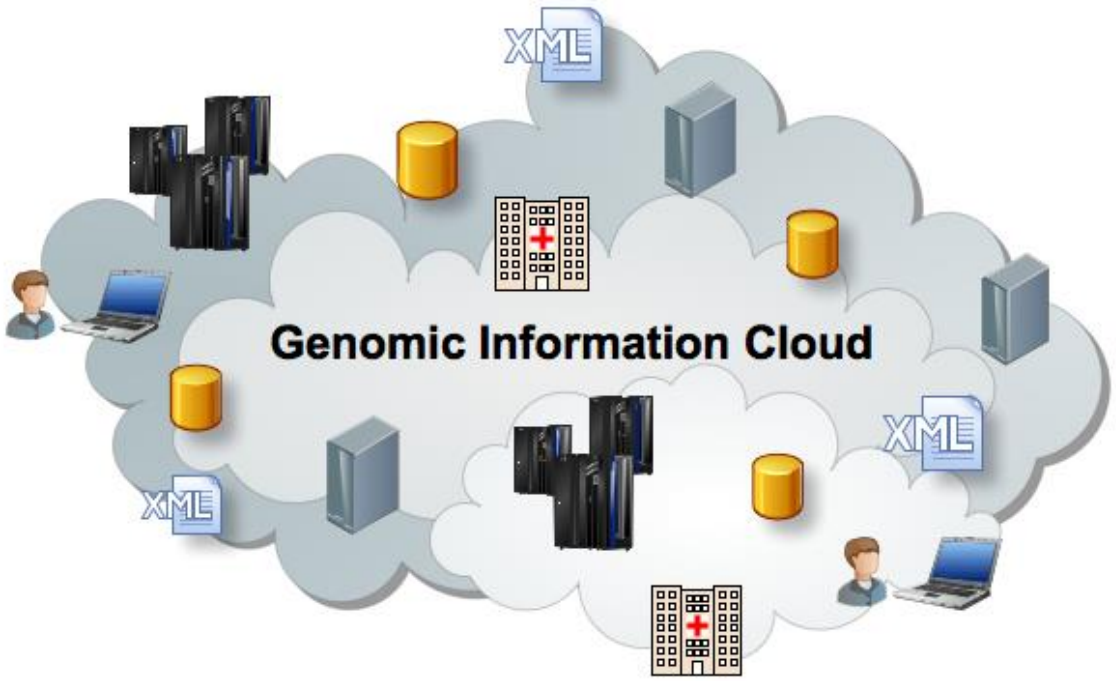
Dispersion of the genomic information


A great variability in the quality of the available information.

# AGENDA

1. Introduction
2. Problem under Investigation
   - Understanding the Genome; From Genome to Reality; The SE/ISE perspective/ The Genome Data Chaos
3. **Treatment Design**
   - The SILE Method; the DELFOS platform
4. Validation
   - Running Projects and Practical Experiences
5. Conclusions

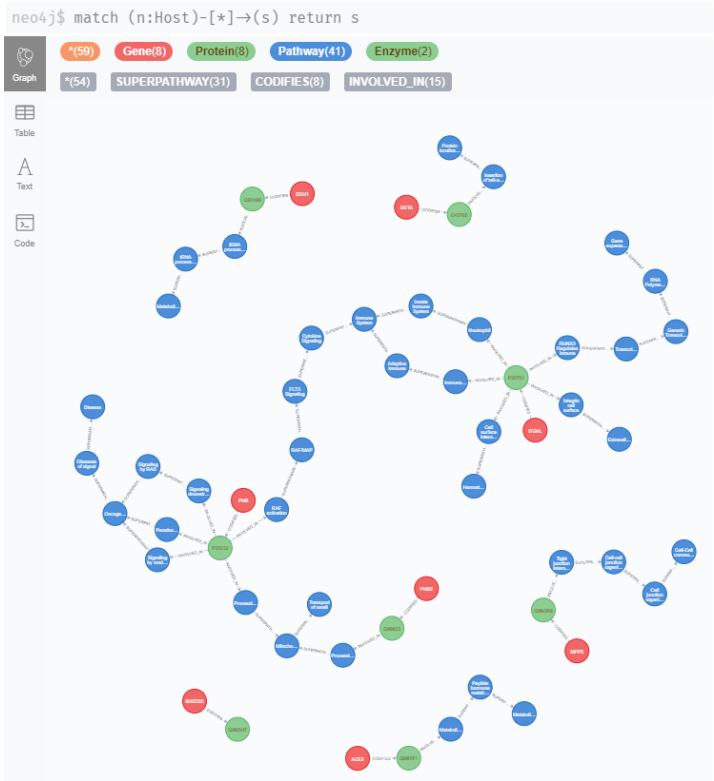
# Fighting the Genome Data Chaos!



