

DS01

FAIRification of preclinical data in radiation oncology through data standardization and formalized representation

PHUSE EU Connect 2024
12th November, 2024

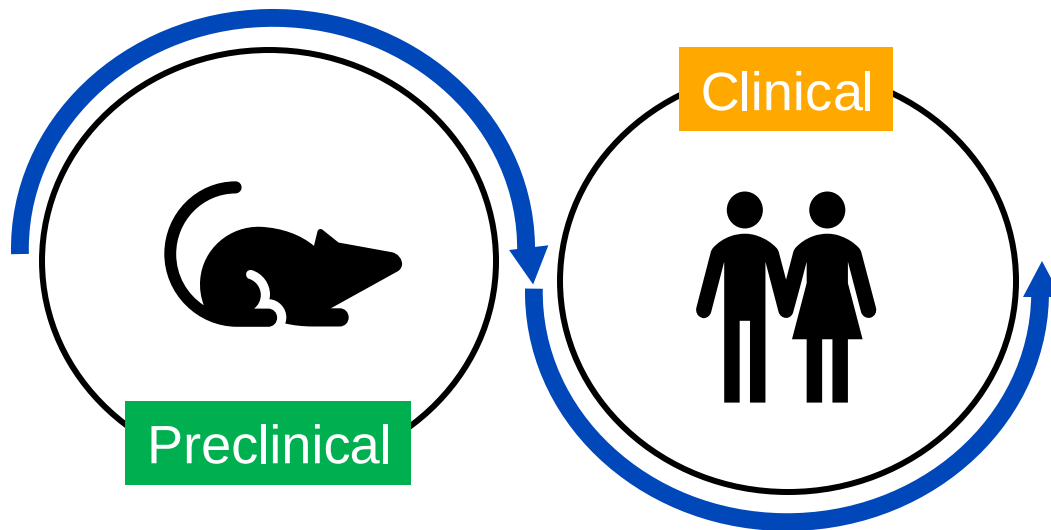
Olga Giraldo
Wahyu Wijaya Hadiwikarta

Agenda

- Background
- Goal
- Methodological Approach
- Results
- Conclusion and Next Steps

