UNIVERSITÀ DI TRENTO

KGE 2025

ECOLOGICAL RELATIONS BETWEEN MEMBERS OF THE MICROBIOME

Presented by
Virginia Leombruni,
Eleonora Giuliani,
Marc Shebaby

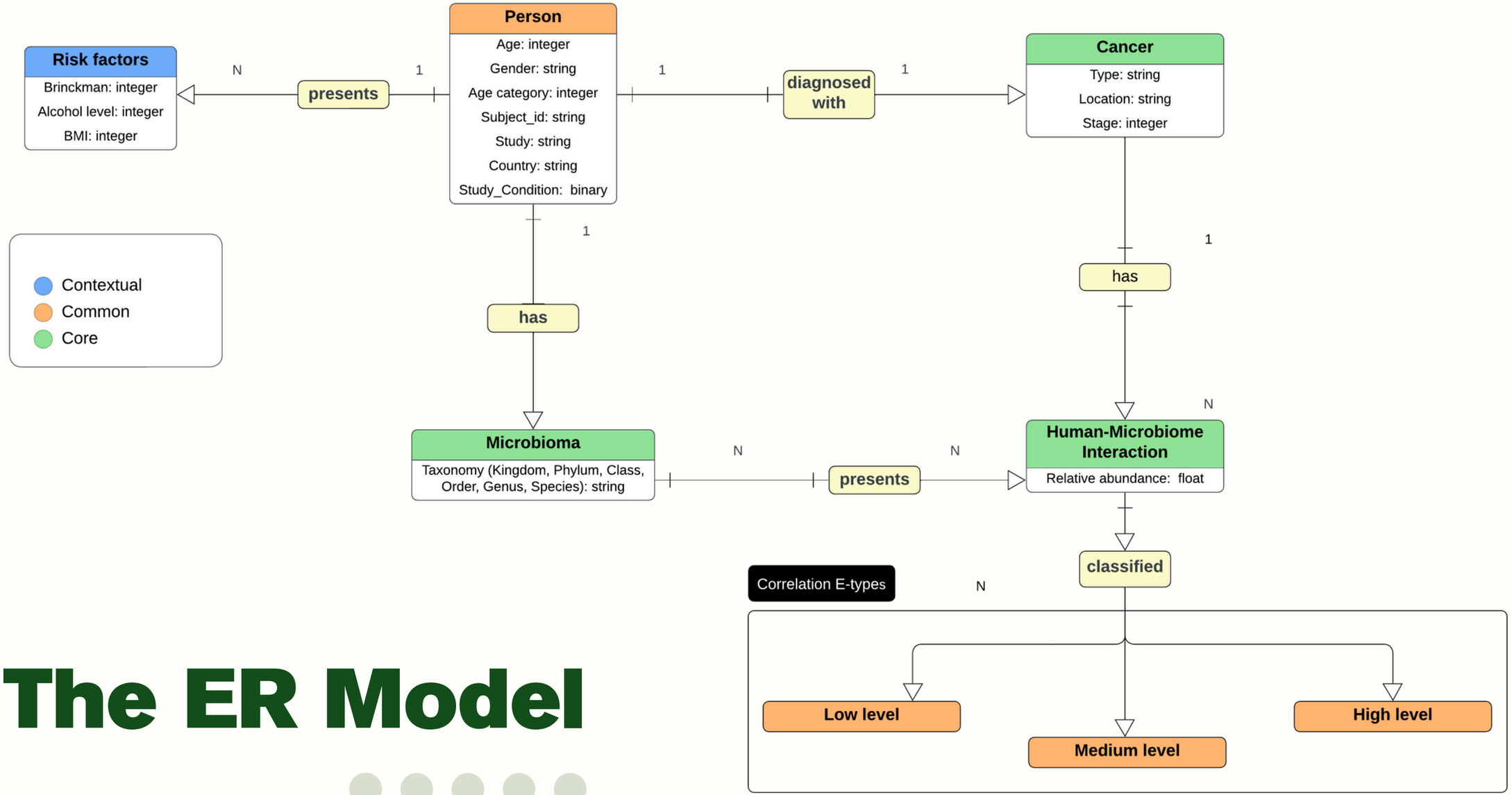
Introduction

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


Person
Age: integer
Gender: string
Age category: integer
Subject_id: string
Study: string
Country: string
Study_Condition: binary
Risk factors
Brinckman: integer
Alcohol level: integer
BMI: integer
presents
diagnosed with
Cancer
Type: string
Location: string
Stage: integer
has
has
Microbioma
Taxonomy (Kingdom, Phylum, Class, Order, Genus, Species): string
presents
Human-Microbiome Interaction
Relative abundance: float
classified
Correlation E-types
Low level
Medium level
High level
Contextual
Common
Core
N
1
1
1
1
1
N
N
N
N

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.

iTelos methodology:

Enhances reusability and supports future applications.

Language Definition

Problem: Ambiguity from polysemy of words ⟷ **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-681	has_Medical_diagnosis	Identification of a disease from its symptoms
KGE-QCBI-2	has_Species	A person has a taxonomic group of sprecies whose memeber can interbreed
KGE-QCBI-3	has_Interaction	A species correlates with a particular person

