



**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*





**Barcelona  
Supercomputing  
Center**  
*Centro Nacional de Supercomputación*



# **Bienvenida**

**Dr. Josep M. Martorell**

**Prof. Alfonso Valencia**

**12/2019**

**Infoday sobre tecnologías del lenguaje**




# The Evolution of the Research Paradigm




## Numerical Simulation and Big Data Analysis

- Reduce expense
- Avoid suffering
- Help to build knowledge where experiments are impossible or not affordable


# HPC: An enabler for all scientific fields




**Materials,  
Chemistry &  
Nanoscience**



**Engineering**



**Astro,  
High Energy  
& Plasma  
Physics**



**Life Sciences  
& Medicine**



**Earth  
Sciences**

## Advances leading to:

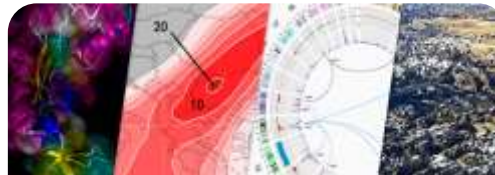
- Improved Healthcare
- Better Climate Forecasting
- Superior Materials
- More Competitive Industry

# Barcelona Supercomputing Center

## Centro Nacional de Supercomputación



Supercomputing services  
to Spanish and  
EU researchers



R&D in Computer,  
Life, Earth and  
Engineering Sciences



PhD programme,  
technology transfer,  
public engagement

BSC-CNS is  
a consortium  
that includes

Spanish Government

60%



Catalonian Government

30%



Univ. Politècnica de Catalunya (UPC)

10%





# MareNostrum4

Total peak performance: **13,7 Pflops**

General Purpose Cluster: 11.15 Pflops (1.07.2017)

CTE1-P9+Volta: 1.57 Pflops (1.03.2018)

CTE2-Arm V8: 0.5 Pflops (????)

CTE3-KNH?: 0.5 Pflops (????)



## MareNostrum 1

2004 – 42,3 Tflops

1<sup>st</sup> Europe / 4<sup>th</sup> World

New technologies

## MareNostrum 2

2006 – 94,2 Tflops

1<sup>st</sup> Europe / 5<sup>th</sup> World

New technologies

## MareNostrum 3

2012 – 1,1 Pflops

12<sup>th</sup> Europe / 36<sup>th</sup> World

## MareNostrum 4

2017 – 11,1 Pflops

2<sup>nd</sup> Europe / 13<sup>th</sup> World

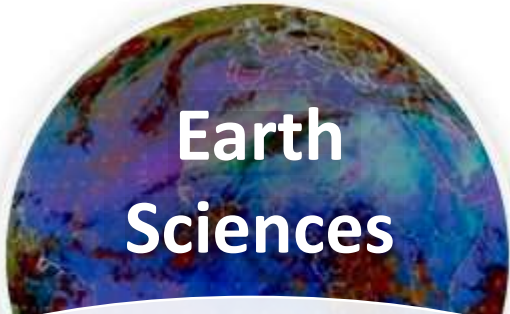
New technologies

# Mission of BSC Scientific Departments



## Computer Sciences

To influence the way machines are built, programmed and used: programming models, performance tools, Big Data, computer architecture, energy efficiency



## Earth Sciences

To develop and implement global and regional state-of-the-art models for short-term air quality forecast and long-term climate applications



## Life Sciences

To understand living organisms by means of theoretical and computational methods (molecular modeling, genomics, proteomics)



## CASE

To develop scientific and engineering software to efficiently exploit super-computing capabilities (biomedical, geophysics, atmospheric, energy, social and economic simulations)

# TOP-10 Spanish Organizations in Horizon 2020

Legal name	EU Contribution (€)	Project Participations
CSIC	230,466,641 €	535
Tecnalia	105,437,290 €	234
<b>Barcelona Supercomputing Center</b>	<b>72,747,561 €</b>	<b>130</b>
Universitat Politècnica de Catalunya	58,045,969 €	157
ATOS Spain	57,879,825 €	145
ICFO	56,517,896 €	78
Universitat Pompeu Fabra	55,709,232 €	108
CIEMAT	55,274,573 €	70
Universidad Politécnica de Madrid	52,993,088 €	153
Universitat Autònoma de Barcelona	52,860,835 €	115



# Collaborations with Industry



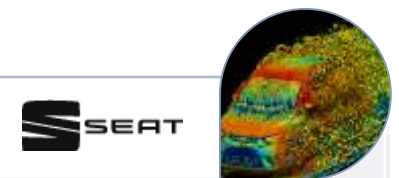
Research into advanced technologies for the exploration of hydrocarbons, subterranean and subsea reserve modelling and fluid flows



Research on wind farms optimization and wing energy production forecasts



Collaboration agreement for the development of advanced systems of deep learning with applications to banking services



Simulations to improve the understanding of the rotating wheels flow physics and its impact over the aerodynamic performance



Advanced statistical methods to the optimization of maintenance, energy usage, and control of the city's water treatment and supply processes.



Research on efficient data sensing, algorithms for analysis of industrial processes and visualization of large datasets of industrial data



Artificial Intelligence and Big Data techniques to improve the quality of care and personalized diagnosis



BSC's dust storm forecast system licensed to be used to improve the safety of business flights.

