IDEA4RC

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

FHIR-oriented
Cohort Builder

___

MAY 21st, 2025

Component

FHIR-oriented Cohort Builder

Partner

ENG: Engineering – Ingegneria Informatica

Brief description

The **Cohort Builder** is designed to enable biomedical researchers to effortlessly and securely define, query, and return the cardinality of patient cohorts across distributed healthcare data repositories using the **FHIR standard**.

Intended users

Different categories of users can benefit from the Cohort Builder:

1. **Clinical and epidemiology Researchers**: can define and execute cohort queries to identify multi-centres patient groups for studies.
2. **Biomedical Data Scientists**: can integrate the Cohort Builder with AI frameworks – such as INFER or the IDEA4RC federated learning framework – to benefit from structured data extraction and train AI models for predictive analysis.
3. **IT & System Administrators in Healthcare data holding organizations**: manage and maintain a privacy-aware cohort builder ready for integration in a secure infrastructure for query execution and data transport.

Functionality

It consists of the following key components:

1. **User Interface (UI) for Cohort Query Definition**
   - **Need Addressed:** Simplifies the creation and refinement of complex cohort queries by allowing researchers to browse an ontology tree and use Boolean logic.
   - **Key Feature:** Drop: down menus and automatic retrieval of search parameters to guide query building.
2. **Middleware for Secure Query Transport**
   Need Addressed: Ensures securely send queries to and retrieve results from the central platform and distributed hospital data repositories (capsules).
   Key Feature: Uses a **pull mechanism** to retrieve queries without requiring hospitals to open firewalls.
3. **FHIR Capsule Query Execution Module**
   - **Need Addressed:** Converts defined cohort criteria into structured **FHIR search queries** that can be executed within hospital research data repositories, ensuring compliance with local access permissions.
   - **Key Feature:** Ensures that only **patient IDs** are returned while maintaining Boolean logic relationships between criteria.
4. Privacy & Compliance Mechanism
   - **Need Addressed:** Prevents potential patient re-identification by enforcing a **minimum patient count threshold** for query results.
   - **Key Feature:** Ensures query results are only returned if the number of matching patients exceeds a defined limit.

5. **Query Results Aggregation and Analysis Module**
   - **Need Addressed:** Collects, consolidates, stores, and presents query results (cohort cardinality) across multiple hospital data repositories into a central platform.
   - **Key Feature:** Allows researchers to **review historical execution results** for future analysis.
6. **FHIR Resource Navigation Support**
   - **Need Addressed:** Enables queries and supports constraints across linked FHIR resources, such filtering patients based on Diagnostic Reports
   - **Key Feature:** Allows users to define constraints on related FHIR resources in a structured way.
7. **User Access Control:**
   - **Need Addressed**: Comply with privacy regulations while querying the underlying repositories
   - **Key Feature**: Ensures that only authorized users with data access permits can execute cohort queries.
8. **Interoperability with FHIR Standard:**
   - **Need Addressed**: Ensure interoperability across multiple clinical centres
   - **Key Feature**: Enables seamless integration with different healthcare systems by leveraging FHIR APIs.

Comparison with competing approaches (business value)

With respect to the current state of the art, the component offers the following advantages:
1. **Seamless Interoperability**: uses the FHIR standard, ensuring compatibility across diverse hospital systems and EHR vendors, including the INFER environment and the IDEA4RC Capsules
2. **Decentralized & Secure Data Access**: queries execute locally within hospital repositories, avoiding direct patient data transfers and reducing privacy risks.
3. **Automated FHIR Query Translation**: converts complex cohort definitions into structured FHIR queries, eliminating the need for manual query writing.
4. **Simplified Query Building Interface**: offers an ontology-based UI with Boolean logic support, making it easier for researchers to define complex queries without technical expertise.
5. **Privacy Protection Mechanisms**: enforces minimum patient count thresholds to prevent potential re-identification from small query results.
6. **Historical Query Execution Analysis**: stores past query results, allowing researchers to track changes over time and refine their cohorts.
7. **Secure Middleware-Based Query Transport**: uses a pull mechanism to retrieve queries and results without requiring hospitals to open firewalls.
8. **Support for Linked FHIR Resources**: allows researchers to apply constraints on nested and referenced resources, enabling more precise cohort selection (e.g., selecting patients based on linked diagnostic reports).
9. **Collaboration & Reusability**: researchers can reuse query definitions, improving efficiency and standardizing cohort selection across multiple studies.