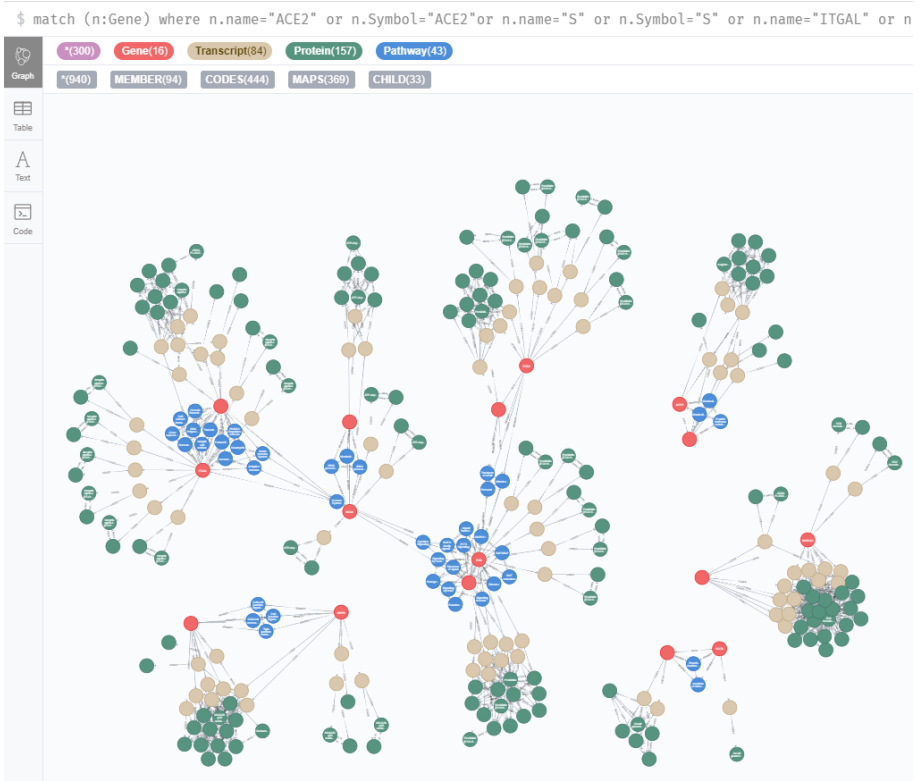

l Modeling

Centro de Investigación en Métodos de Producción de Software

# CM Benefits




CovProt




CovidGraph

VRAIN
Valencian Research Institute for Artificial Intelligence

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

# The SILE Method:
## Towards a Genome DELFOS Oracle

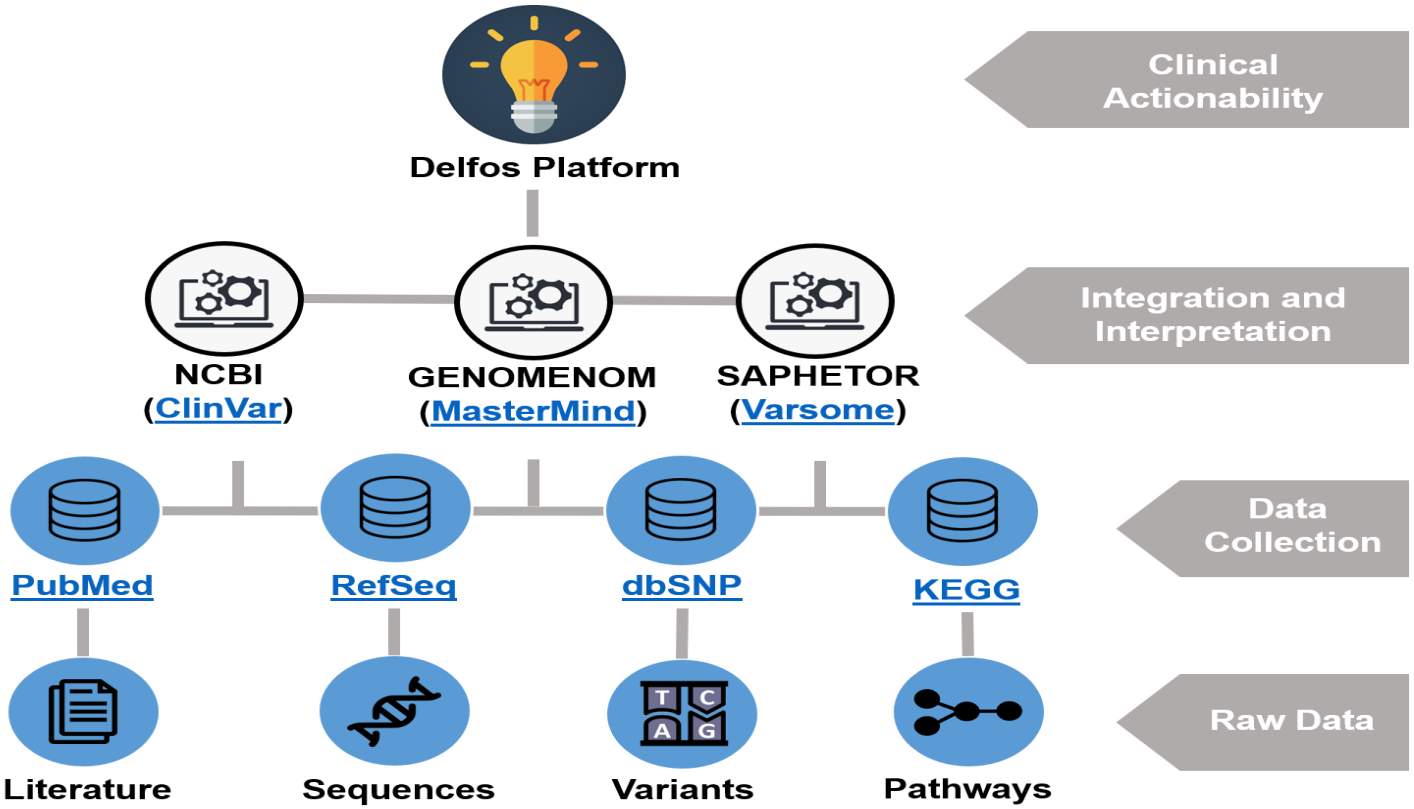