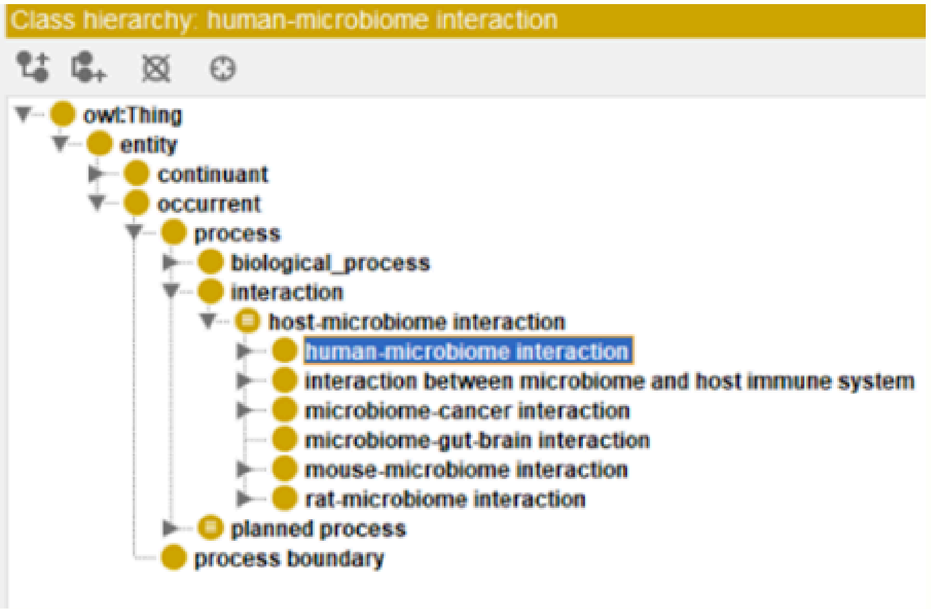
Knowledge Definition

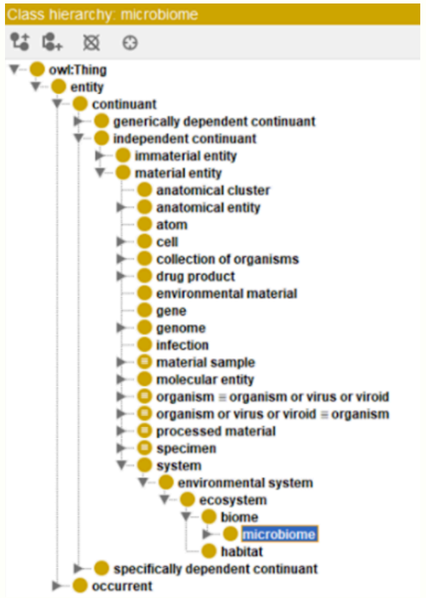
Goal

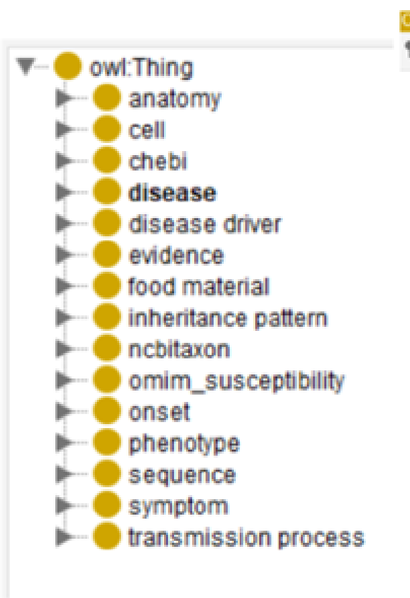
Develop a teleontology for the final KG to enhance interoperability and reusability

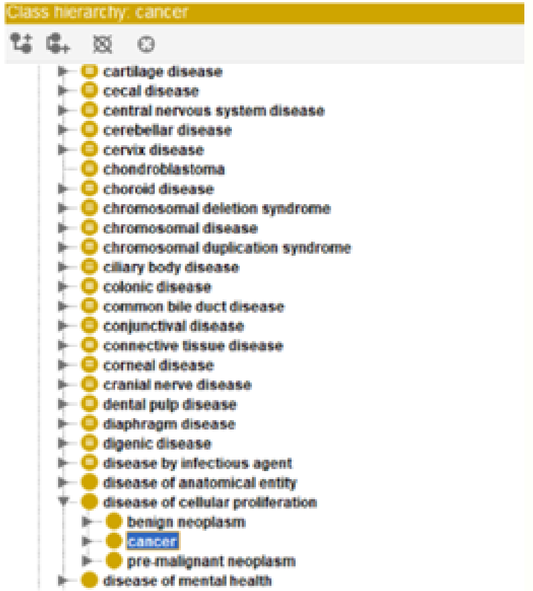
1. Producer Side: → Create interoperable ontologies for datasets:

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





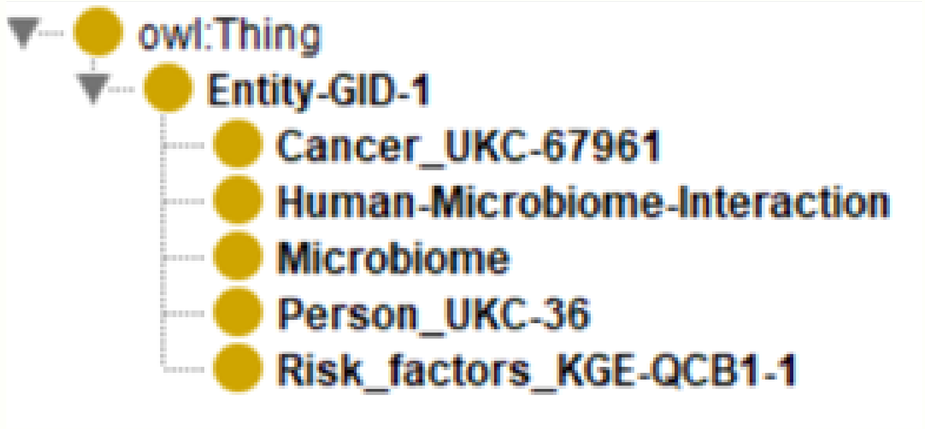




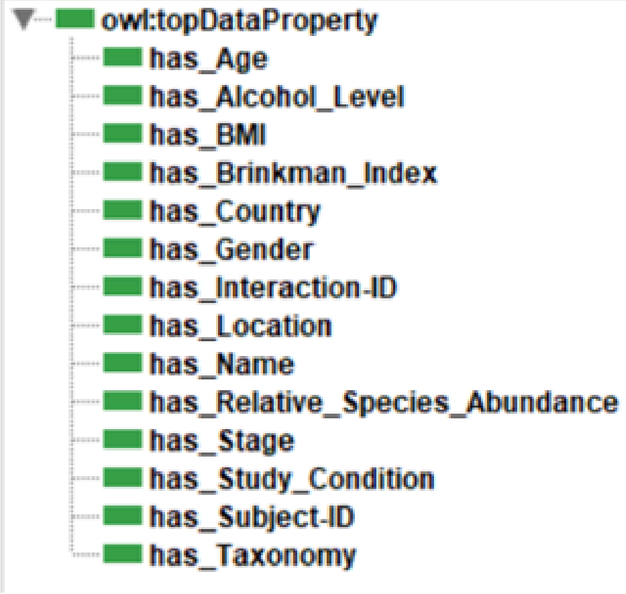
Knowledge Definition

2. Consumer Side

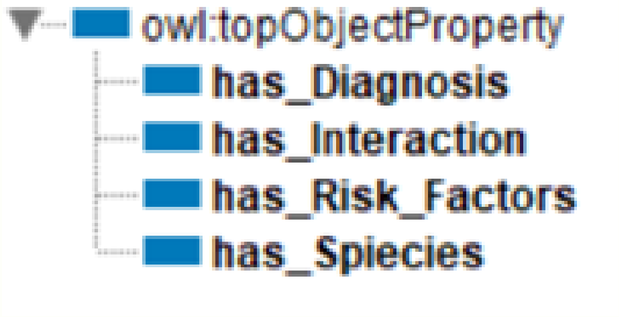