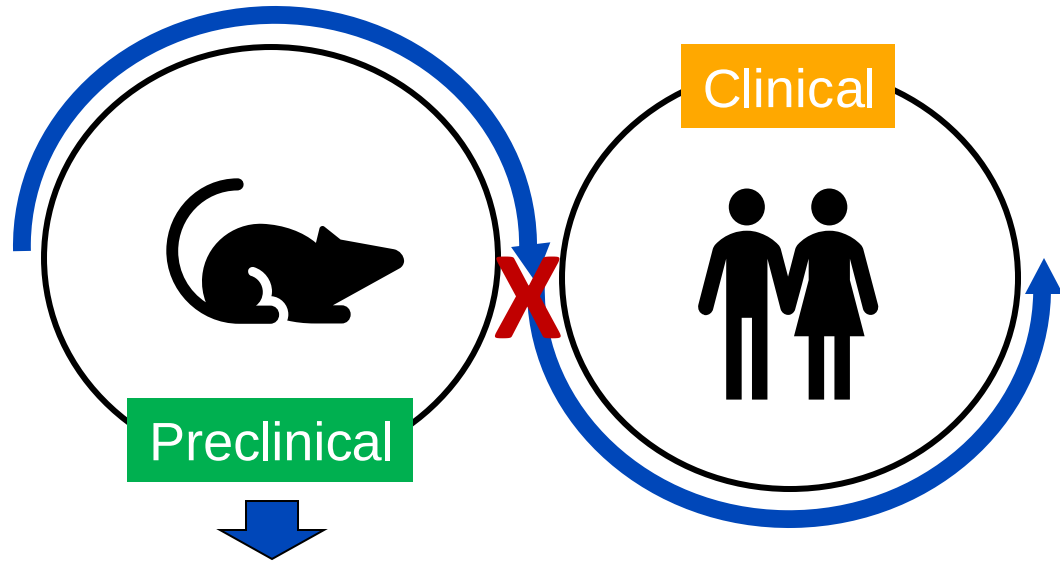
Background

Failures in translating
preclinical treatments
into clinic



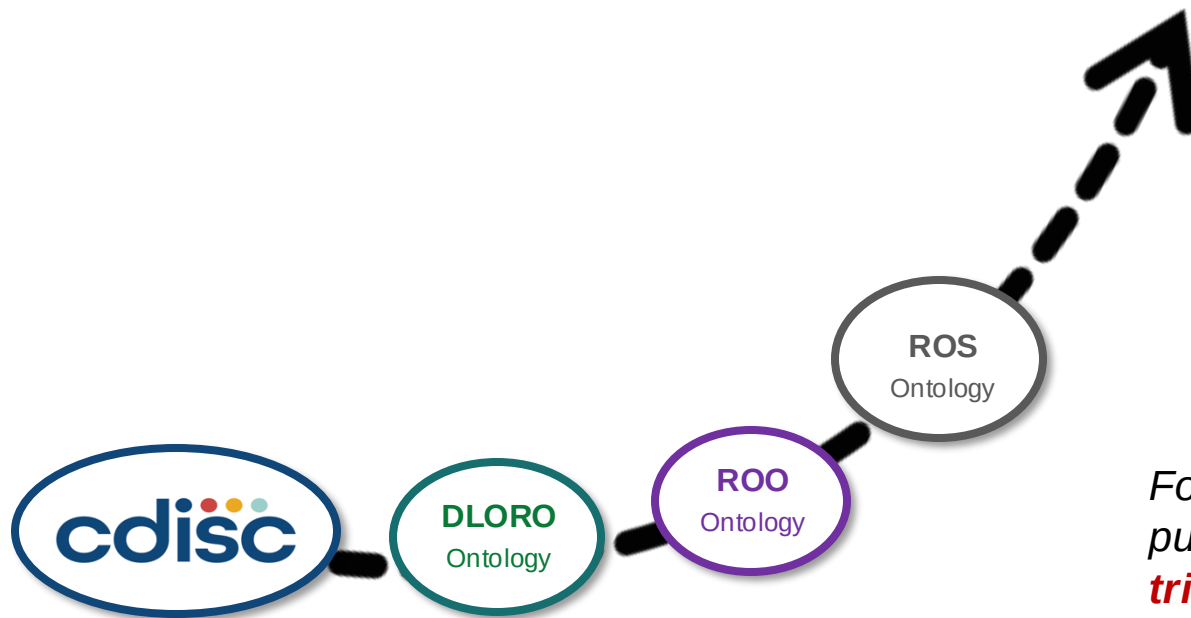
Background

Failures in translating
preclinical treatments
into clinic

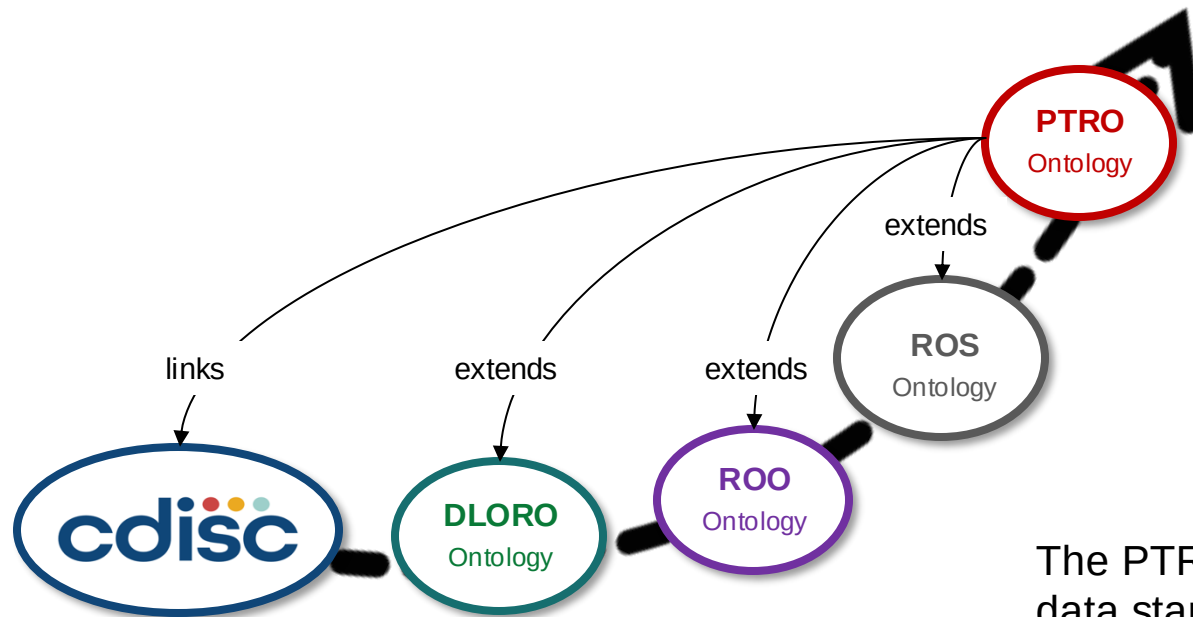


Lack of Reproducibility

- Bias in reporting
- Diverse experimental procedures
