

# **ECOLOGICAL RELATIONS BETWEEN MEMBERS OF THE MICROBIOME**

**Presented by  
Virginia Leombruni,  
Eleonora Giuliani,  
Marc Shebaby**



# Introduction



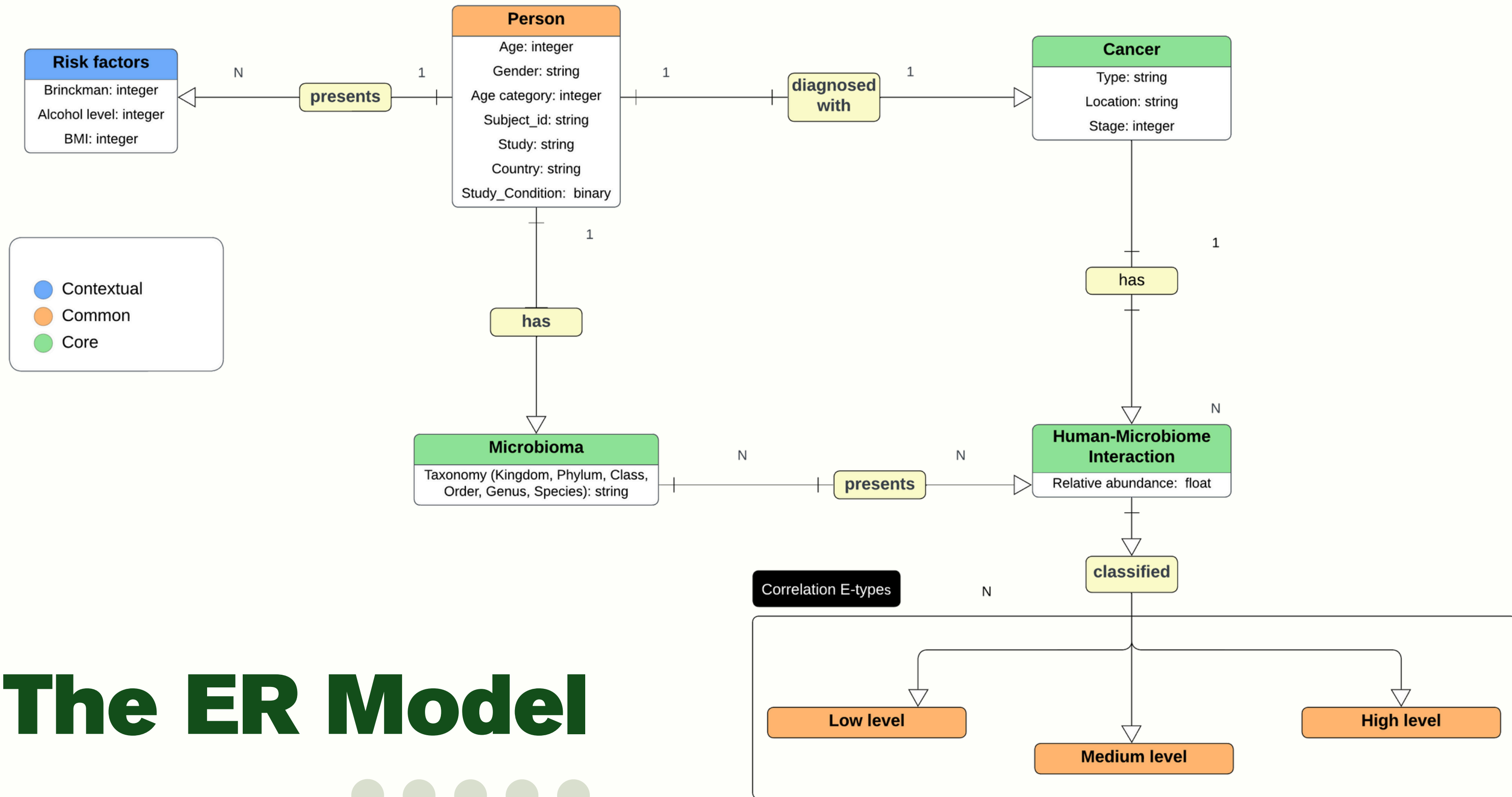
The **purpose** is to construct a Knowledge Graph (KG) to investigate the interactions between microorganisms and their associations with diseases.

## Scenario Definition:

- **Species Interactions (S1):** Ecological relationships between species and their impact on microbial community stability.
- **Microbiome Epidemiology (S2):** How different factors influence microbial profiles and disease outcomes.
- **Microbiome Risk Factors Analysis(S3):** How risk factors affect the microbiome.



