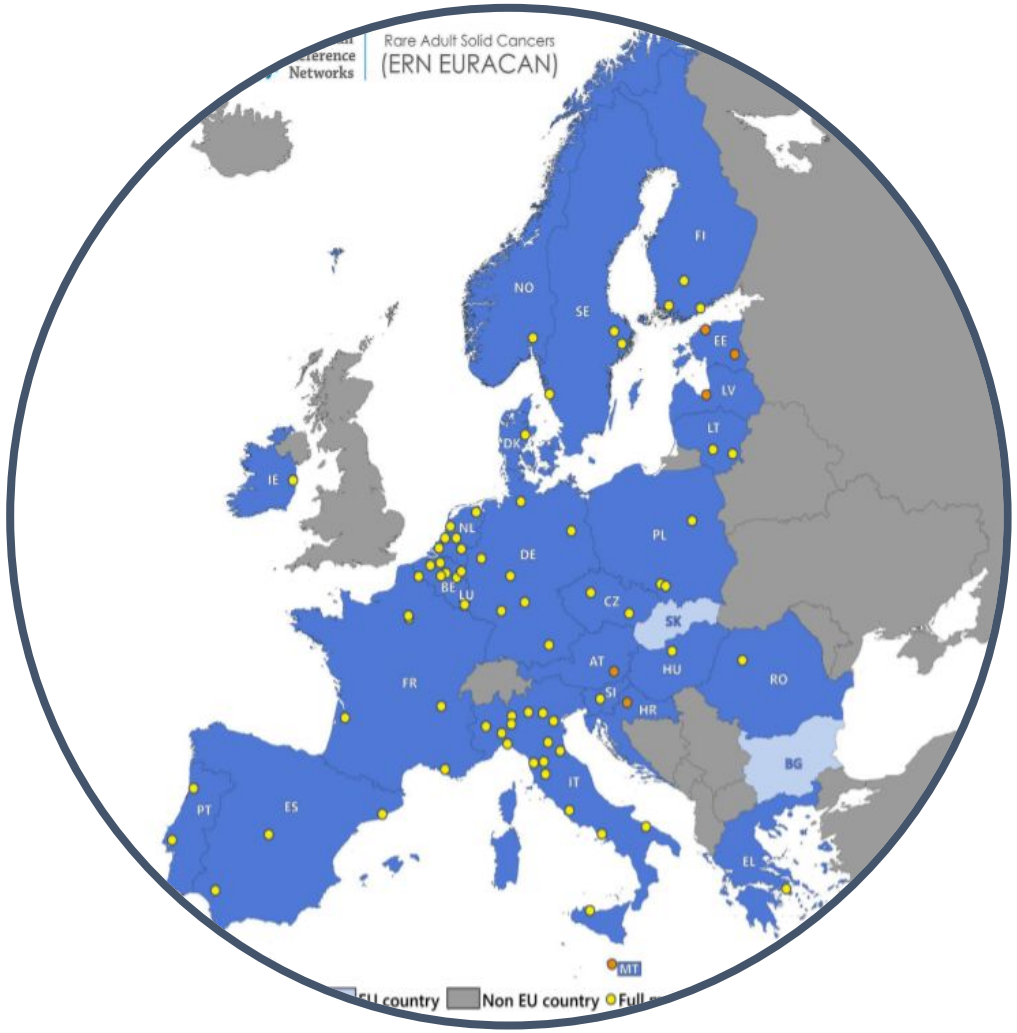
IDEA4RC vision

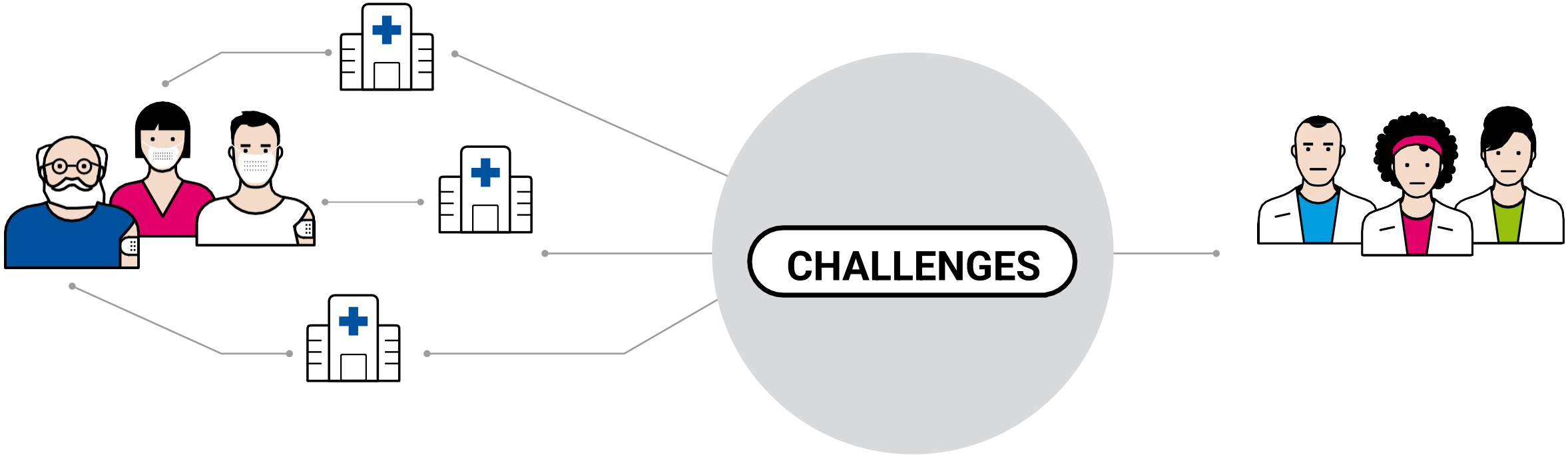
- All observational data for all adult patients with rare cancer in Europe is made available for research, innovation, and clinical decision making
- With no registration burden for clinicians
- With a simple and transparent mechanism to make data available for all stakeholders
- While protecting privacy of individuals and maintaining autonomy and data control of data providers

European Reference Networks

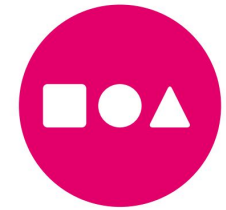
EURACAN



IDEA4RC



Data heterogeneity and availability



Data findability and understanding

Data security and trust

Data access

IDEA4RC

Challenge 1

Data heterogeneity and availability

- Electronic Health Records (EHRs) contain structured and unstructured data
- EHR are written in different EU languages Data
- Quality is highly variable
- Different formats, often incompatible
- Different standards and data models

IDEA4RC

Challenge 2

Data findability and Understanding

- Lack of knowledge about the possibility that a dataset already exists
- Metadata without details: difficult to understand whether the dataset meets the needs
- Different languages
- Different vocabulary/terminology used

IDEA4RC

Challenge 3

Data security and trust

- Traditional perimeter-based security is not enough in the modern IT landscape
- Complying with the security requirements stated in the GDPR, EHDS, and NIS2 entails an increase in cost and expertise that are generally not available at hospitals