- Lack of standardized data collection



*For reproducibility purposes **preclinical trials** serve as initial validation for treatment effectiveness.*

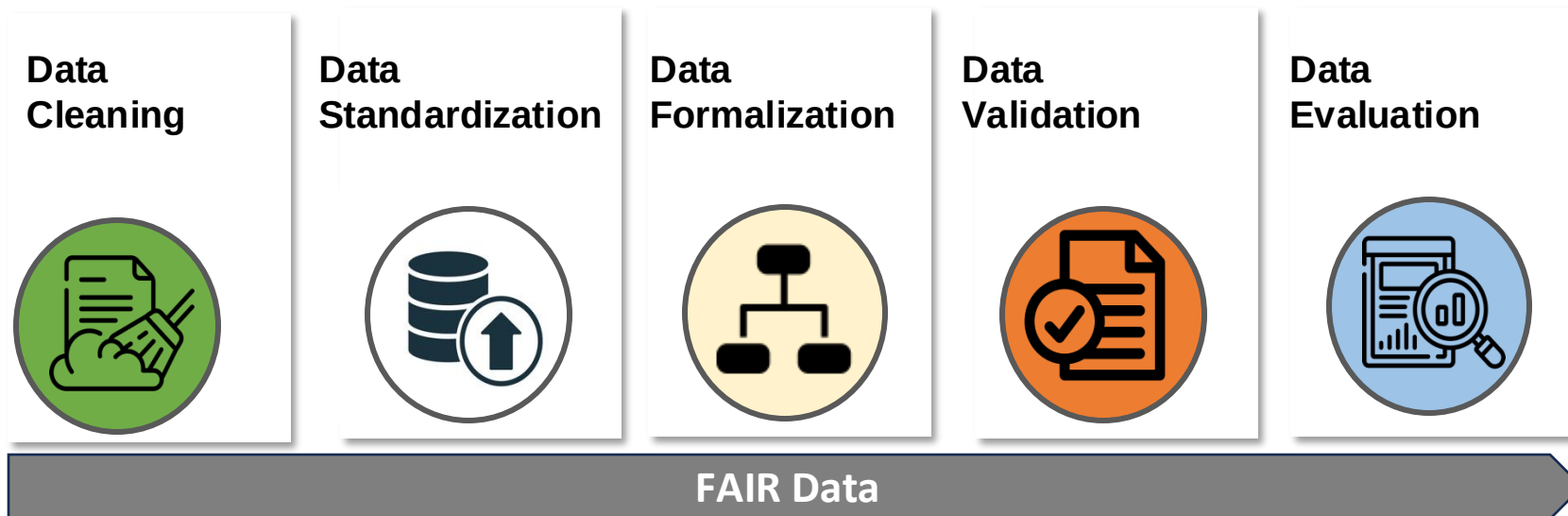


The PTRO ontology can link data standardized with SEND and has the potential to expand existing biomedical and clinical ontologies.