# MareNostrum4

Total peak performance: **13,7 Pflops**

General Purpose Cluster:	11.15 Pflops	(1.07.2017)
CTE1-P9+Volta:	1.57 Pflops	(1.03.2018)
CTE2-Arm V8:	>0.5 Pflops	(31.12.2019)
CTE3-AMD:	>0.5 Pflops	(1.12.2019)



## MareNostrum 1

2004 – 42,3 Tflops

1<sup>st</sup> Europe / 4<sup>th</sup> World

New technologies

## MareNostrum 2

2006 – 94,2 Tflops

1<sup>st</sup> Europe / 5<sup>th</sup> World

New technologies

## MareNostrum 3

2012 – 1,1 Pflops

12<sup>th</sup> Europe / 36<sup>th</sup> World

## MareNostrum 4

2017 – 11,1 Pflops

2<sup>nd</sup> Europe / 13<sup>th</sup> World

New technologies

# The new MareNostrum5

## La UE instalará en Barcelona uno de los tres superordenadores más veloces del continente

Apuesta por la ciencia

### Llega una nueva generación de superordenadores a BCN

El Barcelona Supercomputing Center albergará esta joya de la tecnología a partir del 2020

La Unión Europea aportará 100 millones de euros para el proyecto del MareNostrum 5

Europa confía en España para «supercompetir» con EE.UU. y Asia

- La Comisión Europea elige Barcelona para instalar uno de los nuevos superordenadores, que contará con una inversión de 200 millones de euros



Europa destina 100 millones al Centro Nacional de Computación, su mayor aportación a una infraestructura de investigación española

## Un superordenador a Barcelona



# MareNostrum 5

## A European pre-exascale supercomputer

200 Petaflops peak performance ( $200 \times 10^{15}$ ):

Experimental platform to create supercomputing technologies  
“made in Europe”

223 M€ of investment:



Hosting Consortium:

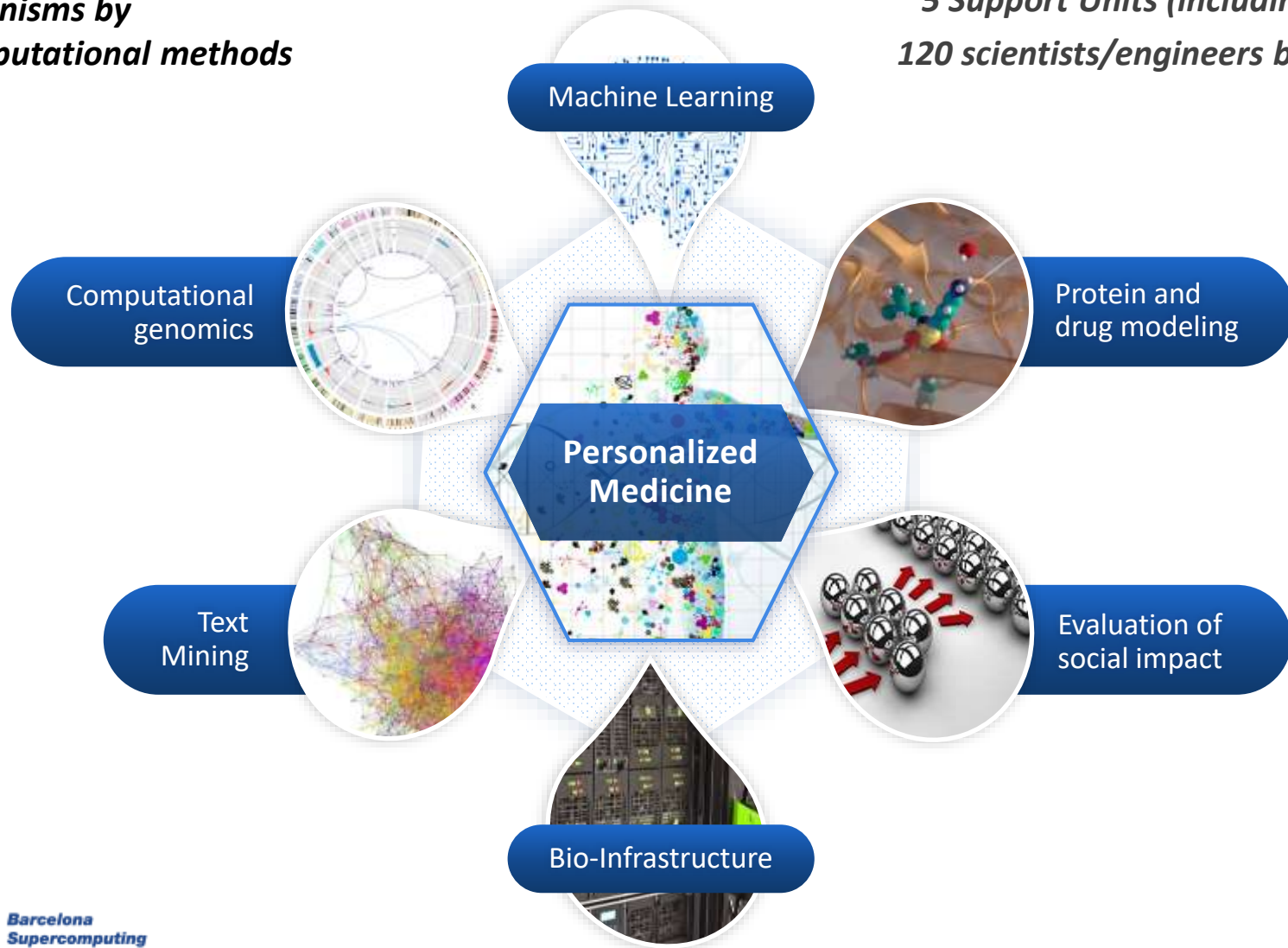
Spain | Portugal | Turkey | Croatia



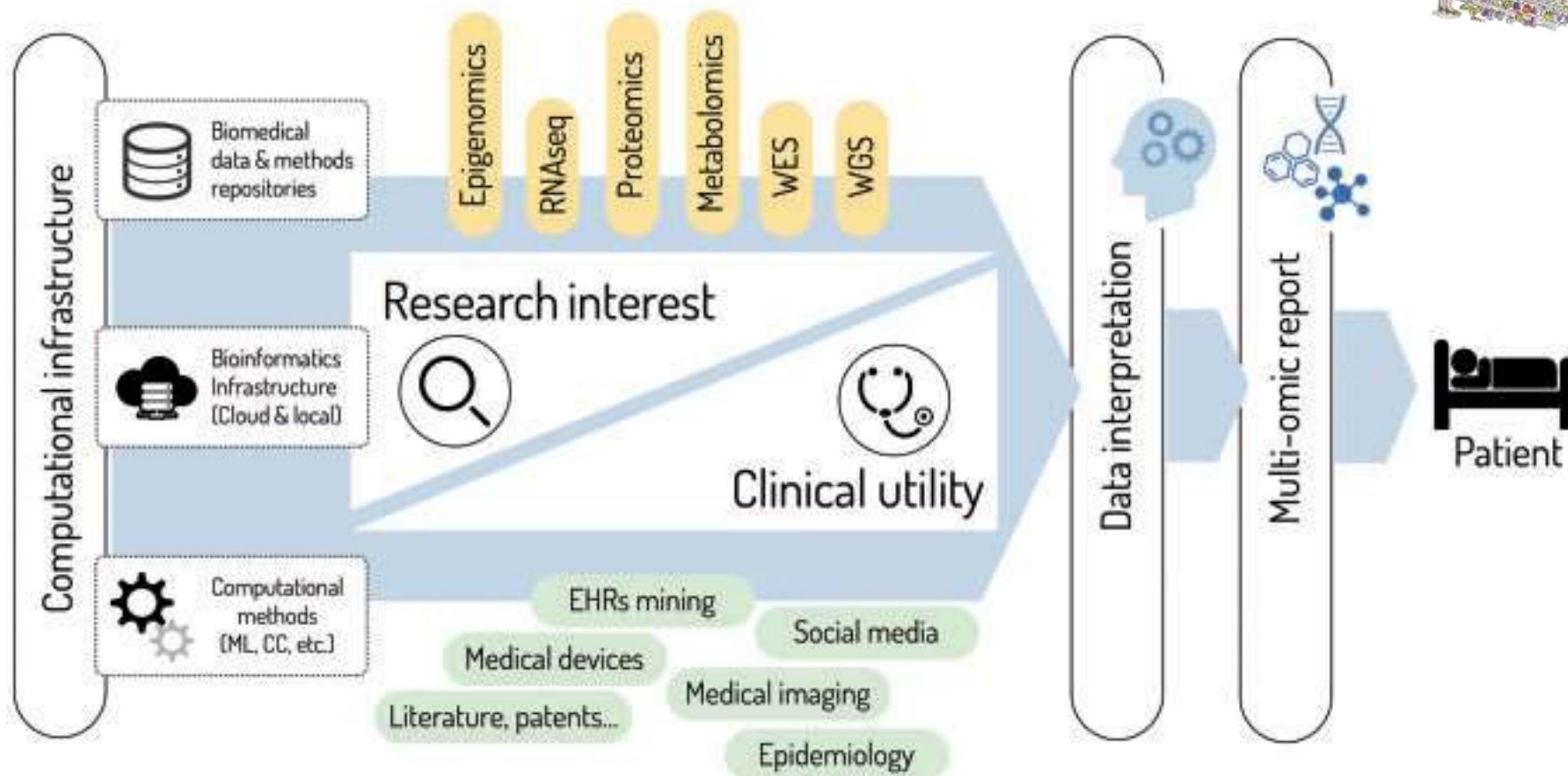
# Life Sciences Department BSC

*Understanding living organisms by computational methods*

*7 research groups  
5 Support Units (including INB)  
120 scientists/engineers by 2020*