# The ER Model





# Information Gathering



**Objective:** Standardize and prepare data for analysis.

## Main Sources:

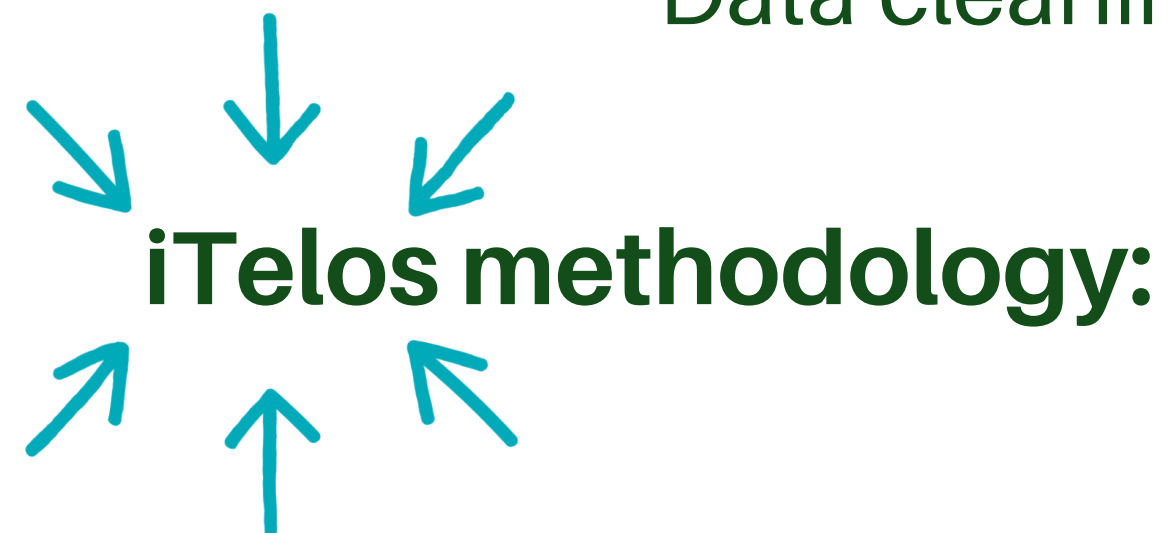
- MetaPhlAn3: Taxonomic abundances
- HUMAnN3: Functional metabolic potential

## Chosen Dataset:

YachidaS\_2019 dataset

## Process:

Data cleaning and creation of a CSV file.



Enhances reusability and supports future applications.

# Language Definition



## Problem:

Ambiguity from polysemy of words  $\longleftrightarrow$

## Solution:

Use UKC to define concepts

## Mapping:

Concepts mapped to UKC ontologies and BioPortal;  
New IDs for specific biological terms (e.g., Brinkman Index)

**Advantage:** Resources already aligned, no further filtering needed

ConceptID	Word-en	Gloss-en
UKC-67961	Cancer	Any malignant growth or tumor caused by abnormal and uncon trolled cell division
OHMI_0000003	Microbiome	A biome that consists of a collection of microorganisms (i.e., microbiota)
KGE_QCB1-1	Risk factor	Something that makes a person more likely to get a particular disease or condition