Main Goal

“Standardize and formalize preclinical data in radiation oncology studies to facilitate alignment with FAIR principles”

Methodological Approach



Manual digitalization of data from paper forms about:

- Project overview of the tumour model (treatment arms)
- List of the mouse numbers
- Data of Measurements (tumour size, weight, etc.)
- Treatment schemes (irradiation and chemotherapy details)



Correction of errors and inconsistencies.

The digitized data was organized and structured according the SEND standard for the tabulation of nonclinical datasets

Key features:

- The best-known standard for representing preclinical data
- Enhanced regulatory compliance and review efficiency



Data is part of the RadPlanBio platform operated under the German Cancer Research Center (DKFZ), in Heidelberg.

Paper forms

experiment:	N150a	tumor from:	full TP + 9GT	responsible scientist:	LK	page:	1	working group:	Bureau	
animal no.	date of birth	sex	date of WBL date of TP	15.12.11 30	15.12.11 30	22.12.11 40	26.12.11 60	27.12.11 70	28.12.11 70	31.12.11 70
1	23.3.11	♂	15.12.11 TP 6.12.11	30	30	30	30	30	30	30
2				30	30	30	30	30	30	30
3				30	30	30	30	30	30	30
4				30	30	30	30	30	30	30
5				30	30	30	30	30	30	30
6				30	30	30	30	30	30	30

Digitization and standardization process

Challenge

- Labor-Intensive and Time-Consuming

SEND-standardized data

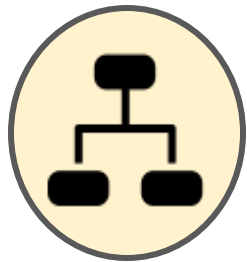
[illegible]

The SEND-standardized data is represented in the Ontology for Preclinical Trials in Radiation Oncology (PTRO).

This ontology builds on a previous ontology model¹

Development process stages:

- Collection of preclinical concepts available in RadPlanBio
- Semantic analysis of existing vocabularies
- Ontology extension