- How to maintain the control over data within each data owner?
- How to enable personal health data analysis if data cannot be transferred across different institutions in context like rare cancers or rare diseases

Challenge 4

Data access

- Difficult to understand which is the organisation/institution responsible to grant data access
- Lengthy, heterogeneous and complex approval processes
- Different governance restrictions to data access
- Lack of interoperability/collaborations across data access applications
- Lack of trust: what happens after access is granted? Who are the actors involved?

IDEA4RC solutions

1. NLP for Medical Entity Extraction
2. NLP Query/cohort Builder for Sarcomas and Head & Neck Cancer Data Models
3. Conversational AI in IDEA4RC Applications
4. Federated AI in IDEA4RC

NLP for Medical Entity Extraction

An example

Me llamo Ignacio, tengo 58 años. En junio de 20254 sufrí un infarto de miocardio por el que fui atendido de urgencia en el hospital. Me realizaron un cateterismo y estuve ingresado varios días en la unidad coronaria. Tras el alta, me prescribieron medicación diaria: ácido acetilsalicílico (aspirina) como antiagregante plaquetario, atorvastatina para el control del colesterol, bisoprolol como betabloqueante, y ramipril. En la revisión de enero de 2025, me diagnosticaron hipertensión arterial. Actualmente estoy en seguimiento con el cardiólogo y el médico de cabecera para controlar la tensión y prevenir nuevos eventos cardiovasculares.