# Knowledge Definition



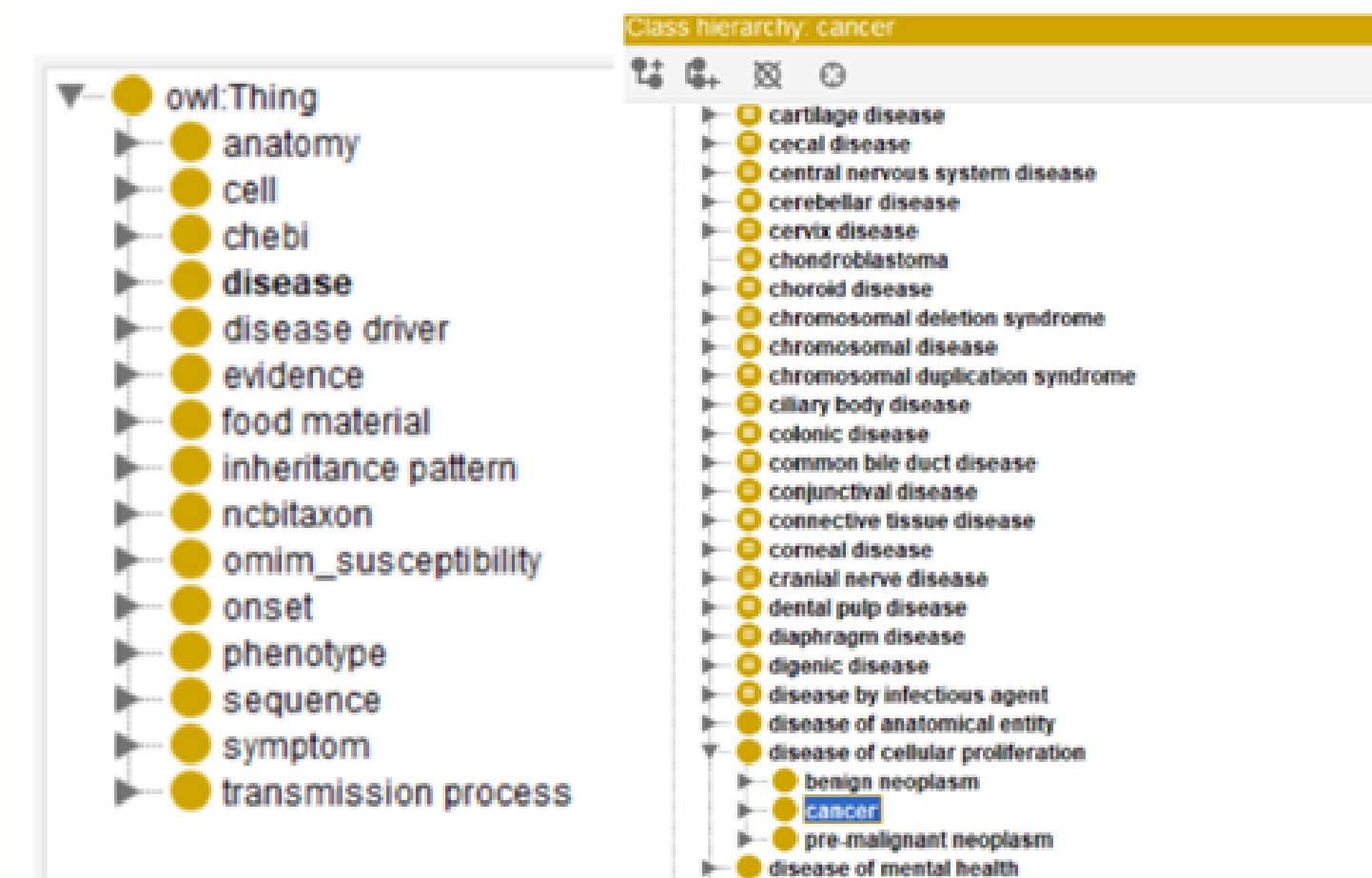