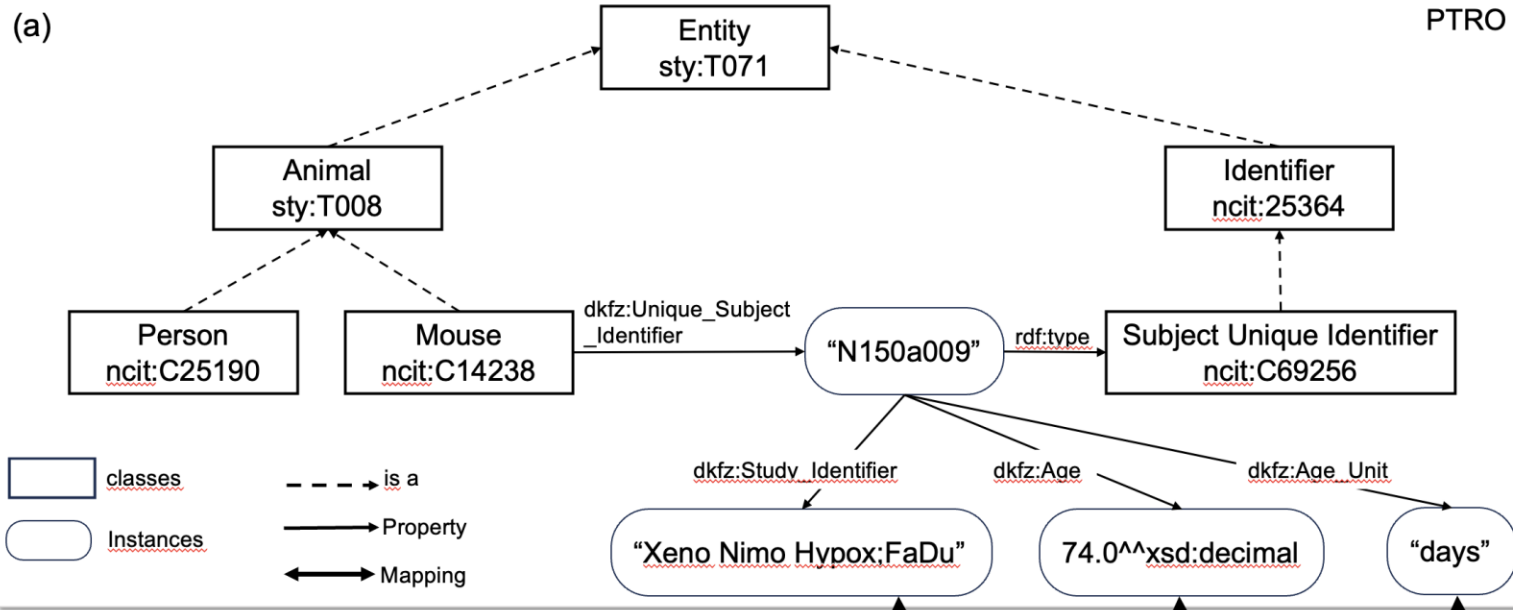


1. Olga Giraldo, Abumansur Sabyrrakhim, Mareike Roscher et al. Semantic Representation of Preclinical Data in Radiation Oncology., 03 November 2024, PREPRINT (Version 1) available at Research Square [<https://doi.org/10.21203/rs.3.rs-5373454/v1>]

Verify if the ontology can represent and capture the knowledge and data from the preclinical databases.

- Formulating SPARQL queries based on competency questions





(b)

Preclinical database schema
(Mouse information)

Subject Id	Study identifier	Age	Age Unit
N150a009	Xeno Nimo Hypox:FaDu	74.0	days
...			



Results

RadPlanBio

Studies



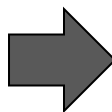
Tumour
volume
measurements



Biomarker
expression



Treatment
protocols



SENDIG v3.1.1



Animal demographics



Animals' treatment exposures



Records of animal body weights



Causes of mortality



Laboratory results



Clinical observations



Procedures



Observational notes and relevant
remarks

- Metrics
 - 958 classes
 - 55 data properties
 - 8 object properties
- We import the Semantic Types Ontology (STY) as a framework for broad concept categorization
- Reuse terminology that come primarily from the National Cancer Institute Thesaurus (NCIt)
- Use OWL format
- Available on GitHub



- “Retrieve all male mice of a specific age (74 days)”

```
PREFIX ptro: <http://www.radplanbio.org/ptro/>
PREFIX ncit:
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#>
PREFIX uo: <http://purl.obolibrary.org/obo/UO>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?Subject_Unique_Identifier

WHERE
{
  ?Subject_Unique_Identifier ptro:DKFZ000040 ?Sex .
  FILTER (?Sex = 0).
  ?Subject_Unique_Identifier ptro:DKFZ000033 ?Age .
  FILTER (?Age = 74.0)
}
```

SPARQL query



?Subject_Unique_Identifier
ptro:instanceDKFZ_N150a009

Result



Knowledge Captured in the Ontology

- “Retrieve the mice with a death diagnosis result **terminal sacrifice** and exposed to a treatment with **cisplatin**”

SPARQL query

```
PREFIX ptro: <http://www.radplanbio.org/ptro/>
PREFIX ncit: <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#>
PREFIX uo: <http://purl.obolibrary.org/obo/UO>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?Subject_Unique_Identifier ?Treatment_Session_Identifier
?Name_of_Actual_Treatment ?Death_Diagnosis_Result_Category

WHERE
{
  ?Subject_Unique_Identifier ptro:DKFZ000077
  ?Treatment_Session_Identifier .
  ?Treatment_Session_Identifier ptro:DKFZ000013
  ?Name_of_Actual_Treatment .
  FILTER (?Name_of_Actual_Treatment = "CISPLATIN")
  ?Subject_Unique_Identifier ptro:DKFZ000021
  ?Death_Diagnosis_Result_Category .
  FILTER (?Death_Diagnosis_Result_Category = "TERMINAL SACRIFICE")
}
```