Search and selection of the adequate data sources to extract information from


Identification of the relevant information to satisfy a knowledge requirement

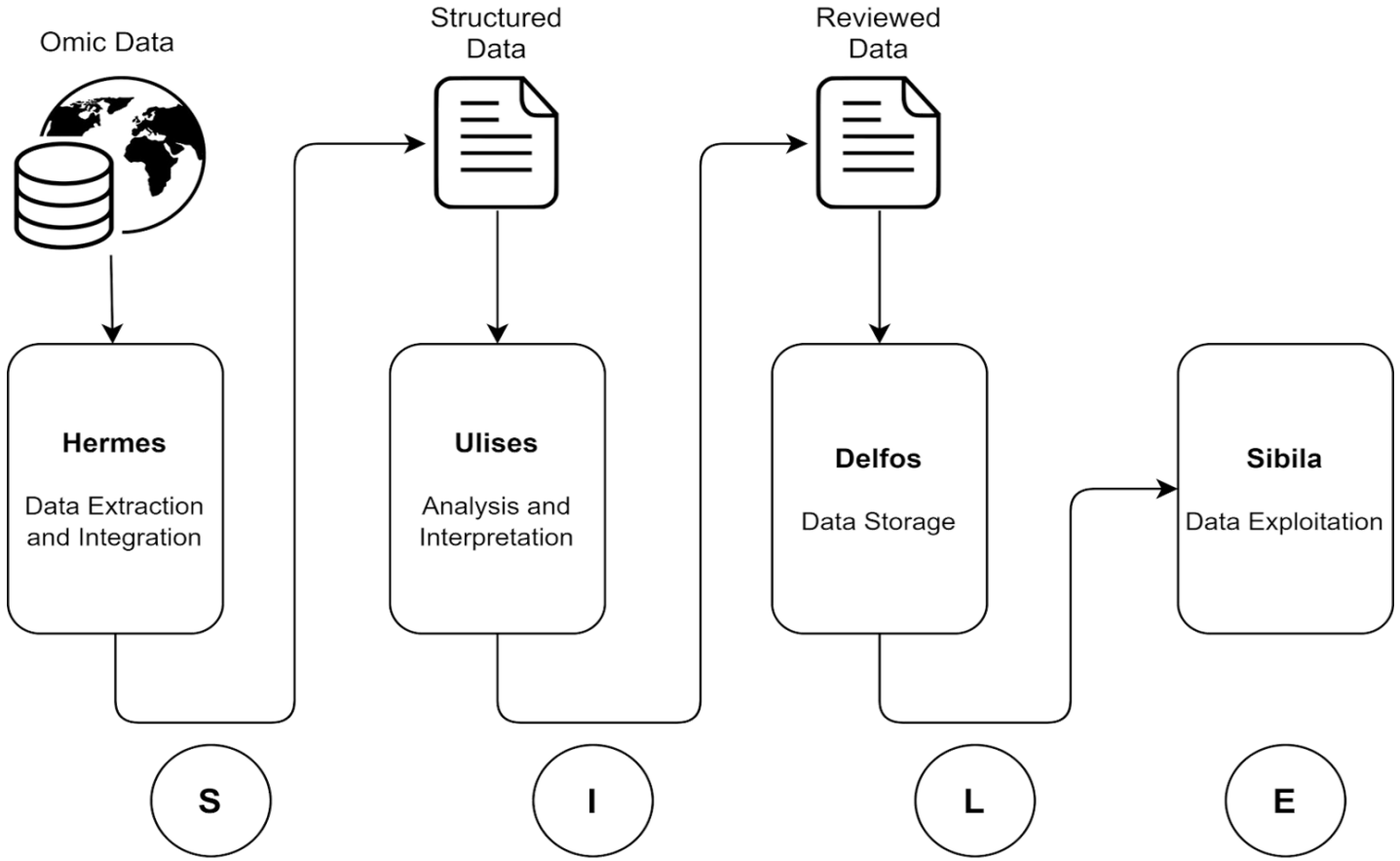

Load the information into a database for its further analysis and exploitation


Extraction of knowledge from the database by using specific tools to analyze and interpret genomic data

