



# Clinical Natural Language Processing and health interoperability to support knowledge management and governance in Rare Cancers

LifeSTech Beatriz Merino,  
Laura Lopez-Perez, Eugenio Gaeta



UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

# Clinical cancer research situation and limitations

**Health data** constitutes **30%** of the world's data

## Rare cancers

**Incidence** rate < 6 in 100,000 but the majority (74%): < **0.5** in 100,000

**200 types** of rare cancers exist **~20-25%** of all cancer diagnoses

In Europe, rare cancers have a **5-year OS rate** of **47%** vs 67% for common cancers

↓ **Scientific attention and financial support**    **clinical trials challenge**    **hard evidence & access**



Priority area for the EHDS

Introduction & Goal

Methodology

Results

Conclusions



**Goal:** Provide equal access to high-quality specialist care throughout the EU

*Clinical Natural Language Processing and health interoperability to support knowledge management and governance in Rare Cancers. In the 16<sup>th</sup> MEDICON and the 5<sup>th</sup> CMBEIH.*

# Clinical cancer research situation and limitations



Intelligent Ecosystem to improve the governance, the sharing and the re-use of health Data for Rare Cancers

1 SEP. 2022 – 31 AGO. 2026

To establish a **Data Space for rare cancers** that will make possible the **re-use of existing multisource health data** (cancer registry data, national registries, data from biobanks etc.) **across EU healthcare systems leveraging emerging interoperability technologies and AI approaches.**

The project approach will be experienced in the framework of the **European reference network for rare adult solid cancers (EURACAN)**.

**Sarcoma** and **Head and Neck** use cases  
**11 pilot sites** in Italian, French, Spanish, Polish, Czech, Norwegian and Swedish

Introduction & Goal

Methodology

Results

Conclusions

*Clinical Natural Language Processing and health interoperability to support knowledge management and governance in Rare Cancers. In the 16<sup>th</sup> MEDICON and the 5<sup>th</sup> CMBEBIH.*