Result

?Subject_Unique_Identifier	?Treatment_Session_Identifier	?Name_of_Actual_Treatment	?Death_Diagnosis_Result_Category
ptro:instanceDKFZ_N150a009	ptro:instanceDKFZ_N150a009_EX_20.0_2	CISPLATIN^^xsd:string	TERMINAL SACRIFICE^^xsd:string
ptro:instanceDKFZ_N150a011	ptro:instanceDKFZ_N150a011_EX_13.0_2	CISPLATIN^^xsd:string	TERMINAL SACRIFICE^^xsd:string

RDF triples produced for the first Subject Id of the database.

```
<http://www.radplanbio.org/ptro/instance/N150a009> <http://www.radplanbio.org/ptro/DKFZ000033> "74.0".  
<http://www.radplanbio.org/ptro/instance/N150a009> <http://www.radplanbio.org/ptro/DKFZ000034> "days".  
<http://www.radplanbio.org/ptro/instance/N150a009> <http://www.radplanbio.org/ptro/DKFZ000040> "0".  
<http://www.radplanbio.org/ptro/instance/N150a009> <http://www.radplanbio.org/ptro/DKFZ000041> "Nude Mouse".  
<http://www.radplanbio.org/ptro/instance/N150a009> <http://www.radplanbio.org/ptro/DKFZ000010> "9".  
<http://www.radplanbio.org/ptro/instance/N150a009> <http://www.radplanbio.org/ptro/DKFZ000009> "N150a009".  
<http://www.radplanbio.org/ptro/instance/N150a009> <http://www.radplanbio.org/ptro/DKFZ000008> "Xeno Nimo Hypox,FaDu".
```

Database transformed into CSV

Subject Id	Study Identifier	Subject Identification	Age	Age Unit	Sex	Species	Strain/Substrain
N150a009	Xeno Nimo Hypox,FaDu	9	74.0	days	0	Mouse	Nude Mouse
N150a011	Xeno Nimo Hypox,FaDu	11		days	0	Mouse	Nude Mouse

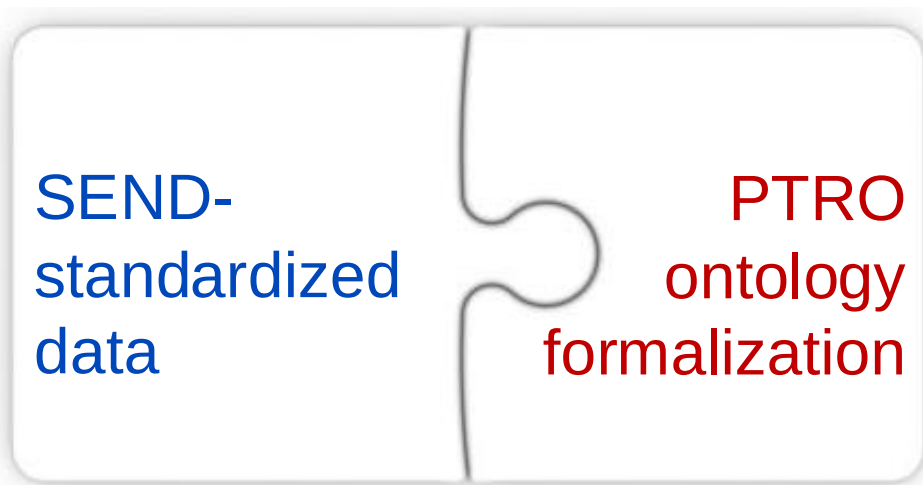
```
<http://www.radplanbio.org/ptro/instance/N150a011> <http://www.radplanbio.org/ptro/DKFZ000034> "days".  
<http://www.radplanbio.org/ptro/instance/N150a011> <http://www.radplanbio.org/ptro/DKFZ000043> "72-74".  
<http://www.radplanbio.org/ptro/instance/N150a011> <http://www.radplanbio.org/ptro/DKFZ000040> "0".  
<http://www.radplanbio.org/ptro/instance/N150a011> <http://www.radplanbio.org/ptro/DKFZ000041> "Nude Mouse".  
<http://www.radplanbio.org/ptro/instance/N150a011> <http://www.radplanbio.org/ptro/DKFZ000010> "11".  
<http://www.radplanbio.org/ptro/instance/N150a011> <http://www.radplanbio.org/ptro/DKFZ000009> "N150a011".  
<http://www.radplanbio.org/ptro/instance/N150a011> <http://www.radplanbio.org/ptro/DKFZ000008> "Xeno Nimo Hypox,FaDu".
```