# The Delfos Platform

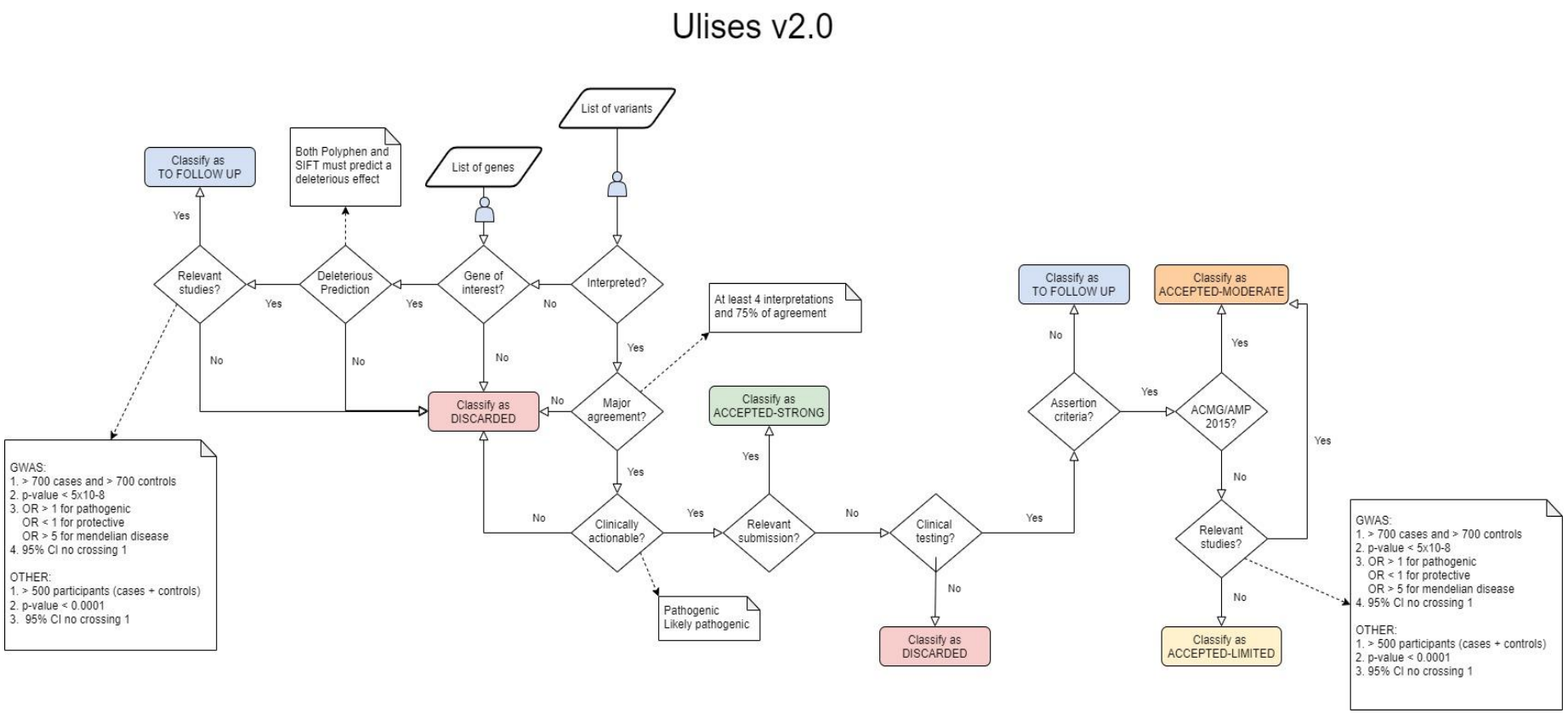


# Delfos Scope

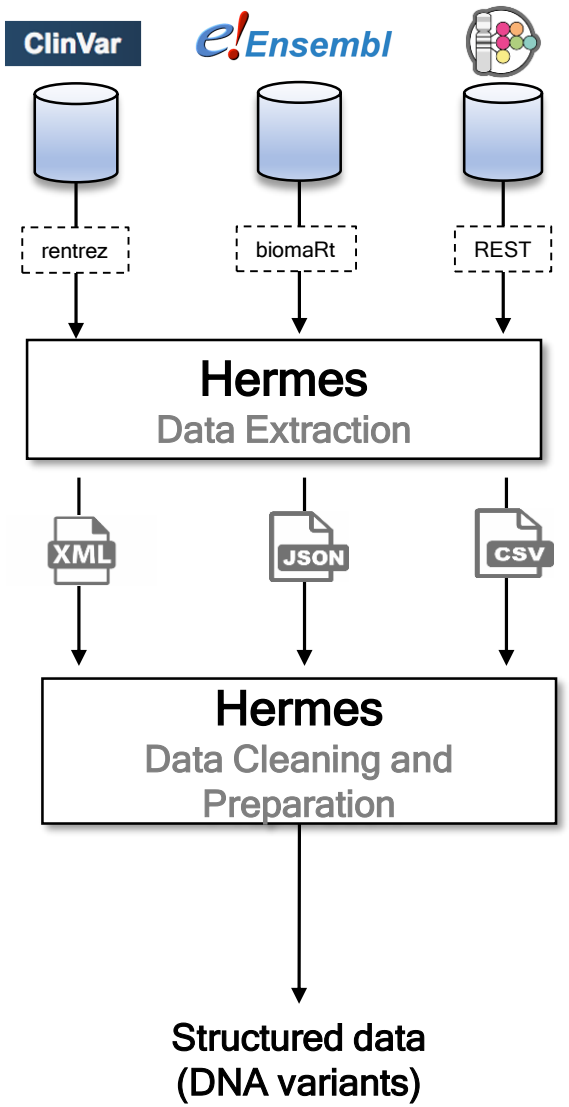


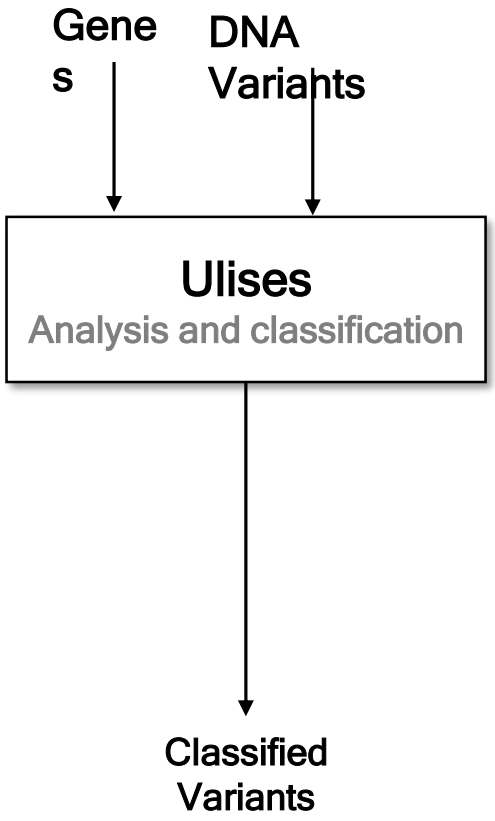

Conceptual Schema of the Human Genome

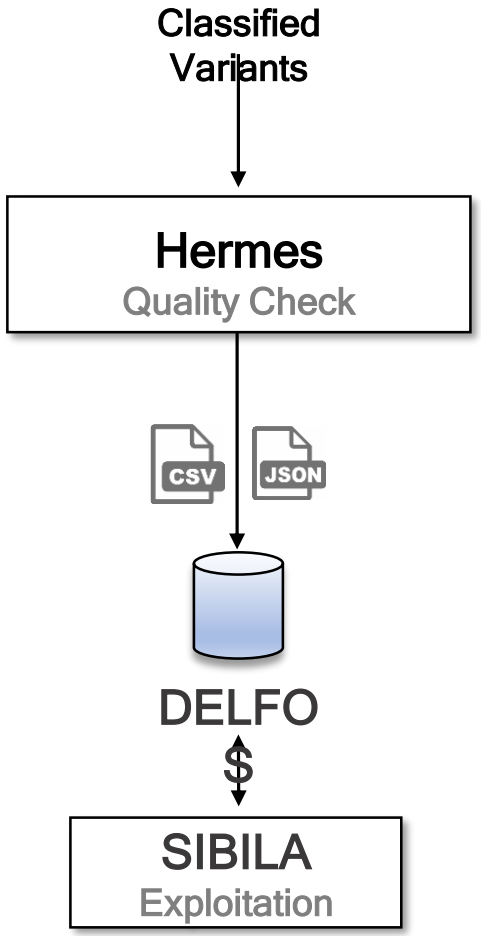
# Delfos Scope
## Ulises



# Current State



# Are you sure your data is always reliable?

Don't run the risk of missing or misidentifying critical details in your data

# The Opportunity

"Genomic medicine has opened the possibility to transform health and wellness of people around the world with life-changing diagnosis and treatments that were previously impossible".

# The Challenge

... To have a breakthrough technology that helps scan all the available repositories and evidence, automatically connecting, integrating and interpreting complex genomic data, and reducing analysis average times from hours to minutes.

# The Delfos Platform

Delfos "enables end-to-end automated clinical decision support for rapid interpretation of NGS data to make a huge difference to people's lives" by providing a holistic conceptual characterization of the domain intended to integrate diverse "omics" dimensions.

# The Delfos Platform

**Here's what Delfos is intended to provide:**

- The world's largest resource for finding disease-causing mutations
- Automatically assessed data derived from the most well-known and reliable data sources and evidence