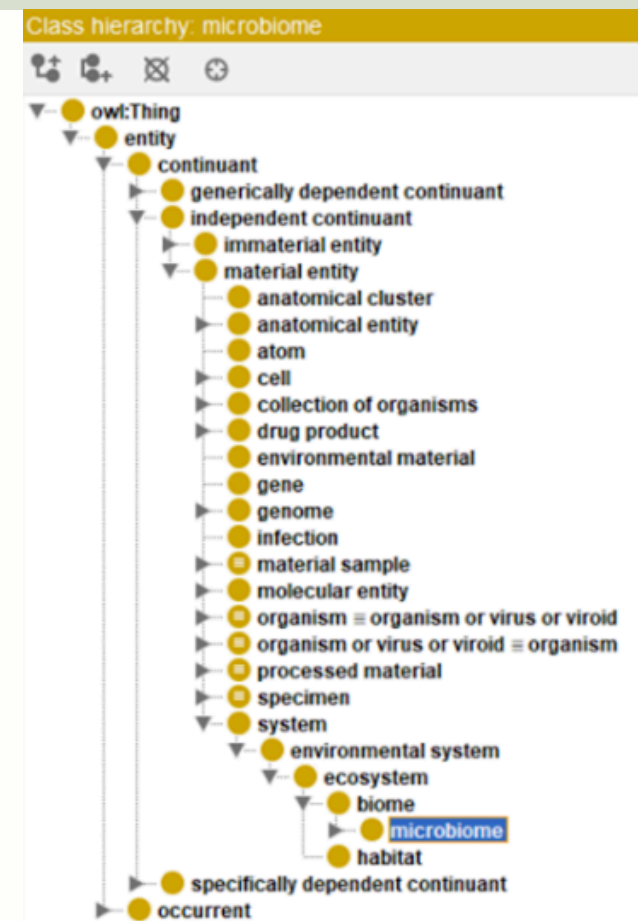
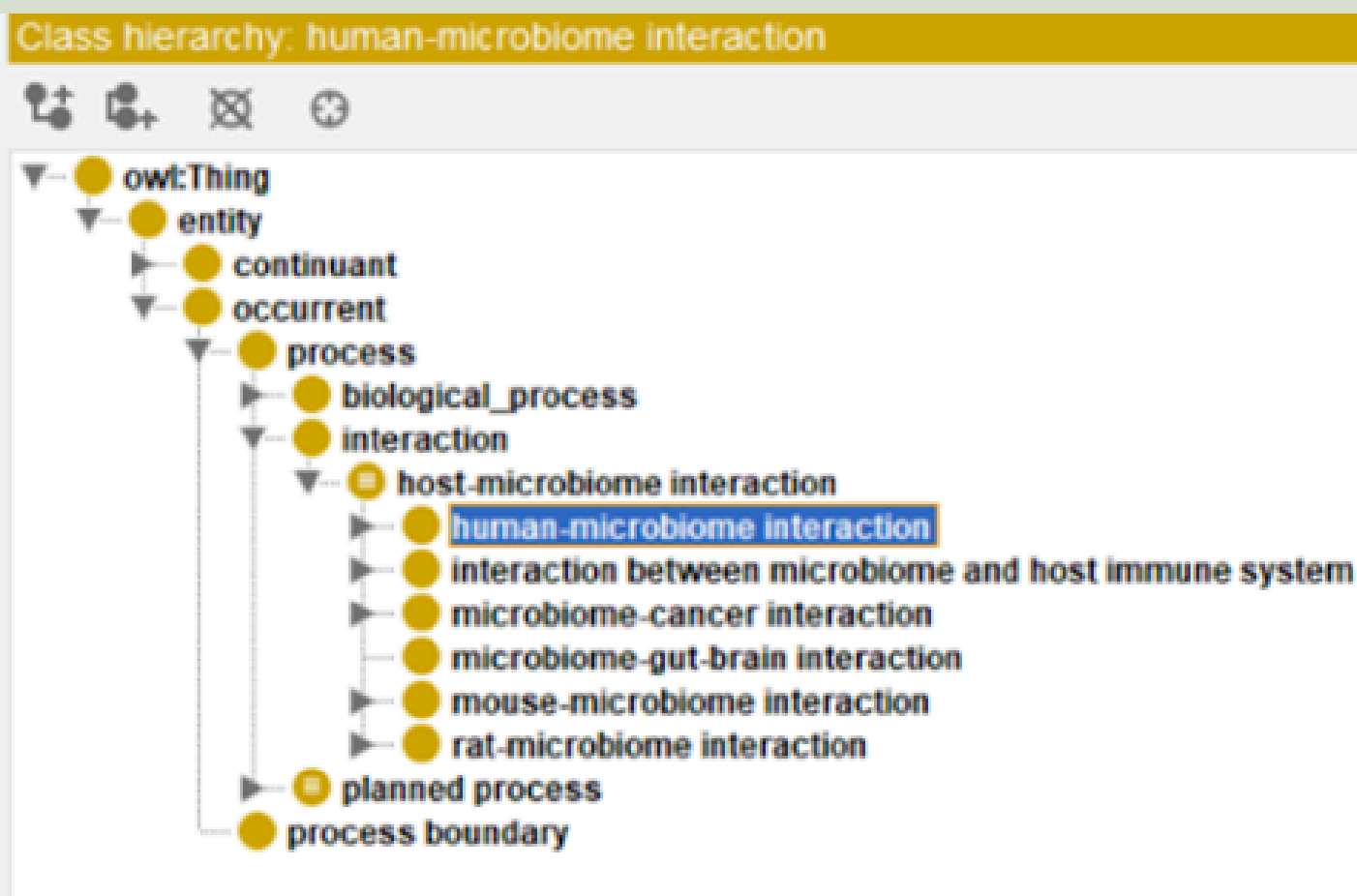
## Goal