# TEHDAS

Towards  
European  
Health  
Data  
Space



European  
Commission

EUROPEAN HEALTH DATA SPACE

#EUDigitalHealth

IDEA4RC user journey: from requirements to an architecture proposal

EHDS requirements for a secure and trustworthy platform

Data search

Data permit

Data finalization



Feasibility study

Data Use

- Privacy
- Isolation
- Interoperability
- Security
- Trust
- Quality

Introduction & Goal

Methodology

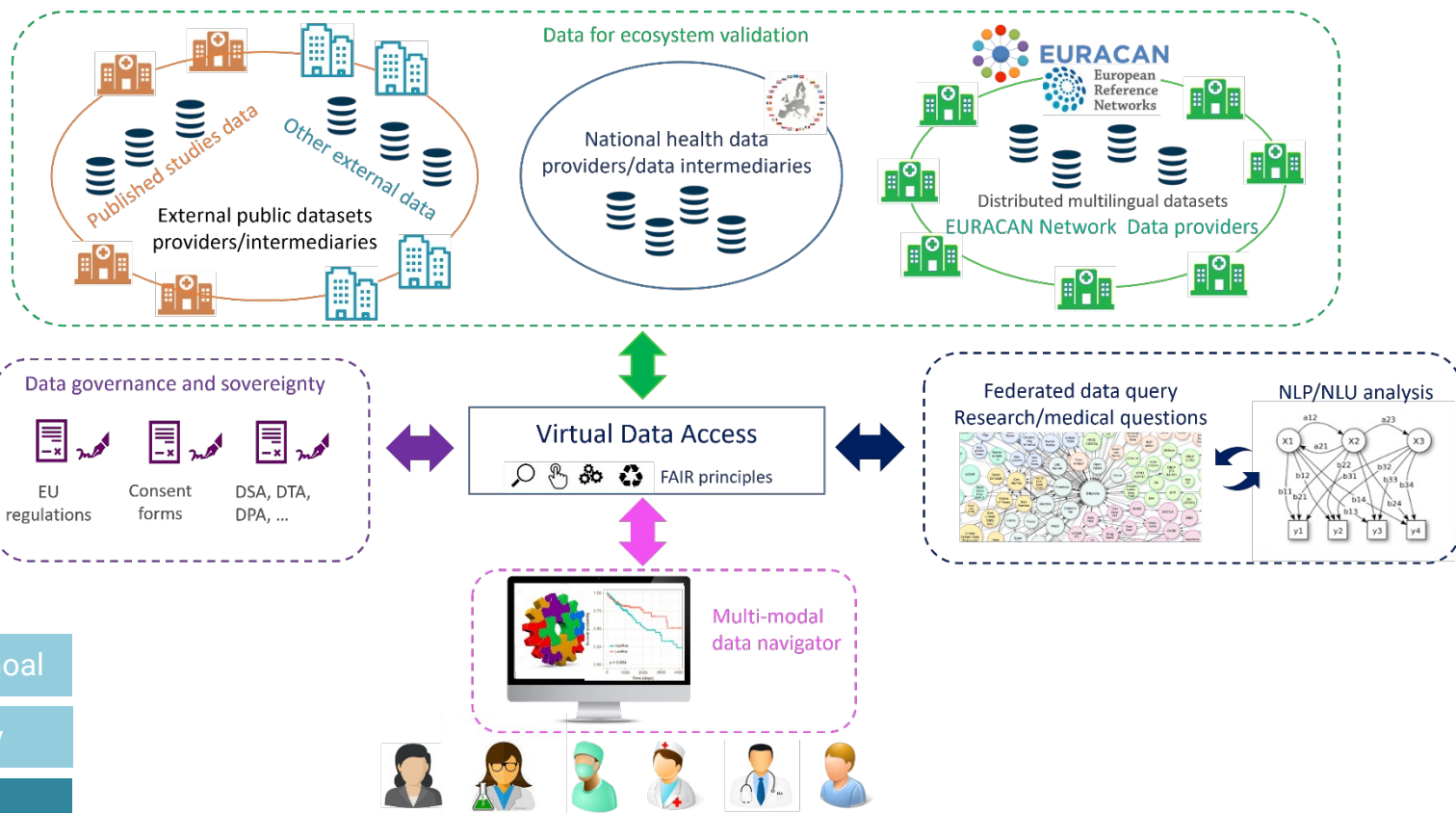
Results

Conclusions

**IDEA4RC ecosystem: innovative health services and solutions, protecting privacy & security of health data and fostering trust**

*Clinical Natural Language Processing and health interoperability to support knowledge management and governance in Rare Cancers. In the 16<sup>th</sup> MEDICON and the 5<sup>th</sup> CMBEIH.*

# IDEA4RC Data Ecosystem concept



Introduction & Goal

Methodology

Results

Conclusions

*Clinical Natural Language Processing and health interoperability to support knowledge management and governance in Rare Cancers. In the 16<sup>th</sup> MEDICON and the 5<sup>th</sup> CMBEIH.*

# IDEA4RC Data Ecosystem concept

## Ethics guidelines for trustworthy AI



## WHO Guidance



## European Principles for Digital Health



Introduction & Goal

Methodology

Results

Conclusions

## Data governance and sovereignty



EU regulations



Consent forms



DSA, DTA, DPA, ...

## Regulations revised



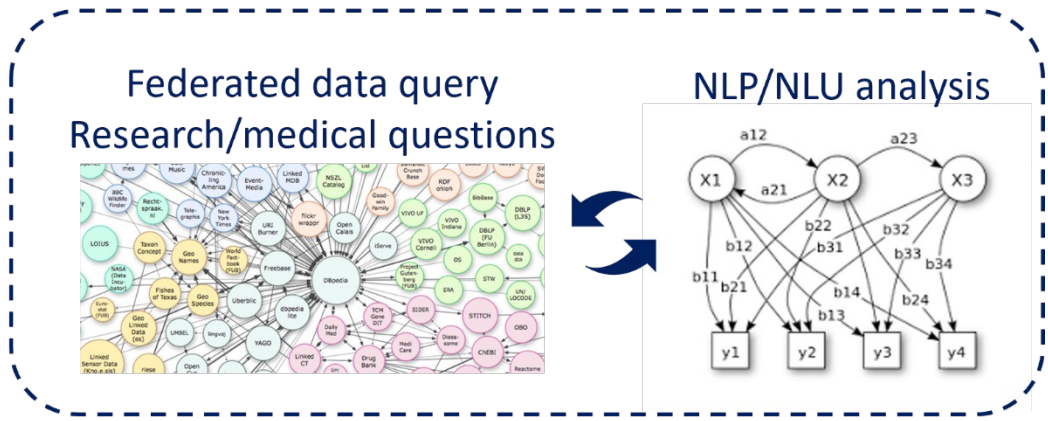
Developments on common European Data Spaces (EDS) for the best use of health data to advance knowledge, care, and research for rare cancers

*Clinical Natural Language Processing and health interoperability to support knowledge management and governance in Rare Cancers. In the 16<sup>th</sup> MEDICON and the 5<sup>th</sup> CMBEIH.*

# IDEA4RC Data Ecosystem concept

From the Data governance layer:  
FHIR capsules (privacy-preserving environments)

↳ Anonymized data generation



## Challenges

- Multiple language. Extraction of structured information from the Electronic Health Records (EHRs) free text
- Help clinical and epidemiological researchers to explore data

**MLOps:** continuous integration and deployment practices with proper monitoring, validation, and governance of ML models.