- Translation of genomic data into clinically actionable insights with complete traceability of the decisions made by our XAI algorithms
- Up-to-date content and functionalities to ensure you remain informed on the latest findings

# AGENDA

1. Introduction
2. Problem under Investigation
   - Understanding the Genome; From Genome to Reality; The SE/ISE perspective/ The Genome Data Chaos
3. Treatment Design
   - The SILE Method; the DELFOS platform
4. **Validation**
   - Running Projects and Practical Experiences
5. Conclusions

PROS
Centro de Investigación en Métodos de Producción de Software

# Projects: DataME, DELFOS, OGMIOS, SREC and CARDIOVAL

**OGMIOS**. Sistema Inteligente de apoyo a la toma de decisiones clínicas en medicina de precisión.
*Proyectos estratégicos en cooperación. Convocatoria 2021*

**DELFOS**. Plataforma Delfos: Sistema de Información para la gestión de variaciones genómicas.
*Convocatoria 2021 - «Proyectos Pruebas de Concepto»*

**SREC.** Desarrollo ágil de sistemas desde requisitos a código.
*Convocatoria AEI -«Proyectos de Generación de Conocimiento»*

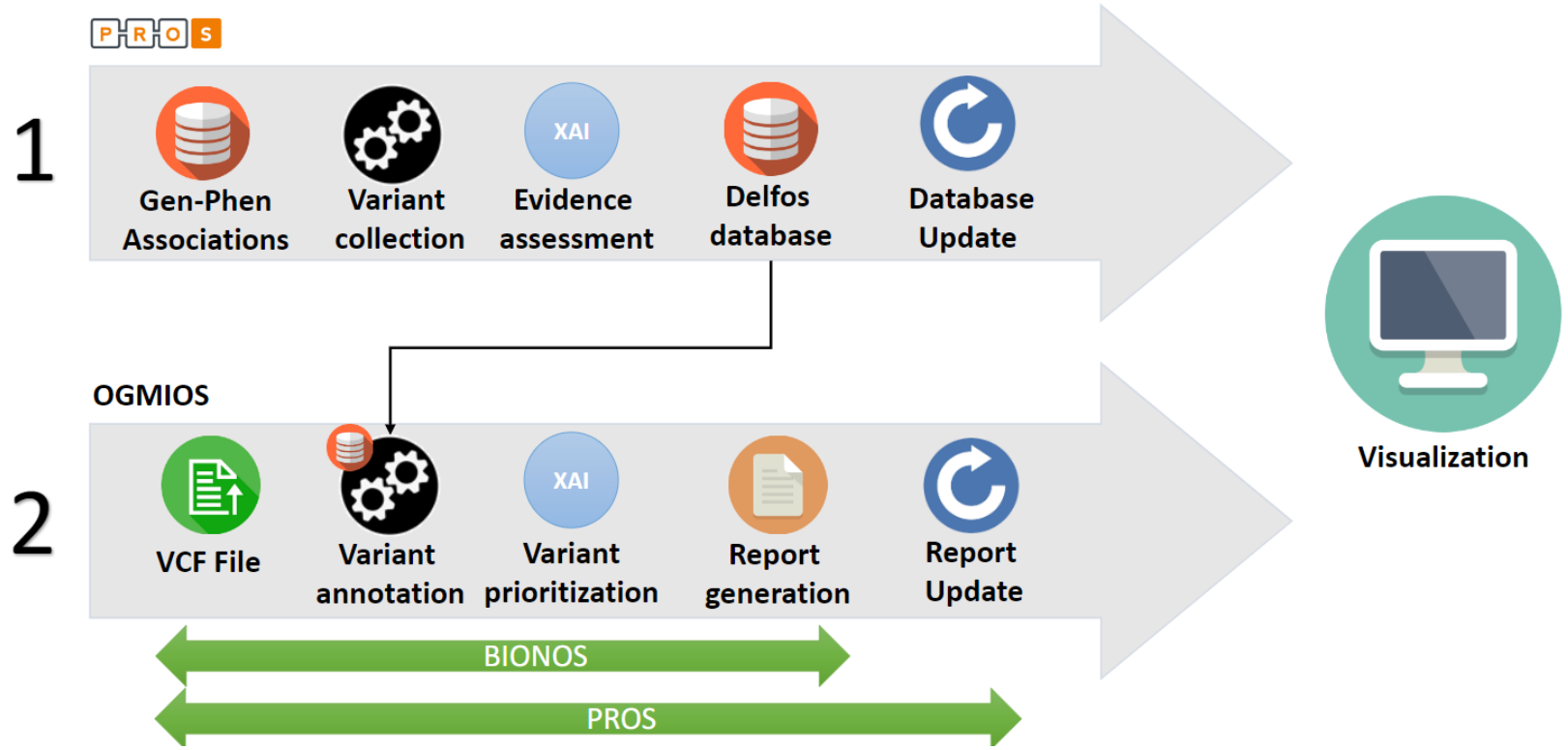
**DataME.** Un Método de producción de software dirigido por modelos para el desarrollo de aplicaciones Big Data. *Convocatorias 2016 - Proyectos EXCELENCIA y Proyectos RETOS* **(FINALIZADO)**