Develop a teleontology for the final KG to enhance interoperability and reusability

### 1. Producer Side: →

Create interoperable ontologies for datasets:

- **OHMI**: Host-Microbiome Interactions
- **DOID**: Disease Ontology (cancer focus)



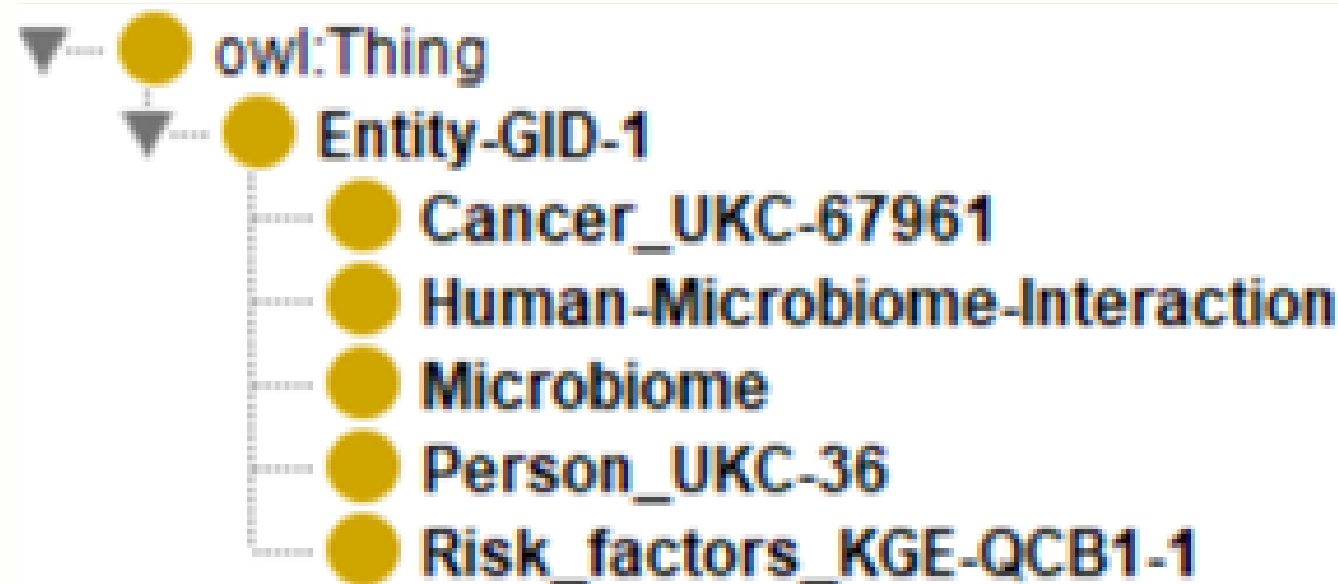
# Knowledge Definition



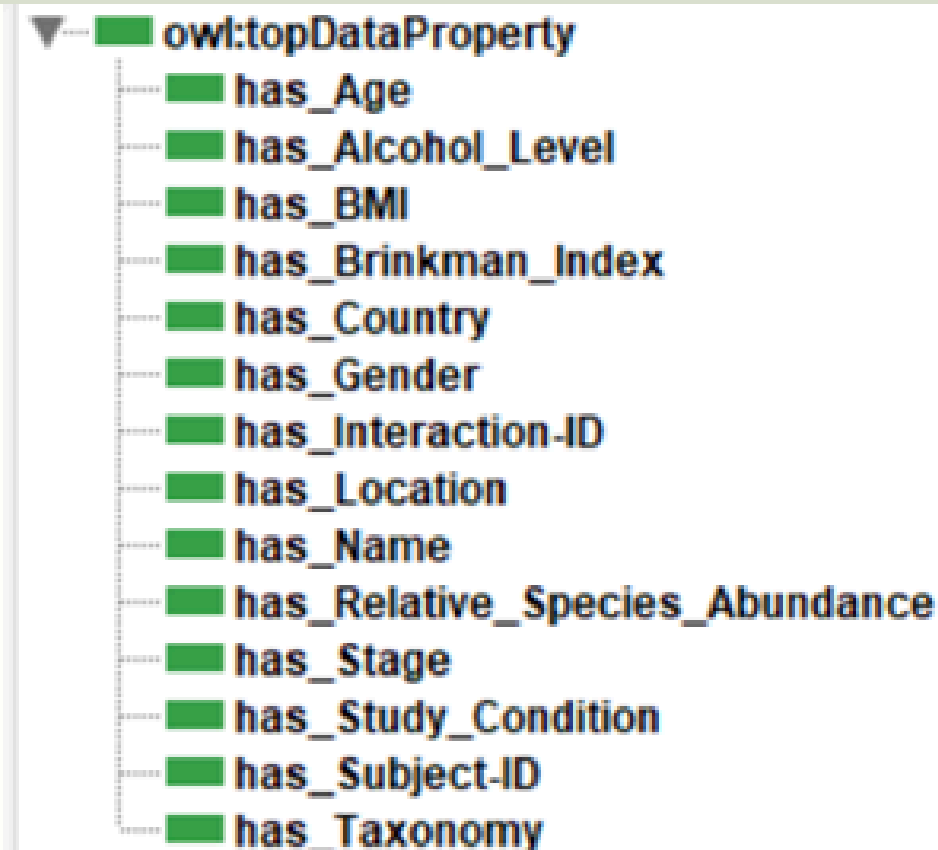
## 2. Consumer Side



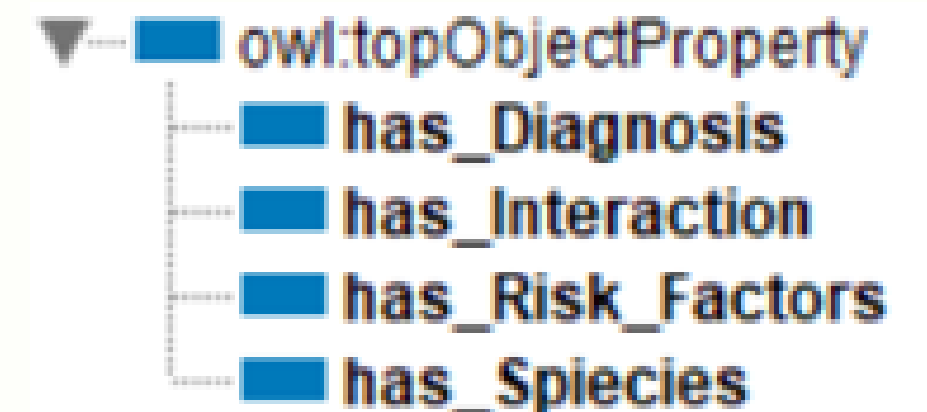