Patient

Condition

Procedure

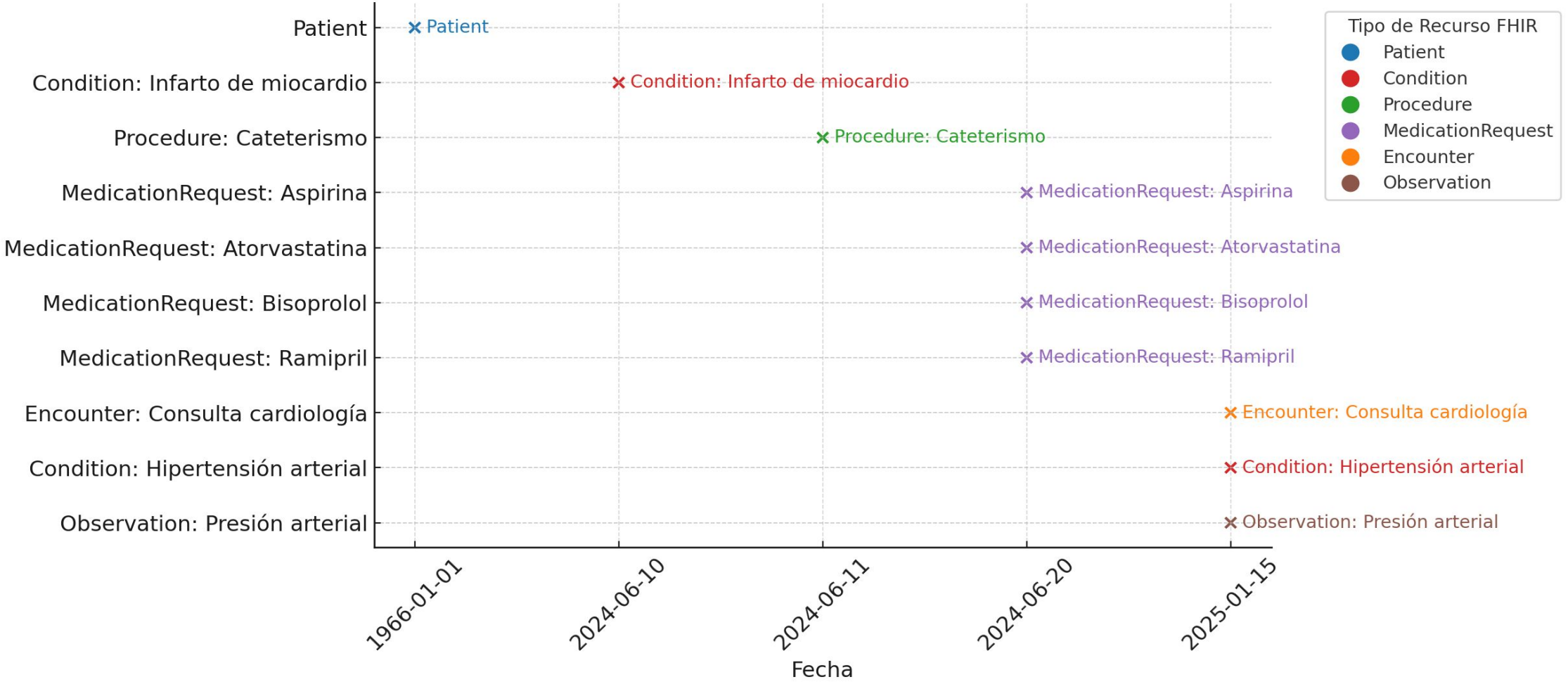
MedicationRequest

Encounter

Observation

NLP for Medical Entity Extraction

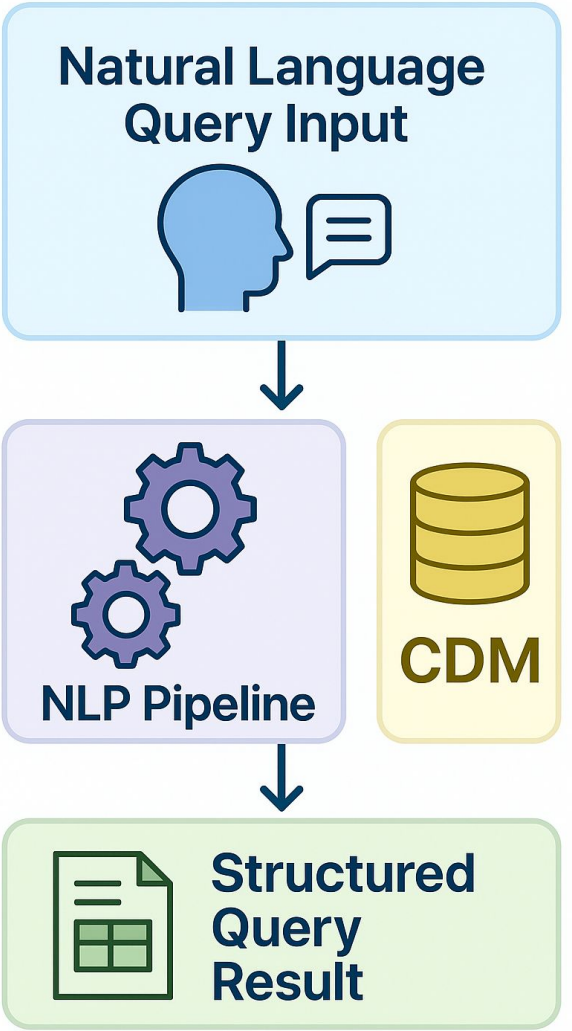
An example



NLP Query/cohort Builder for Cancer Data Models

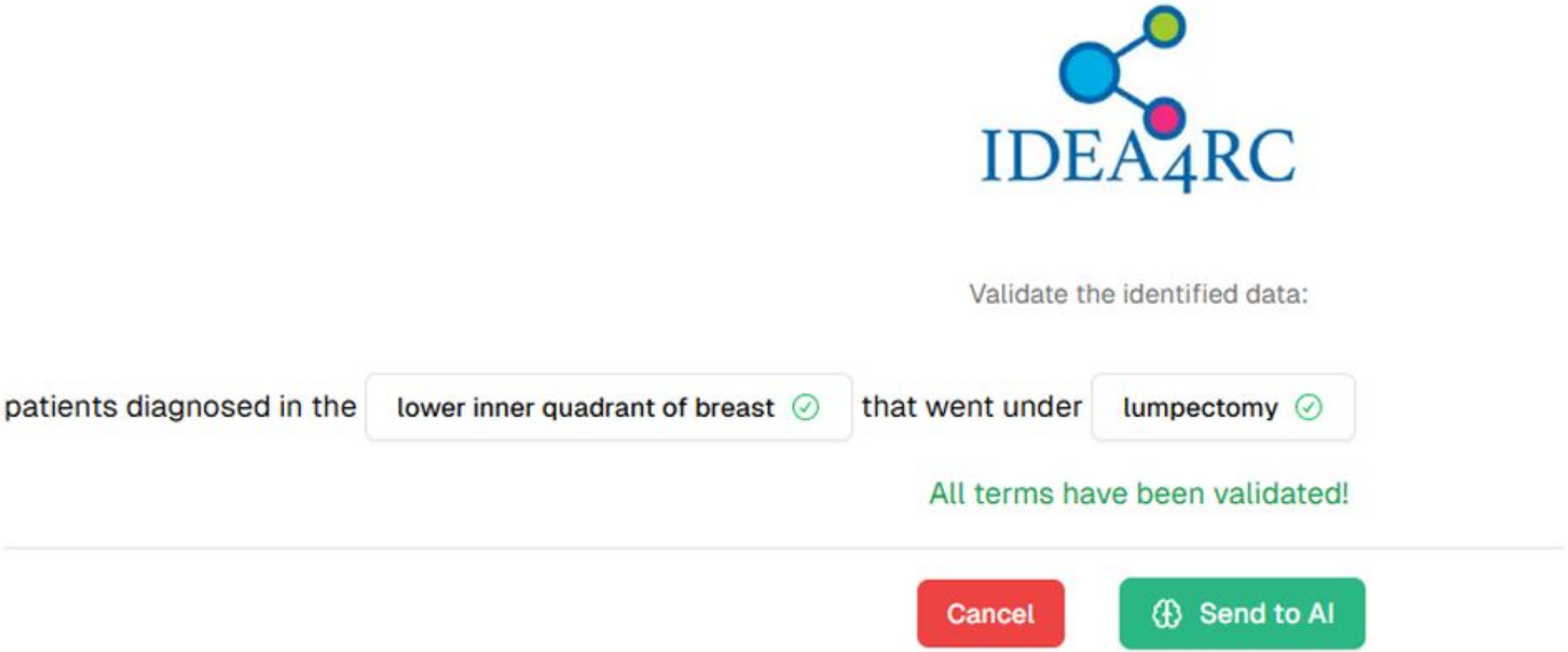
The Scenario in IDEA4RC

- Clinicians wants to use directly the tools
- No previous knowledge about OMOP or FHIR data structure need to be asumed
- The tools need to provide easy access to the most common queries



NLP Query Builder for Cancer Data Models

Innovation in IDEA4RC



Federate Learning

what is it?