**CardioVAL.** Diseño y desarrollo de un prototipo basado en Inteligencia Artificial Explicable para la gestión de información genética relacionada con el riesgo de sufrir muerte súbita de origen cardiaco.
*PROGRAMA INBIO 2021, SUBPROGRAMA DE FOMENTO DE ACCIONES PREPARATORIAS (AP)* **(FINALIZADO)**

VRAIN
Valencian Research Institute for Artificial Intelligence

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

# Real use case in OGMIOS



# Case Study: Familial Heart Diseases

Group of cardiovascular diseases that have a **genetic basis**, a familial presentation, and that can be related with sudden death.

Three groups:

- Cardiomyopathies
- Channelopathies
- Aortic disease

*Catecholaminergic polymorphic ventricular tachycardia

| Cardiomyopathies | Channelopathies | Aortic disease |
|---|---|---|
| Hypertrophic | Long QT | Marfan |
| Dilated | Short QT | Loeys-Dietz |
| Restrictive | Brugada | |
| Noncompaction | CPVT* | |
| Arrhythmogenic | | |



# Conceptual Modeling-based Cardiopathies Data Management

Conceptual Modeling for Life Science (CMLS)@ER2022
Oct 17, 2022

Mireia Costa micossan@vrain.upv.es
**Alberto García S.** algarsi3@pros.upv.es
Oscar Pastor opastor@pros.upv.es

UNIVERSITAT POLITÈCNICA DE VALÈNCIA
VRAIN Valencian Research Institute for Artificial Intelligence
PROS Centro de Investigación en Métodos de Producción de Software

# Practical Experiences

# A Comparative analysis of the completeness and concordance of data sources with cancer-associated information

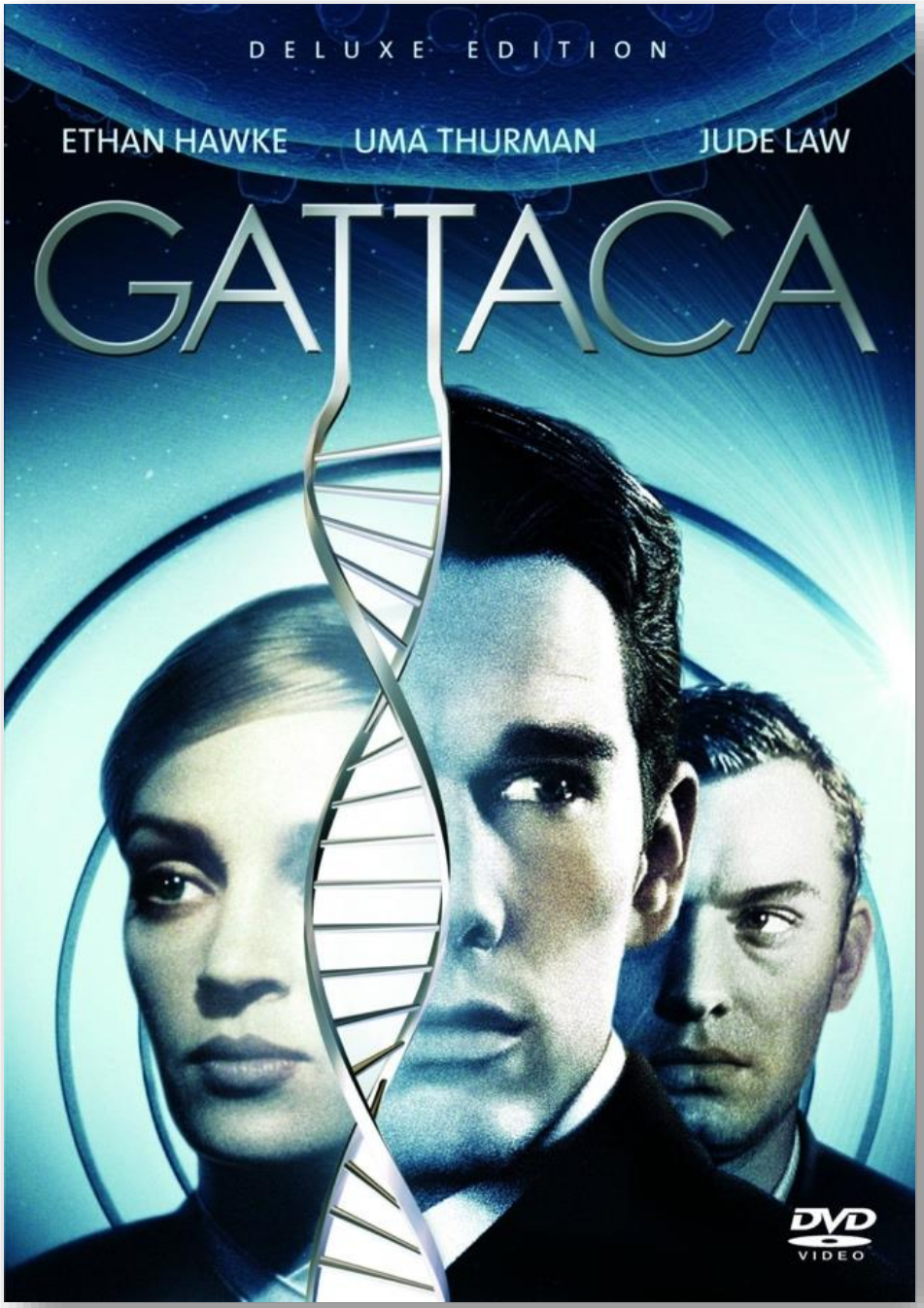
Conceptual Modeling for Life Science (CMLS)@ER2022
Oct 17, 2022