**NLP tools:** to unify and facilitate the identification and exploitation of relevant information related to rare cancers.

Clinical Natural Language Processing and health interoperability to support knowledge management and governance in Rare Cancers. In the 16<sup>th</sup> MEDICON and the 5<sup>th</sup> CMBEIH.

- Introduction & Goal
- Methodology
- Results
- Conclusions

# IDEA4RC Data Ecosystem concept

**Multimodal interaction framework** will involve a virtual assistant that uses **natural modes of communication aligned with the needs** of the healthcare researchers to provide an **advanced augmented analytical system**



synthetic voice, voice recognition, or NLU will be considered

A **multilingual UI** for assistant will be provided to

- a) explore available **dictionaries**
- b) use selected **canonical sentences**

Introduction & Goal

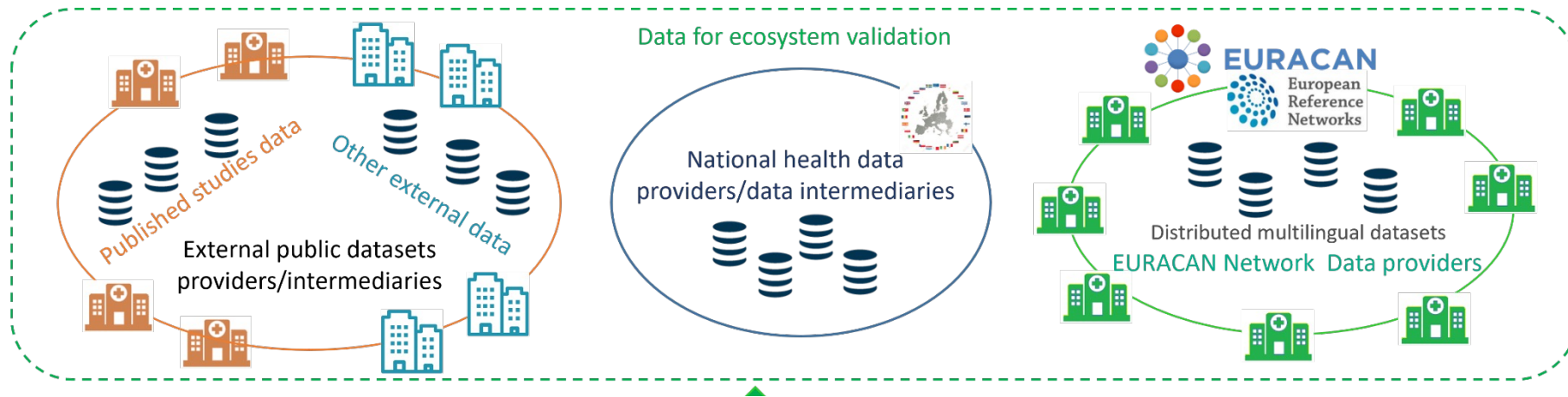
Methodology

Results

Conclusions



# IDEA4RC Data Ecosystem concept



The capabilities of the ecosystem will be assessed in real use cases addressing: (1) the provision of **more evidence-sustained** information on the natural history of the disease, (2) the identification of **potential prognostic and predictive factors** and to evaluate **treatment effectiveness**, and (3) the evaluation of the **quality of care**

Introduction & Goal

Methodology

Results

Conclusions

*Clinical Natural Language Processing and health interoperability to support knowledge management and governance in Rare Cancers. In the 16<sup>th</sup> MEDICON and the 5<sup>th</sup> CMBEIH.*

# Intelligent ecosystem to improve the governance, the sharing, and the re-use of health data for rare cancers



With the implementation of the Rare Cancer Data Ecosystem, we aim to create a **cross-EU health data space for rare cancers** that will contribute to the EU's research and industry leadership in the domain

Our **ambition** is to address

- ✓ the negative impacts of **scarcity of data**
- ✓ improve **quality of care, research opportunities, and health policy making**
- ✓ enable the benefits of innovative **analytical tools** and AI for rare cancer patients

**The main goal is to enable better, accessible, and inclusive care for rare cancer patients by advancing knowledge through research, creating data-driven indicators for quality of care and identifying care inequalities, and transforming and making more efficient the referral and virtual consultation processes.**

Introduction & Goal

Methodology

Results

Conclusions