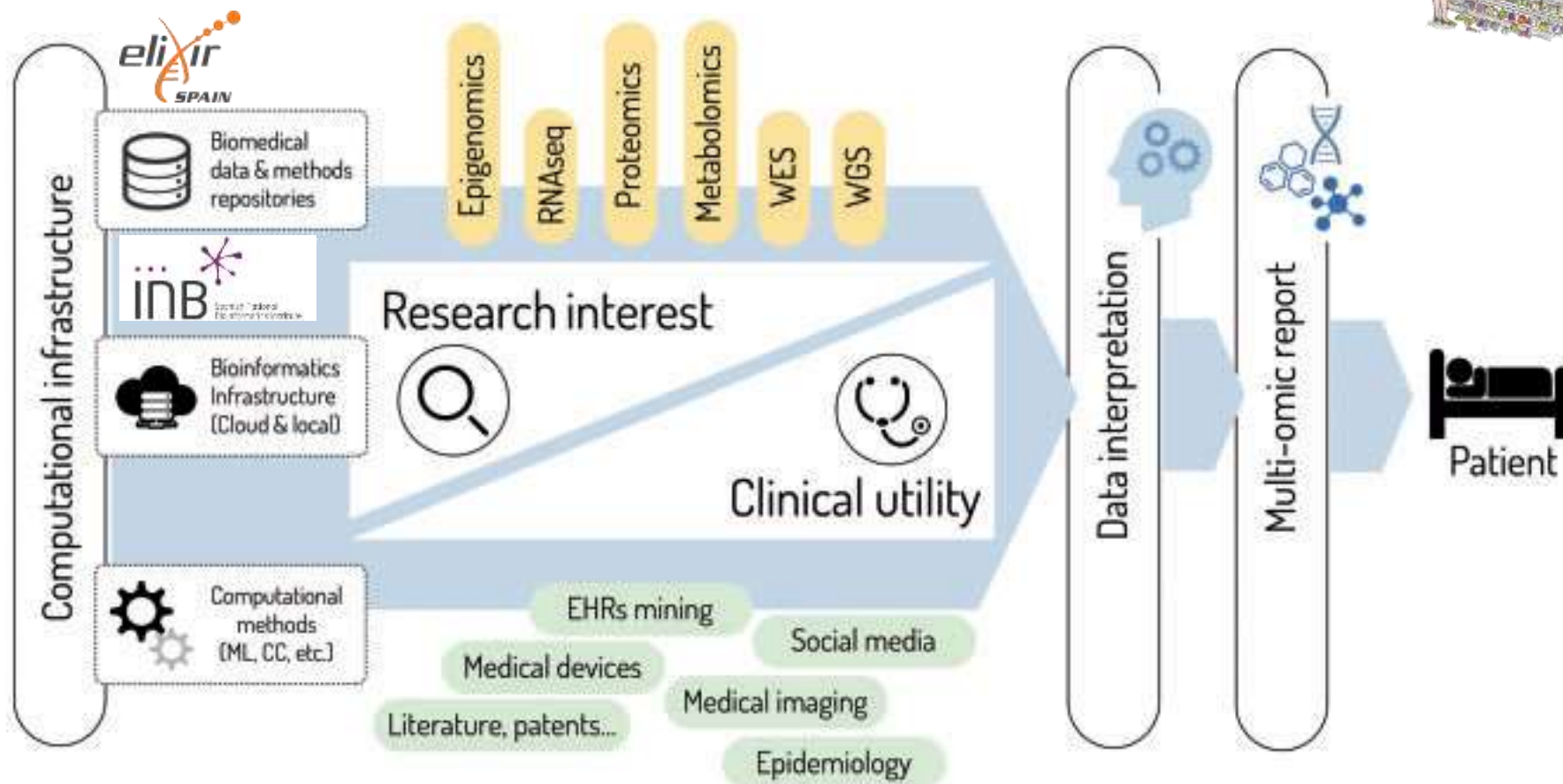
# Personalized Medicine



Gomez-Lopez 2019

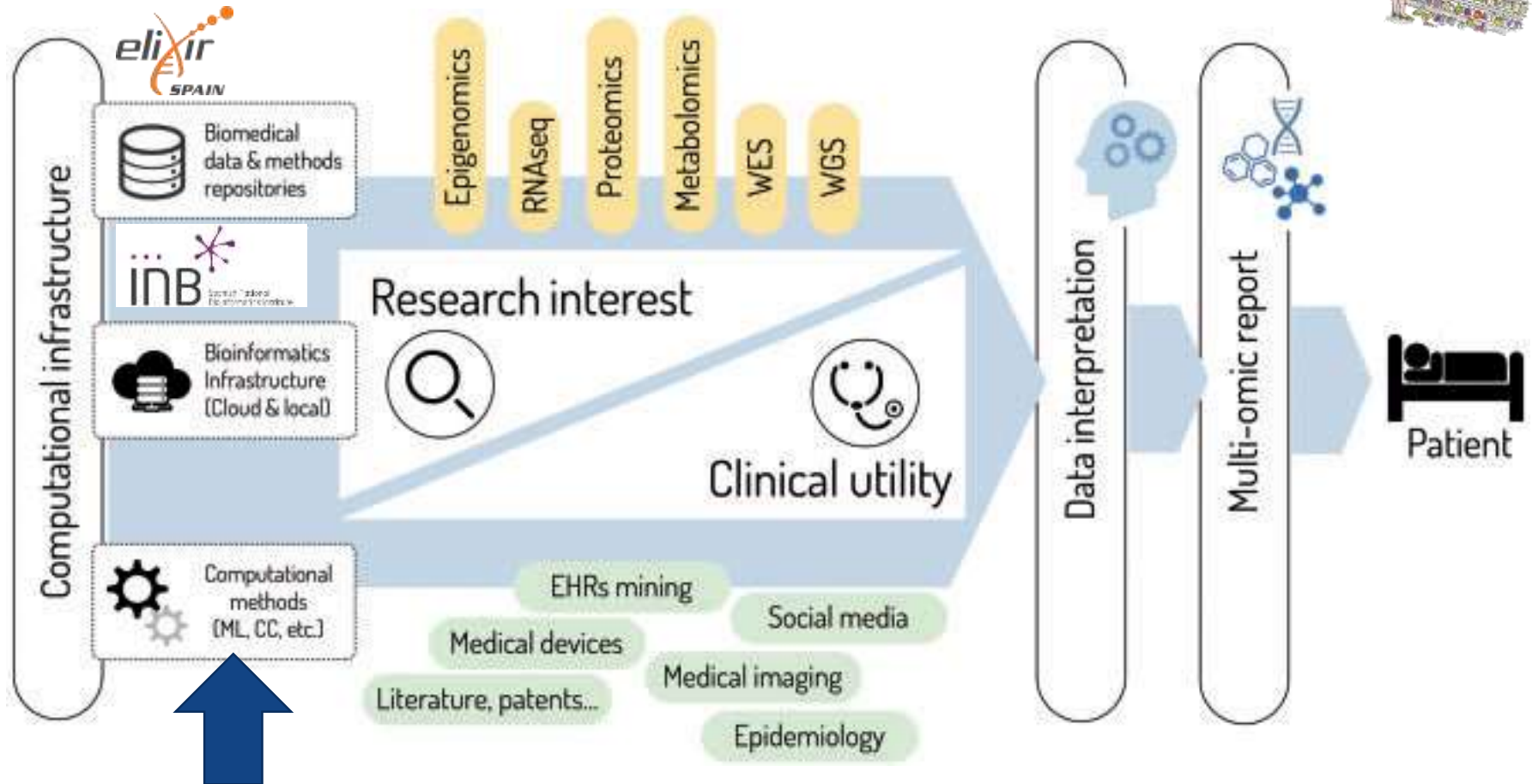


# Personalized Medicine



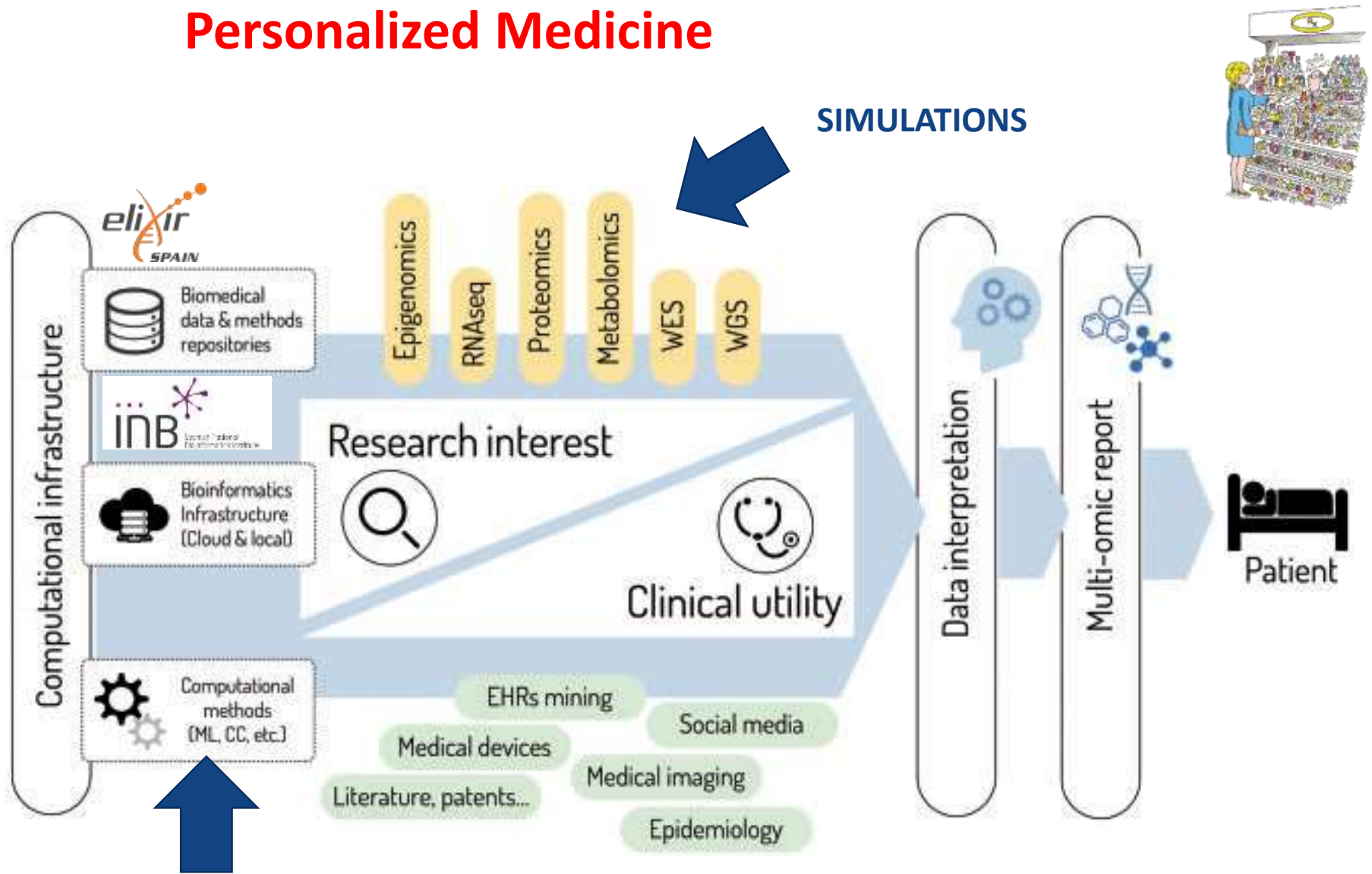
Gomez-Lopez 2019

# Personalized Medicine



Gomez-Lopez 2019

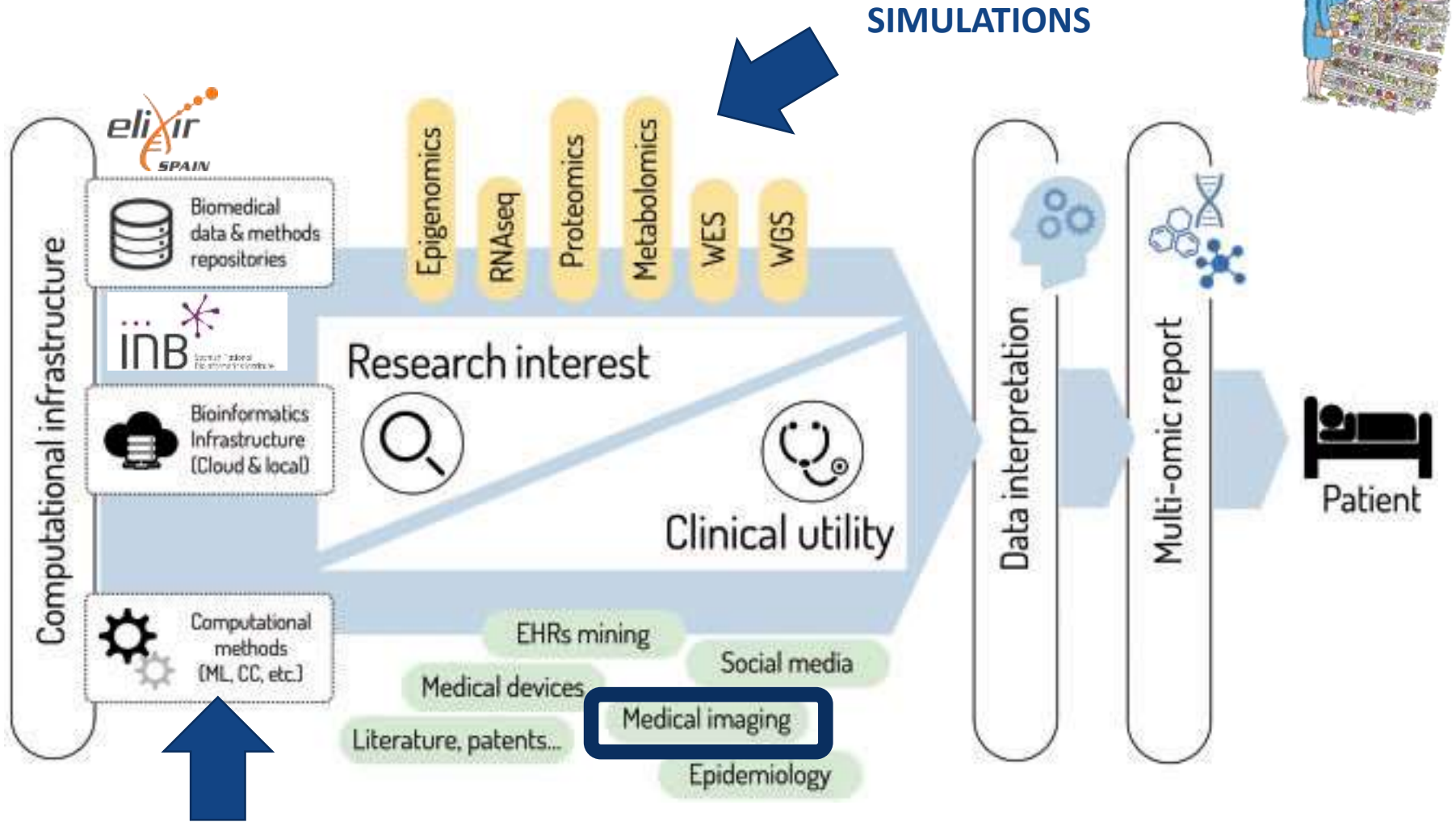