**Mireia Costa** micossan@vrain.upv.es
Alberto García S. algarsi3@pros.upv.es
Oscar Pastor opastor@pros.upv.es

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

VRAIN
Valencian Research Institute for Artificial Intelligence

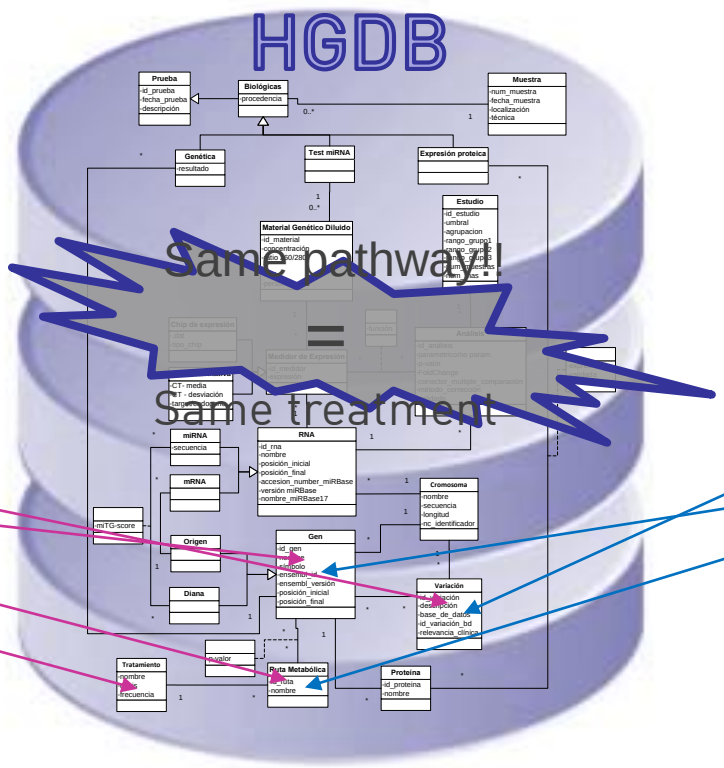
PROS
Centro de Investigación en Métodos de Producción de Software


DELUXE EDITION
ETHAN HAWKE
UMA THURMAN
JUDE LAW
GATTACA
DVD
VIDEO


Friendship is the ultimate upgrade
A NETFLIX FILM
NEXT GEN
7 SEPT | NETFLIX

# Conceptual Model of the Human Genome

possible scenarios of use...



- Variation
- Gene
- Pathway
- Treatment

- Variation
- Gene
- Pathway

# AGENDA

1. Introduction
2. Problem under Investigation
   - Understanding the Genome; From Genome to Reality; The SE/ISE perspective/ The Genome Data Chaos
3. Treatment Design
   - The SILE Method; the DELFOS platform
4. Validation
   - Running Projects and Practical Experiences
5. **Conclusions**

# Conclusions

- "We" have to be active and essential actors in the immense challenge of understanding life, that is leading to a Medicine of Precision revolution.
- Inventors have long dreamed of creating machines that think… (Goodfellow et als., Deep Learning, 2020)
- It all starts by agreeing on the meaning of the important concepts in your domain… (Spreeuwenberg,S., Artificial Intelligence needs explanation, 2020)