Development status
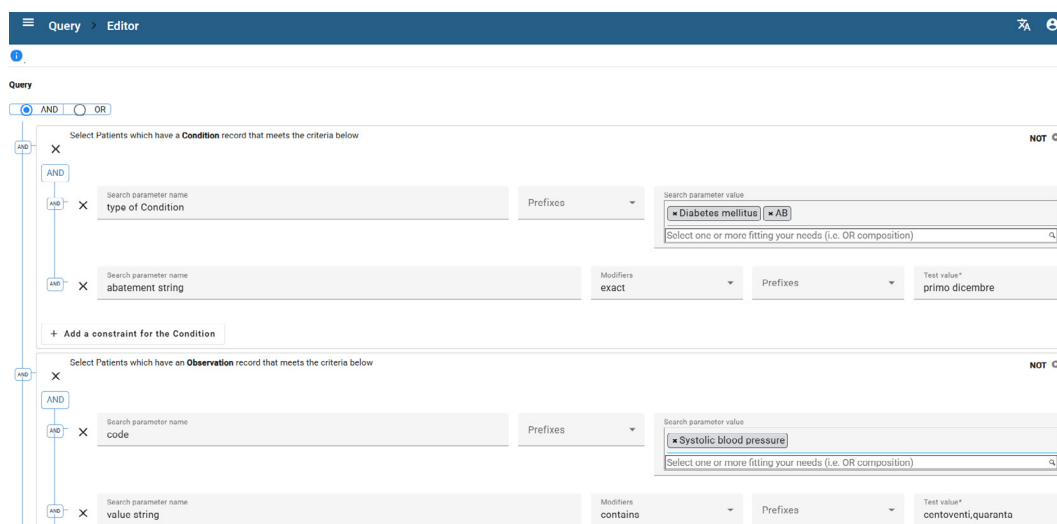
The component has been implemented and alpha-tested.

It is available for registered beta-users.

A demonstration is available on demand.
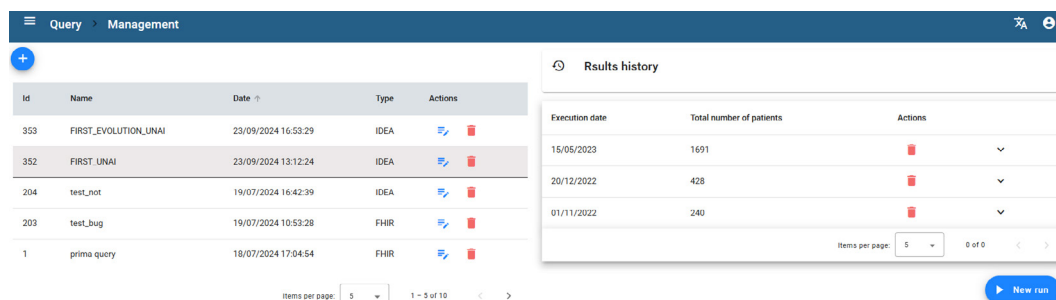
The product is expected to be market available in a 2-years horizon.

Imagery

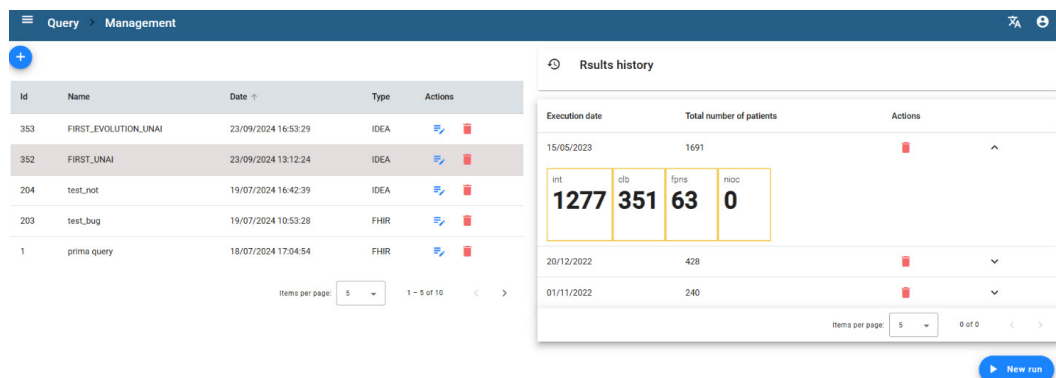Query editor, showing the support of ontology trees and Boolean logic

Result history overview, showing the cardinality of the selected cohorts

Result history details, showing the number of patients retrieved from each different healthcare centres

IDEA4RC