# Personalized Medicine



Gomez-Lopez 2019



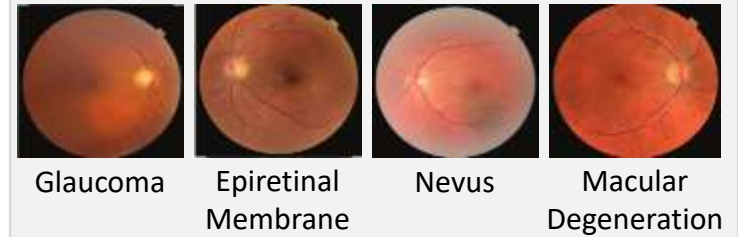
# Personalized Medicine



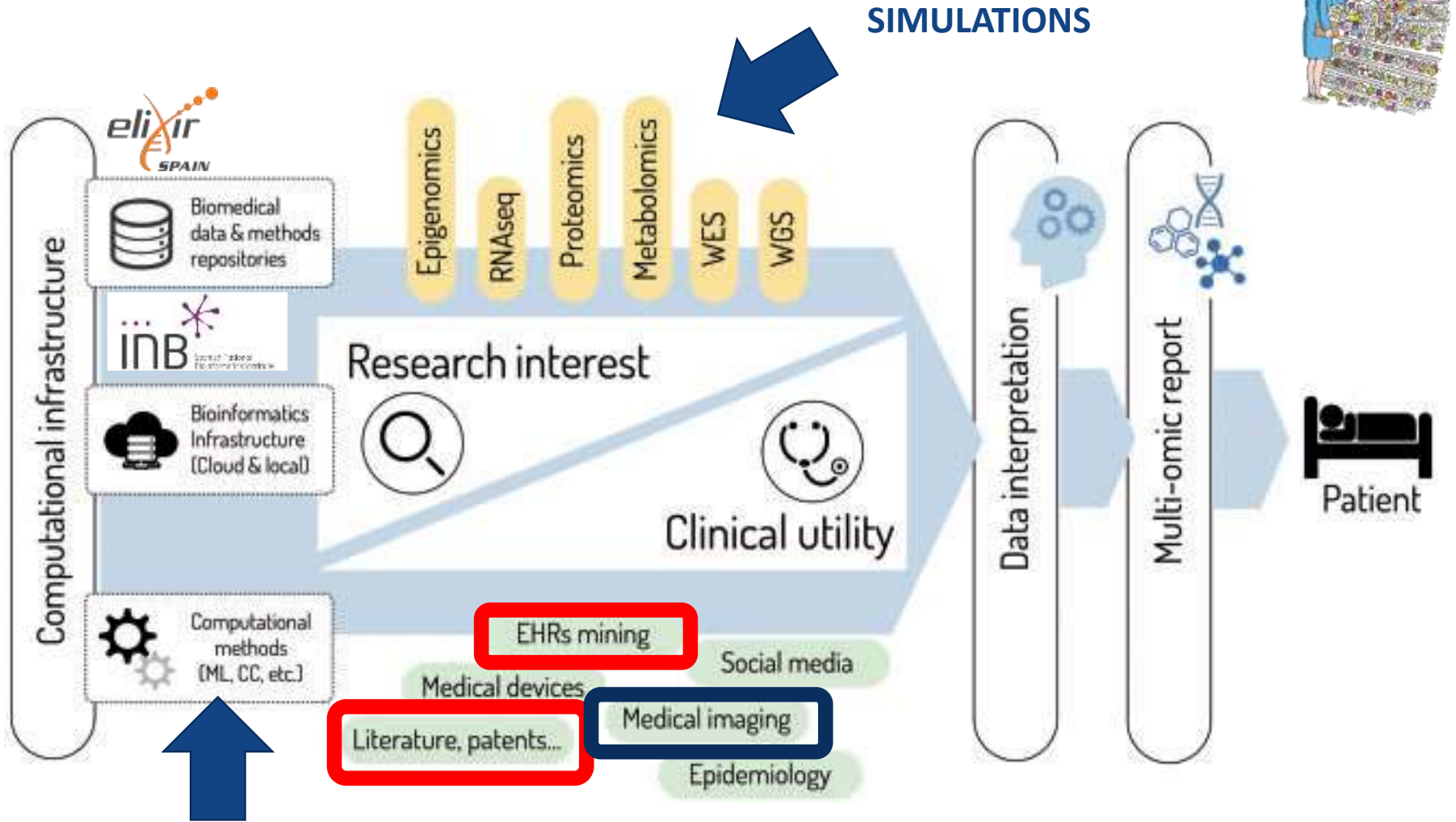
Gomez-Lopez 2019

# BSC works on Medical Imaging

- **Detecting retina pathologies**
  - Trained models competitive with ophthalmologists
  - With Lenovo & Hospital Vall Hebron
- **Learning from liver conditions**
  - Learning about rare diseases
  - With Hospital Clinic
- **Predicting and guiding in-vitro success**
  - Finding the best embryo ASAP
  - With Hospital Clinic
- **Supporting medical doctors on Rx review**
  - Aid for Dr. in rural areas
  - With Asepeyo and ICS