# Conclusions

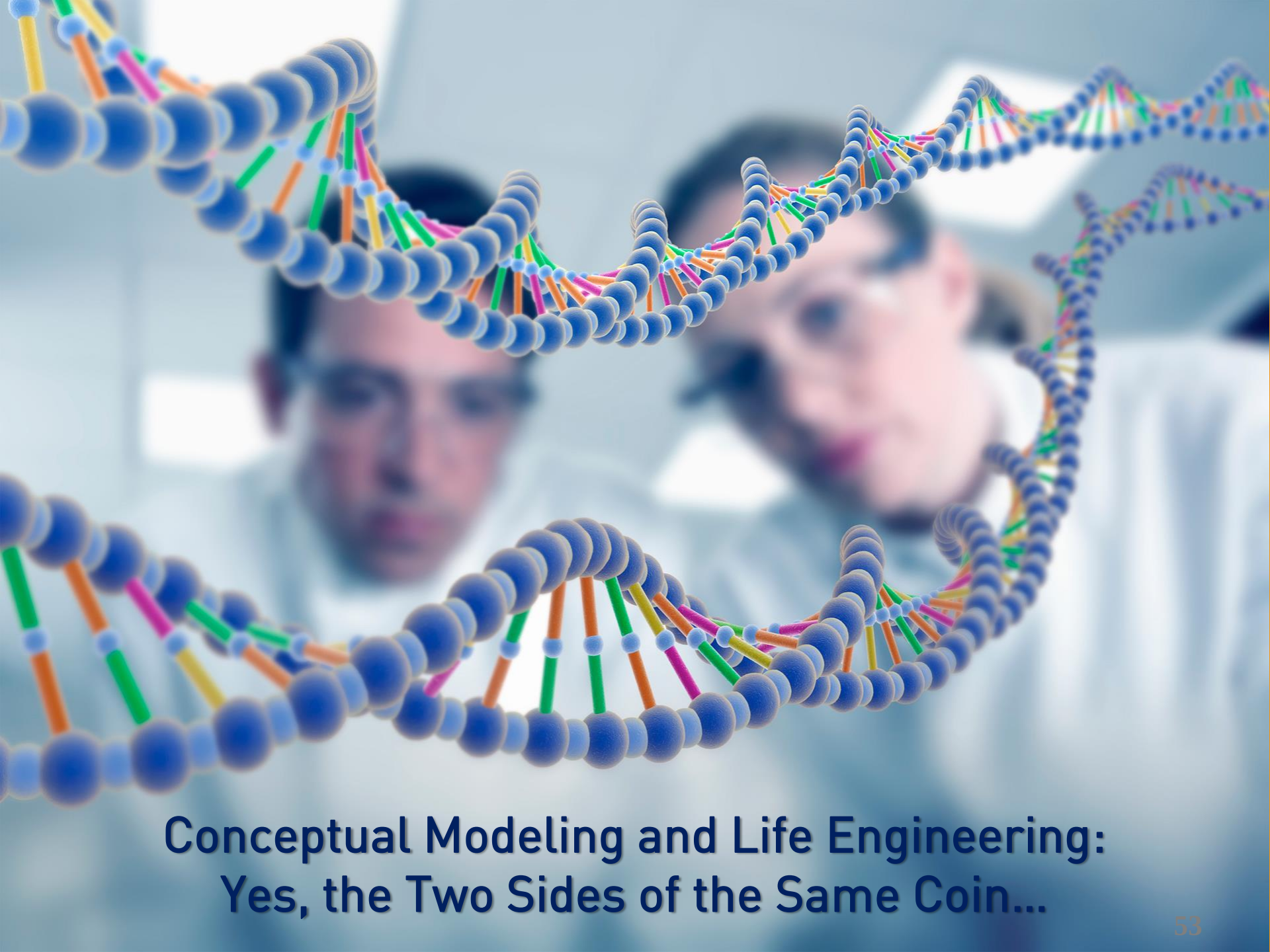
- The Genomic Data Chaos requires methodological solutions to manage genomic data
- The SILE Method provides solutions to the main bottlenecks of genomic data management:
  - Search and selection of data sources
  - Identification of relevant data
  - Load in the adequate repository
  - Tools for data exploitation
- The Delfos Platform is a specific implementation for the use of the SILE Method in Precision Medicine

# Current Work

- Processing of VCF files and automation of genetic reports
- Application of the SILE method to other biological domains: proteomics, pathways, pharmacogenomics…
- Automatic extraction of metadata from the literature using AI Techniques (Natural Language Processing)
- Automation of the ACMG/AMP guidelines for determining the clinical impact of DNA variants.
- Connection of clinical and genomic data

# References

1. **Pastor, O.**, & Molina, J. C. (2007). Model-driven architecture in practice: a software production environment based on conceptual modeling (Vol. 1). New York: Springer.
2. **Pastor, O.** (2016, November). Conceptual modeling of life: beyond the homo sapiens. In International Conference on Conceptual Modeling (pp. 18-31). Springer, Cham.
3. **Pastor, Ó.**, León, A., Reyes Román, J. F., García, A. S., & Casamayor, J. C. R. (2020). Using conceptual modeling to improve genome data management. Briefings in Bioinformatics. DOI: 10.1093/bib/bbaa100
4. Reyes Román, J. F., León Palacio, A., García Simón, A., Beyrouti, R. C., & **Pastor, O.** (2022). Integration of clinical and genomic data to enhance precision medicine: a case of study applied to the retina-macula. Software and Systems Modeling, 1-16. DOI: https://doi.org/10.1007/s10270-022-01039-4



# Conceptual Modeling and Life Engineering: Yes, the Two Sides of the Same Coin…


DANKSCHEEN
TASHAKKUR ATU
YAQHANYELAY
TINGKI
BIYAN
SHUKRIA
GRACIAS
SUKSAMA
EKHMET
THANK
ARIGATO
SHUKURIA
MAAKE
GRAZIE
MEHRBANI
PALDIES
YOU
GOZAIMASHITA
EFCHARISTO
KOMAPSUMNIDA
JUSPAXAR
BOLZÏN
MERCI

PROS
Centro de Investigación en Métodos de Producción de Software

ÓSCAR PASTOR
Director

opastor@pros.upv.es

T. +34 96 387 70 07 Ext. 77353 · F. +34 96 387 73 59 · M. +34 616 467 009

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

VRAIN
Valencian Research Institute for Artificial Intelligence

PROS
Centro de Investigación en Métodos de Producción de Software

# Deciphering the Language of Life: Combining Software Engineering and Life Engineering

SISTEDES SEMINAR SERIES
October 24, 2022

*Oscar Pastor*
*opastor@pros.upv.es*