Build a unified **ontology** in Protégé:  
Class hierarchies, object & data properties



Classes



Data Properties



Object Properties

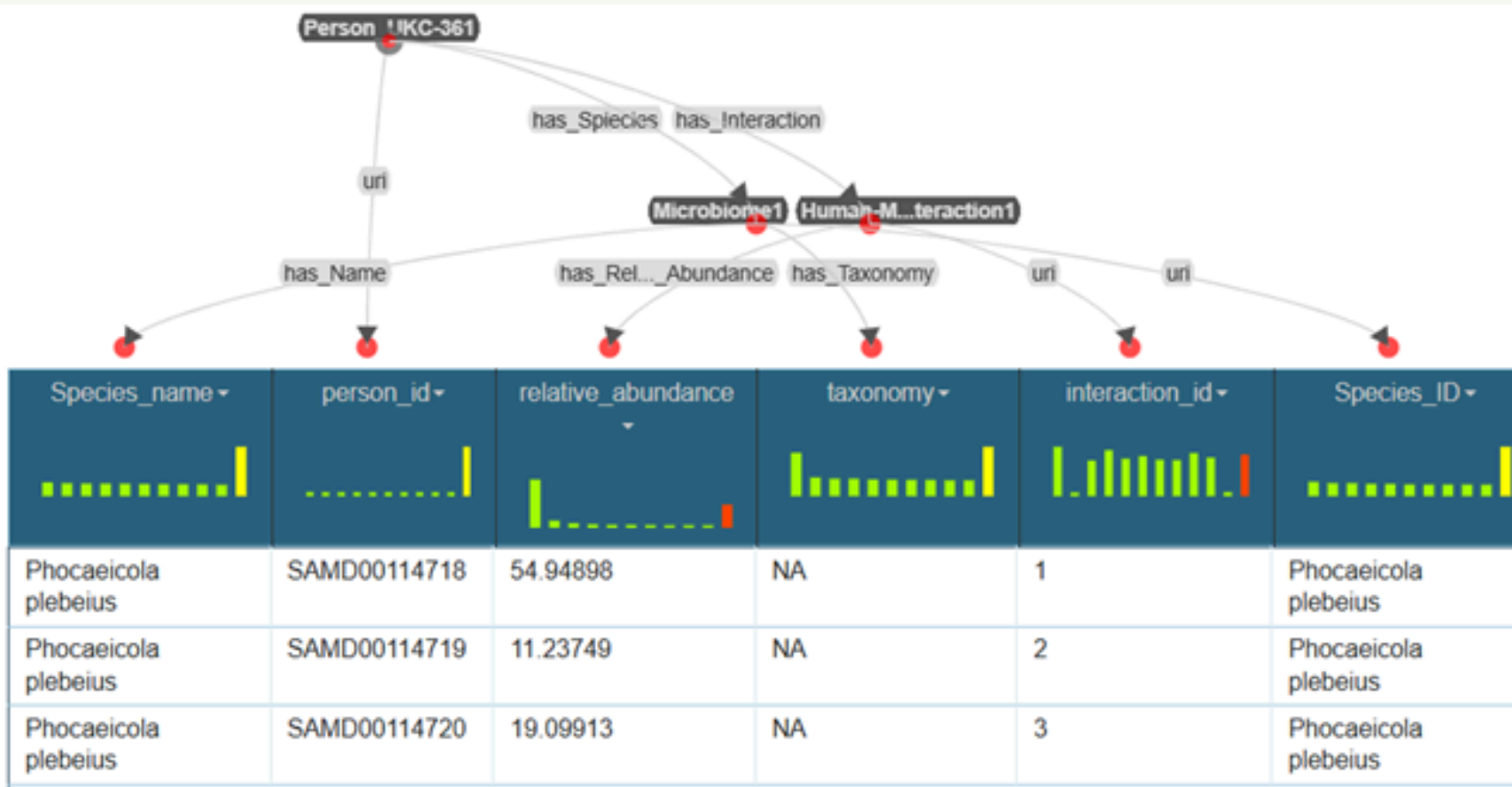
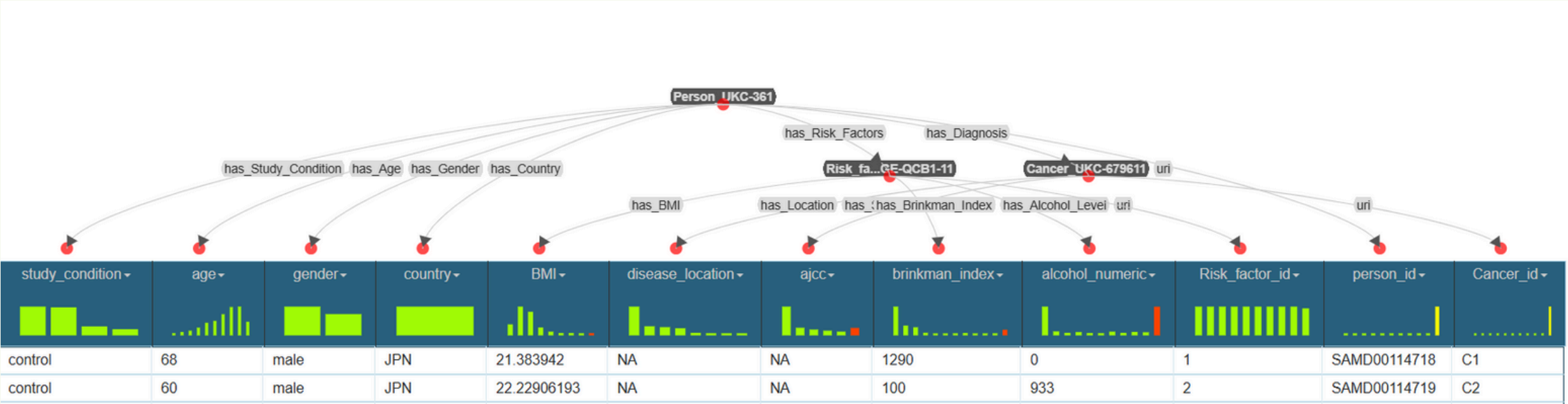