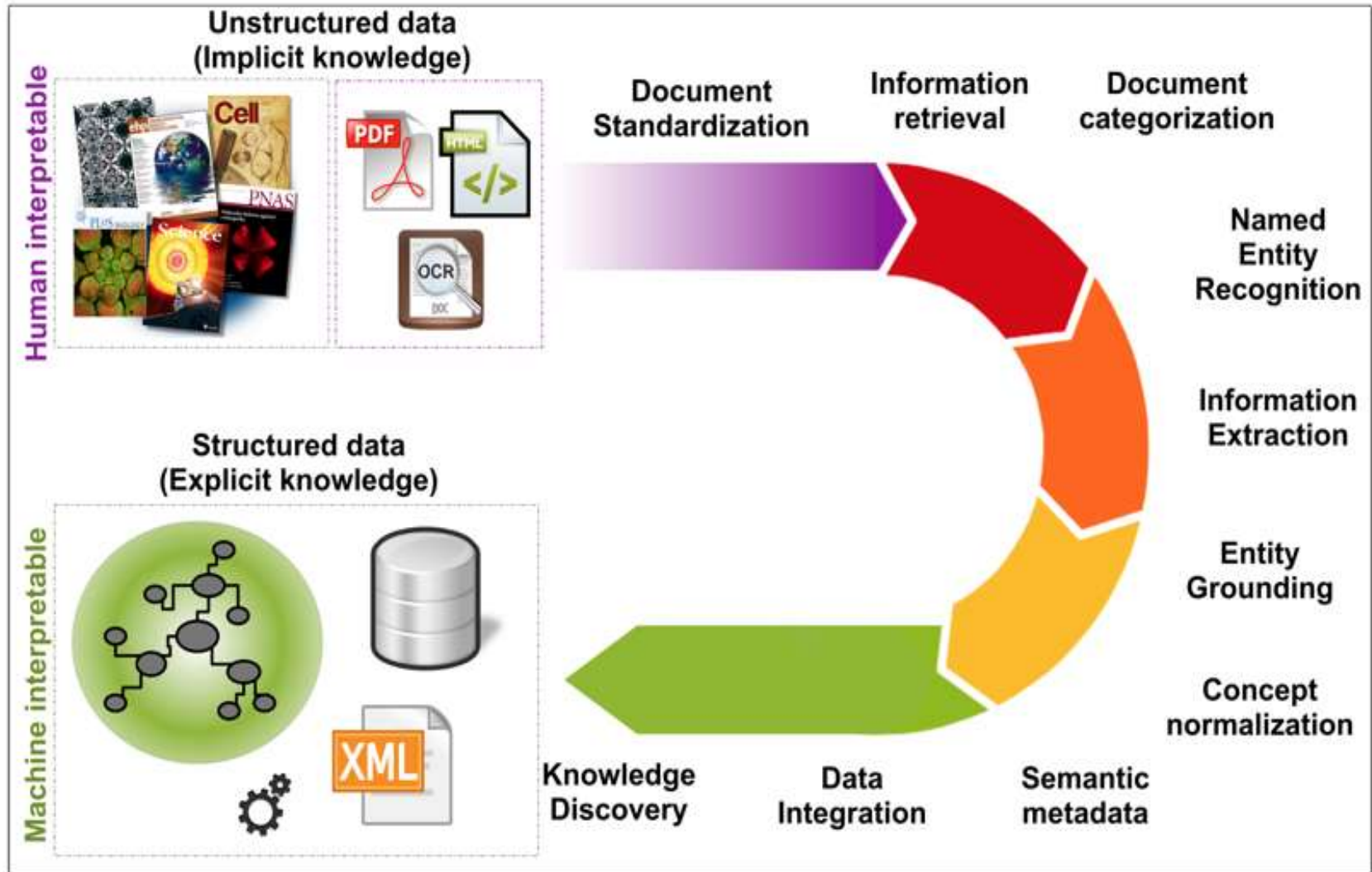
# Personalized Medicine



Gomez-Lopez 2019



# Text mining and NLP in biomedicine



# Word embeddings

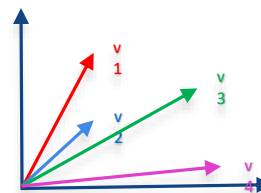


Word2Vec  
fastText

→ |V|

v1	0,78	0,65	0,98
v2	0,23	0,12	0,32
v3	0,90	0,32	0,56
v4	0,08	0,43	0,65
v5	0,77	0,88	0,77

Word vectors



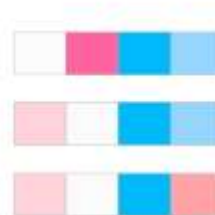
Evaluación intrínseca: cálculo similitud entre términos (sinónimos en SNOMED)

Evaluación extrínseca: comprobar su utilidad en otras tareas PLN (neuroNER)

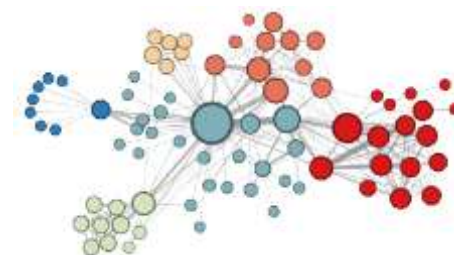


MAPK3  
H3K9  
methylation  
cerebellar granule  
cells

Word2vec



Word embeddings



individualized Paediatric Cure (iPC)  
Cloud-based virtual-patient models  
for precision paediatric oncology

# Spanish Plan for Advancement of Language Technologies

The screenshot shows the homepage of the website. At the top left, there is a header with the Spanish flag, the coat of arms, and the text: "GOBIERNO DE ESPAÑA", "MINISTERIO DE ECONOMÍA Y EMPRESA", and "SECRETARÍA DE ESTADO PARA EL AVANCE DIGITAL". To the right of this header is the title "Plan de Impulso de las Tecnologías del Lenguaje" and navigation links for "Inicio", "Mapa del sitio", and "Contacto", along with a search icon. Below the header is a dark teal navigation bar with three tabs: "Inicio", "Tecnologías del Lenguaje" (which is active), and "Sanidad". The main content area has a light blue background. On the left, there is a large white arrow pointing left. In the center, there is a play button icon inside a speech bubble. The main heading reads "Plan de Impulso de las Tecnologías del Lenguaje". Below it, a paragraph states: "El nuevo plan tiene como objetivo fomentar el desarrollo del procesamiento del lenguaje natural y la traducción automática en lengua española y lenguas cooficiales." Underneath this is a link that says "Conoce el plan". On the right side, there is a circular logo for "Plan TL" which features a globe with network connections. Below the logo, it says "Plan de Impulso de las Tecnologías del Lenguaje". On the far right, there is a large white arrow pointing right. At the bottom center, there are three small white circles, with the first one being filled.