RDF triples produced for the second Subject Id of the database.

1. Olga Giraldo, Abumansur Sabyrrakhim, Mareike Roscher et al. Semantic Representation of Preclinical Data in Radiation Oncology., 03 November 2024, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-5373454/v1]



Summary



- Fosters a more FAIR-compliant preclinical data ecosystem
- Allow translation of preclinical findings to clinical applications
- Supports personalized treatment approaches

Next Steps

- Continue the population, validation and evaluation of the PTRO ontology to ensure robustness
- Develop a knowledge graph (KG) utilizing the PTRO ontology and integrate it into the semantic layer of the RadPlanBio platform
 - To support complex semantic querying, reasoning, and inference capabilities
- Combine KG + LLM

Acknowledgments

- RadioOncology/RadioBiology E220 - DKFZ

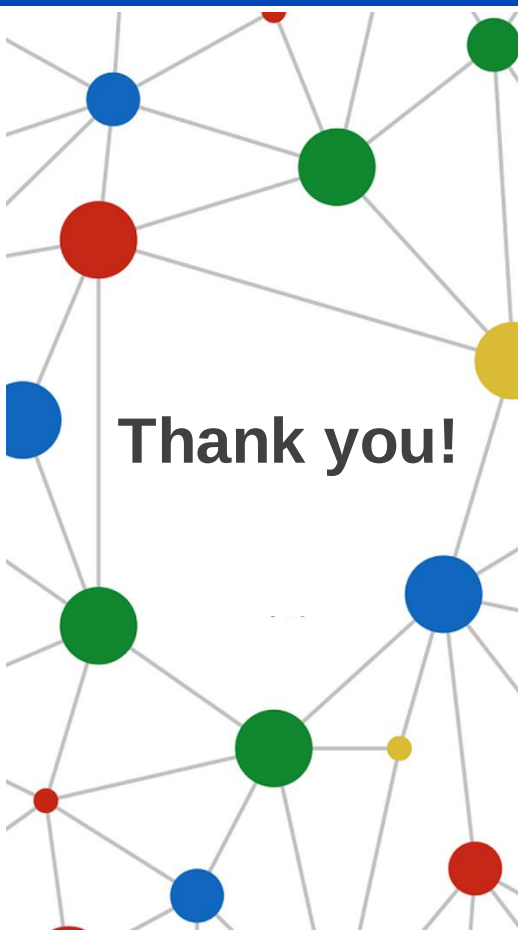
Wahyu Wijaya Hadiwikarta, Ina Kurth, Mareike Roscher, Abumansur Sabyrrakhim, Rosemarie Euler-Lange, Thomas Früchtel, Myta Pristanty, Betül Çakir and Michael Baumann.

- Funding

Research supported in part by the European Federation for Cancer Images (EUCAIM), Project ID: 101100633.

Research supported in part by the DKTK funding for the operation of RadPlanBio platform in DKFZ.





Thank you!

- How have you navigated challenges in representing and standardizing preclinical data?
- What technologies or strategies have worked well in your experience?
 - Specifically, how have these technologies impacted data accessibility and interoperability?