# Entity Definition

## Objective:

- **Merging** the knowledge and the data layers into a single structure
- Handling data value **heterogeneity** through entity matching and entity identification





# Evaluation



## Teleontology vs CQs

## Teleontology vs Reference Ontologies

Entity  
level

$$\text{Cov}_E(\text{CQ}_E) = \frac{|\text{CQ}_E \cap \text{T}_E|}{\text{CQ}_E} = \frac{5}{5} = 1$$

$$\text{Cov}_E(\text{RO}_E) = \frac{|\text{RO}_E \cap \text{T}_E|}{\text{RO}_E} = \frac{3}{19699} = 0.0001$$

Property  
level

$$\text{Cov}_p(\text{CQ}_p) = \frac{|\text{CQ}_p \cap \text{T}_p|}{\text{CQ}_p} = \frac{18}{19} = 0.94$$

$$\text{Cov}_p(\text{RO}_p) = \frac{|\text{RO}_p \cap \text{T}_p|}{\text{RO}_p} = \frac{18}{178} = 0.1$$

# Exploitation



**QUERY 1:** CQs about which microbiome is associated with stage 4 cancer and high risk factors

person	alcohol	cig_level	species	occurrence_of_species	mean_of_species	stage
tp://localhost:8080/ source/SAMD00114750	"638.786"	"820"	http://localhost:8080/ source/ Bacteroides%20uniformis	"12"^^xsd:integer	"11.230103"^^xsd:float	"iv"
tp://localhost:8080/ source/SAMD00114803	"1899"	"630"	http://localhost:8080/ source/ Bacteroides%20uniformis	"12"^^xsd:integer	"11.230103"^^xsd:float	"iv"
tp://localhost:8080/ source/SAMD00114810	"348"	"640"	http://localhost:8080/ source/ Bacteroides%20uniformis	"12"^^xsd:integer	"11.230103"^^xsd:float	"iv"
tp://localhost:8080/ source/SAMD00114817	"360"	"780"	http://localhost:8080/ source/ Bacteroides%20uniformis	"12"^^xsd:integer	"11.230103"^^xsd:float	"iv"
tp://localhost:8080/ source/SAMD00114750	"638.786"	"820"	http://localhost:8080/ source/ Eubacterium%20rectale	"11"^^xsd:integer	"6.4065185"^^xsd:float	"iv"
tp://localhost:8080/ source/SAMD00114810	"348"	"640"	http://localhost:8080/ source/ Eubacterium%20rectale	"11"^^xsd:integer	"6.4065185"^^xsd:float	"iv"
tp://localhost:8080/ source/SAMD00114817	"360"	"780"	http://localhost:8080/ source/ Eubacterium%20rectale	"11"^^xsd:integer	"6.4065185"^^xsd:float	"iv"
tp://localhost:8080/ source/SAMD00114750	"638.786"	"820"	http://localhost:8080/ source/ Parabacteroides%20distasonis	"10"^^xsd:integer	"8.460612"^^xsd:float	"iv"

# Exploitation



person	species	relative_value	cig	bmi	status
localhost:8080/ SAMD00114899	http://localhost:8080/ source/ Helicobacter%20pylori	"0.00157"	"0"	"22.18934911"	"control"
localhost:8080/ SAMD00164772	http://localhost:8080/ source/ Helicobacter%20pylori	"0.00337"	"570"	"25.40281608"	"adenoma"
localhost:8080/ SAMD00164834	http://localhost:8080/ source/ Helicobacter%20pylori	"0.01394"	"360"	"22.14532872"	"CRC"
localhost:8080/ SAMD00164893	http://localhost:8080/ source/ Helicobacter%20pylori	"0.00399"	"0"	"18.7961895"	"adenoma"

**QUERY 2:**  
CQs about association of  
Helicobacter Pylori with cancer

person	species	relative_value	status
ttp://localhost:8080/source/SAMD00164889	http://localhost:8080/source/ Bacteroides%20uniformis	"10.04378"	"CRC"
ttp://localhost:8080/source/SAMD00114811	http://localhost:8080/source/ Prevotella%20sp%20CAG5226	"10.12263"	"CRC"
ttp://localhost:8080/source/SAMD00115010	http://localhost:8080/source/ Prevotella%20sp%20CAG520	"10.12896"	"CRC"
ttp://localhost:8080/source/SAMD00114775	http://localhost:8080/source/ Faecalibacterium%20prausnitzii	"10.13096"	"CRC"
ttp://localhost:8080/source/SAMD00164867	http://localhost:8080/source/ Bacteroides%20uniformis	"10.13915"	"CRC"

**QUERY 3:**  
CQs about patterns of relative  
abundance in control vs cancer

person	status	cig	gender	Escherichia_coli_...	Ruminococcus_g...	Abs_Diff_Ecoli_Gn...	Abs_Diff_CRC	Abs_Diff_NonCRC
http://localhost:8080/source/SAMD00114718	"control"	"1290"	"male"	"1.31309"	"2.26875"	"0.9556599855422974"	"1.2582162618637085"	"0.996005892753601"
http://localhost:8080/source/SAMD00114719	"control"	"100"	"male"	"9.1E-4"	"0.07518"	"0.07427000254392624"	"2.5703961849212646"	"2.308185815811152"
http://localhost:8080/source/SAMD00114720	"control"	"1800"	"male"	"0.00525"	"4.14455"	"4.1392998695373535"	"2.566056251525879"	"2.303845882415775"
http://localhost:8080/source/SAMD00114721	"control"	"300"	"male"	"16.3262"	"0.09351"	"16.232690811157227"	"13.754894256591797"	"14.01710510253902"
http://localhost:8080/source/SAMD00114722	"CRC"	"800"	"male"	"11.08987"	"8.04042"	"3.0494508743286133"	"8.518564224243164"	"8.78077507019043"
http://localhost:8080/source/SAMD00114723	"CRC"	"0"	"male"	"0.02001"	"1.48201"	"1.4620000123977661"	"2.5512962341308594"	"2.28908586502075"
http://localhost:8080/source/SAMD00114724	"CRC"	"0"	"female"	"4.41078"	"14.22575"	"9.814970016479492"	"1.8394737243652344"	"2.10168409347534"

**QUERY 3:**  
CQS about association of E.Coli  
with other species

# Open Issues



- Additional data can be added to cover different types of cancer.
- Time-series data provided by future studies can enhance more insights on species-species interactions within the same individual.
- Machine learning and statistical analysis is needed for precise and accurate results obtained from queries.



**THANK YOU FOR THE ATTENTION**