GOBIERNO DE ESPAÑA  
MINISTERIO DE ECONOMÍA Y EMPRESA  
SECRETARÍA DE ESTADO PARA EL AVANCE DIGITAL

Plan de Impulso de las  
**Tecnologías del Lenguaje**

Inicio Mapa del sitio Contacto

Inicio Tecnologías del Lenguaje Sanidad

Plan de Impulso de las  
**Tecnologías del Lenguaje**

El nuevo plan tiene como objetivo fomentar el desarrollo del procesamiento del lenguaje natural y la traducción automática en lengua española y lenguas cooficiales.

Conoce el plan

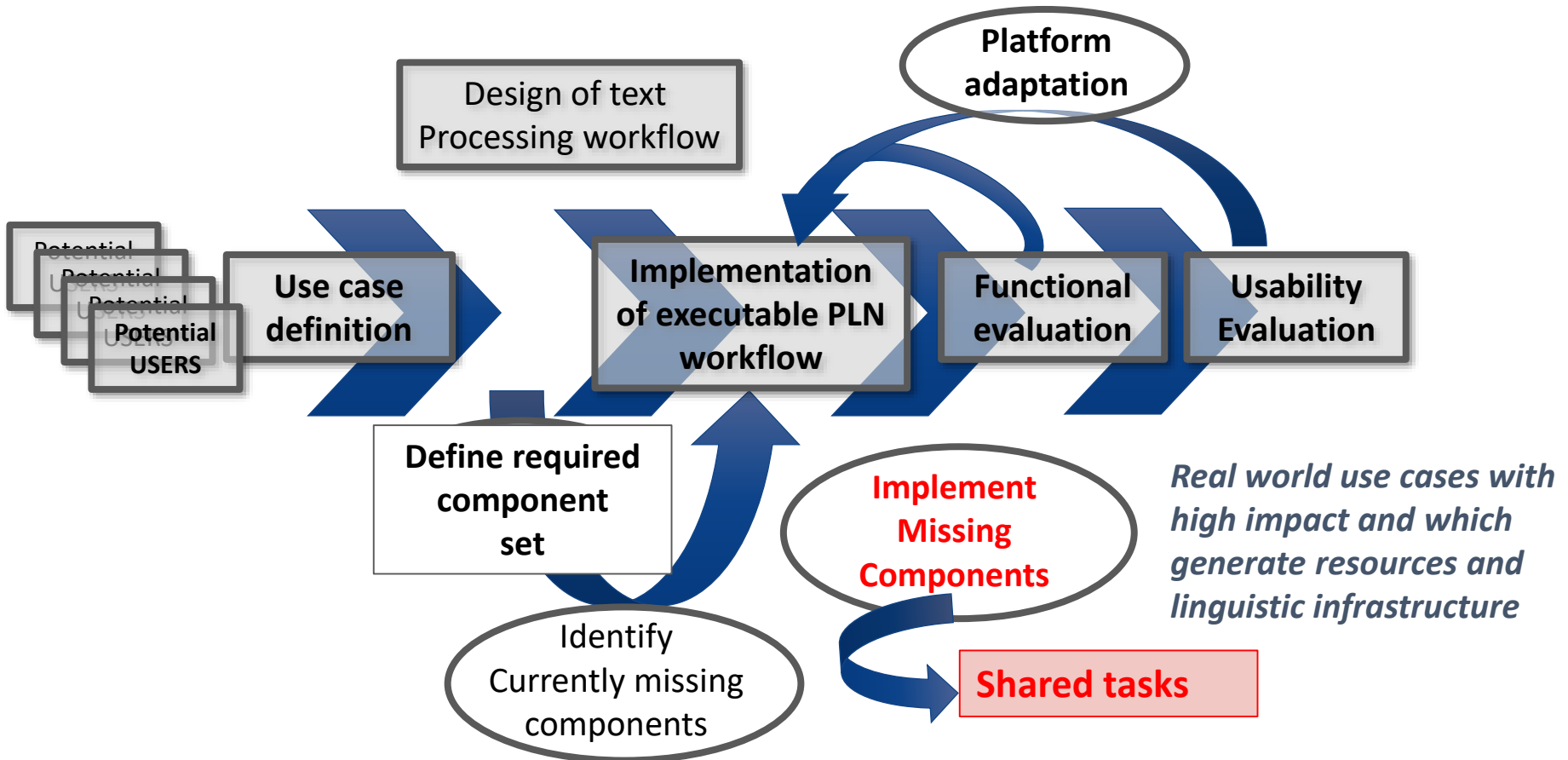
**Plan TL**  
Plan de Impulso de las  
Tecnologías del Lenguaje

https://www.plantl.gob.es

<https://www.plantl.gob.es>



# Use cases of Plan TL and shared tasks



Martin Krallinger  
Marta Villegas

Joaquim Moré  
Claudia Rozas

David Vicente

Renata  
Gimenez



**Barcelona  
Supercomputing  
Center**  
*Centro Nacional de Supercomputación*



# Thank you

[Valencia@bsc.es](mailto:Valencia@bsc.es)

[Martorell@bsc.es](mailto:Martorell@bsc.es)



[@martorellBSC](https://twitter.com/martorellBSC)

[@alfons\\_valencia](https://twitter.com/alfons_valencia)

12/2019