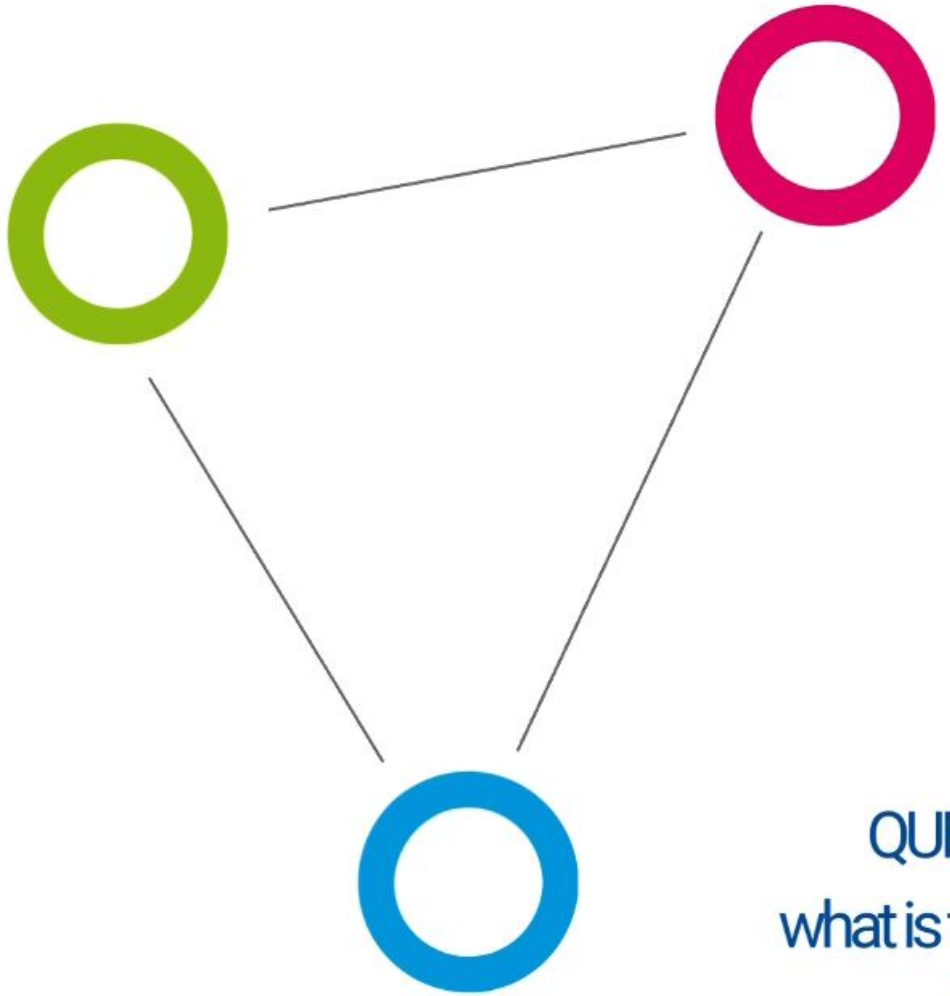
Federated Learning is a machine learning technique that enables multiple devices to collaboratively train a model without sharing raw data.

Instead, each device trains the model locally with its own data, and only the model updates (gradients) are sent to a central server for aggregation.

This approach is useful for privacy-sensitive data, as it keeps individual data secure and private throughout the training process

Federate Learning

An example



QUESTION:
what is the average age of patients?

AI-Powered Innovations in IDEA4RC

🔍 NLP for Clinical Text

- Extracts insights from multilingual free text
- Standard structured data

💬 Conversational AI

- Smart assistant for platform navigation
- Trained on IDEA4RC docs

🧠 Query Builder

- No need to know OMOP/FHIR
- Easy cohort creation for clinicians

https://www.idea4rc.eu/

🌐 Federated Learning

- Enable privacy-preserving AI across 11 centers
- Zero Trust, GDPR-compliant