

Connecting the dots in radiation oncology data:

RadPlanBio use case

Olga Giraldo¹, Thomas Früchtel¹, Mareike Roscher², Rosemarie Euler-Lange², Michael Baumann^{1,3,4}, Ina Kurth^{1,2,3}, Wahyu Wijaya Hadiwikarta^{1,3}

¹Division of Radiooncology/Radiobiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany, ²Service Unit for Radiopharmaceuticals and Preclinical Studies, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany, ³German Cancer Consortium (DKTK), Core Center Heidelberg, 69120 Heidelberg, Germany, ⁴Heidelberg Institute of Radiation Oncology (HIRO), 69120 Heidelberg, Germany

dkfz.

GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION

Background

Radiation oncology research generates heterogeneous datasets. These datasets are often stored in spreadsheets, laboratory systems, or local databases using inconsistent terminology and structures, limiting interoperability, reuse, and cross-project integration.

To address this, we provide an infrastructure that supports:

- Standardisation of radiation oncology and radiobiology data
- Ontology-based semantic modelling
- Integration into an ontology-driven Knowledge Graph (KG)
- Translation of plain text into SPARQL by using Large Language Models (LLMs)
- Multi-institutional research workflows through ongoing collaborations with KITZ, PCTU, and MDC.

This enables structured, machine-readable, and FAIR-compliant research data.

WHAT IS IT NOT INTENDED FOR?

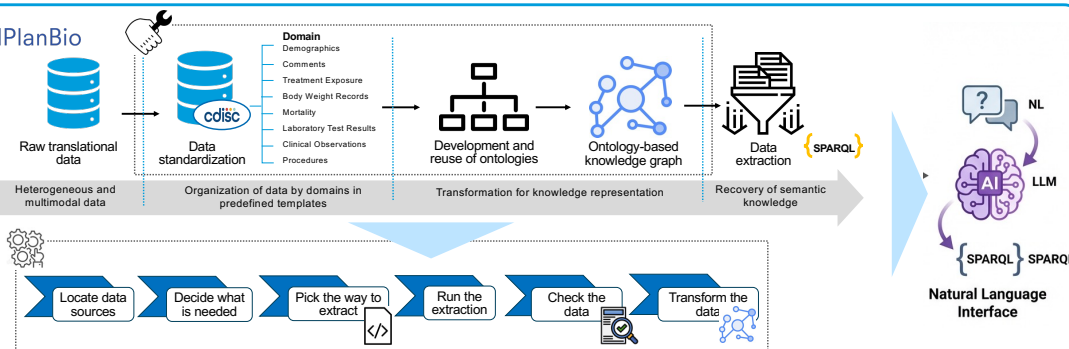
To ensure clarity, this infrastructure does not initially:

- Replace laboratory data management systems (ELN/LIMS)
- Automatically clean unstructured data without domain input
- Fully automate ontology mapping without expert validation
- Replacing human oversight in query formulation – LLMs assist with translating and interpreting plain-text questions into SPARQL

🎯 We focus on semantic structuring and integration, exploring how LLMs can convert natural-language questions into SPARQL queries, enabling researchers to interact intuitively with KGs without requiring expertise in semantic query languages.

Our infrastructure is evolving from manual operations toward semi-automated data extraction, integrating LLMs to enhance access to FAIR biomedical data.

RadPlanBio



WHAT IS THIS INFRASTRUCTURE INTENDED FOR?

This infrastructure supports researchers in:

Standardising preclinical and clinical data

- Harmonising heterogeneous datasets
- Defining consistent metadata structures
- Aligning terminology using controlled vocabularies

Ontology-based data representation

- Mapping experimental entities (demographics, treatments, clinical assays, etc)
- Structuring data using domain ontologies
- Enabling semantic interoperability

Integration into a knowledge graph

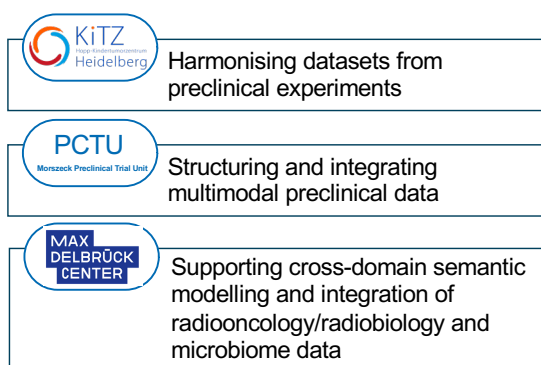
- Connecting datasets across projects and institutions
- Representing relationships between biological entities
- Supporting complex semantic queries
- Enabling AI/ML- ready structured data

LLM-supported natural language interaction

- Translating natural-language questions into SPARQL queries
- Enabling intuitive access to KGs without SPARQL expertise

CURRENT COLLABORATIONS

Our infrastructure is being actively adapted to support ongoing research initiatives, including:



These collaborations help us in refining and standardizing metadata, expand ontologies and validate LLM-supported SPARQL generation across biomedical domains.

ACKNOWLEDGMENTS

- Radiooncology/Radiobiology E220 – DKFZ

Sources:

Ontology for Preclinical Trials in Radiation Oncology:

<https://github.com/DKFZ-E220/PTRO>

Contact:

olga.giraldo@dkfz-heidelberg.de
w.hadiwikarta@dkfz-heidelberg.de

EUCAIM
CANCER IMAGE EUROPE

HMC
HELMHOLTZ METADATA COLLABORATION

RadPlanBio

PCTU
Morzeck Preclinical Trial Unit

KITZ
Heidelberg

MAX
DELBRÜCK
CENTER

DKTK German Cancer Consortium