Build a unified **ontology** in Protégé:
Class hierarchies, object & data properties



Classes



Data Properties

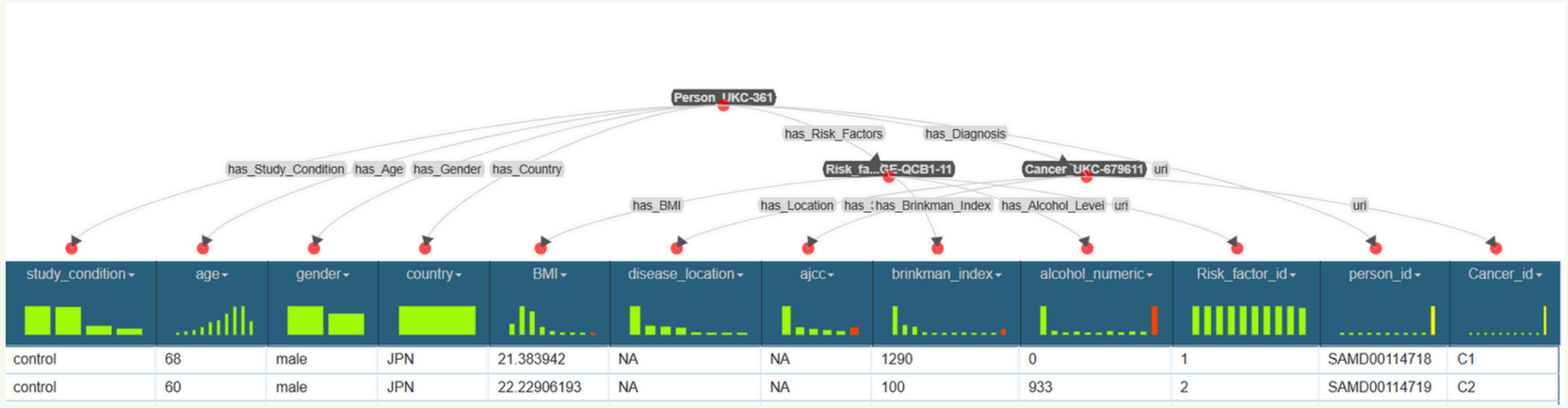


Object Properties

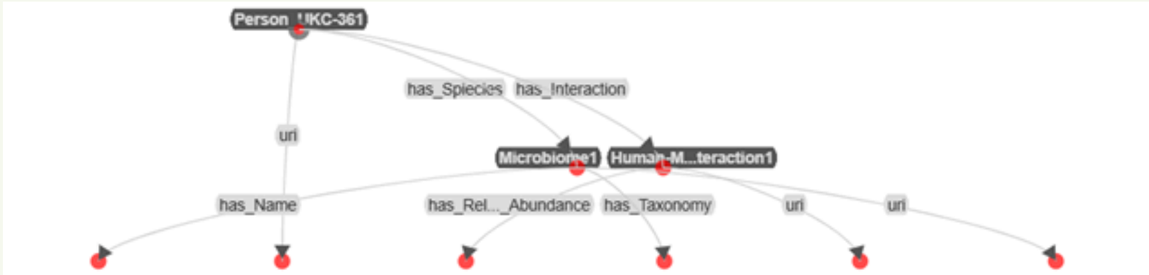
Entity Definition

Objective:

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



study_condition	age	gender	country	BMI	disease_location	ajcc	brinkman_index	alcohol_numeric	Risk_factor_id	person_id	Cancer_id
control	68	male	JPN	21.383942	NA	NA	1290	0	1	SAMD00114718	C1
control	60	male	JPN	22.22906193	NA	NA	100	933	2	SAMD00114719	C2



Species_name	person_id	relative_abundance	taxonomy	interaction_id	Species_ID
Phocaeicola plebeius	SAMD00114718	54.94898	NA	1	Phocaeicola plebeius
Phocaeicola plebeius	SAMD00114719	11.23749	NA	2	Phocaeicola plebeius
Phocaeicola plebeius	SAMD00114720	19.09913	NA	3	Phocaeicola plebeius

Evaluation

Teleontology vs CQs

Teleontology vs Reference Ontologies

Entity level

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

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

Property level

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

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

Exploitation

QUERY 1: CQs about stage 4 cancer and high risk factors

person	alcohol	cig_level	species	occurrence_of_species	mean_of_species	stage
tp://localhost:8080/ ource/SAMD00114750	"638.786"	"820"	http://localhost:8080/ source/ Bacteroides%20uniformis	"12"^^xsd:integer	"11.230103"^^xsd:float	"iv"
tp://localhost:8080/ ource/SAMD00114803	"1899"	"630"	http://localhost:8080/ source/ Bacteroides%20uniformis	"12"^^xsd:integer	"11.230103"^^xsd:float	"iv"
tp://localhost:8080/ ource/SAMD00114810	"348"	"640"	http://localhost:8080/ source/ Bacteroides%20uniformis	"12"^^xsd:integer	"11.230103"^^xsd:float	"iv"
tp://localhost:8080/ ource/SAMD00114817	"360"	"780"	http://localhost:8080/ source/ Bacteroides%20uniformis	"12"^^xsd:integer	"11.230103"^^xsd:float	"iv"
tp://localhost:8080/ ource/SAMD00114750	"638.786"	"820"	http://localhost:8080/ source/ Eubacterium%20rectale	"11"^^xsd:integer	"6.4065185"^^xsd:float	"iv"
tp://localhost:8080/ ource/SAMD00114810	"348"	"640"	http://localhost:8080/ source/ Eubacterium%20rectale	"11"^^xsd:integer	"6.4065185"^^xsd:float	"iv"
tp://localhost:8080/ ource/SAMD00114817	"360"	"780"	http://localhost:8080/ source/ Eubacterium%20rectale	"11"^^xsd:integer	"6.4065185"^^xsd:float	"iv"
tp://localhost:8080/ ource/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
calhost:8080/ SAMD00114899	http://localhost:8080/ source/ Helicobacter%20pylori	"0.00157"	"0"	"22.18934911"	"control"
calhost:8080/ SAMD00164772	http://localhost:8080/ source/ Helicobacter%20pylori	"0.00337"	"570"	"25.40281608"	"adenoma"
calhost:8080/ SAMD00164834	http://localhost:8080/ source/ Helicobacter%20pylori	"0.01394"	"360"	"22.14532872"	"CRC"
calhost:8080/ SAMD00164893	http://localhost:8080/ source/ Helicobacter%20pylori	"0.00399"	"0"	"18.7961895"	"adenoma"

QUERY 2:
CQs about assocoiation of Helicobacter Pylori with cancer

	person	species	relative_value
1	http://localhost:8080/source/ SAMD00164889	http://localhost:8080/source/ Bacteroides%20uniformis	"10.04378"
2	http://localhost:8080/source/ SAMD00114811	http://localhost:8080/source/ Prevotella%20sp%20CAG5226	"10.12263"
3	http://localhost:8080/source/ SAMD00115010	http://localhost:8080/source/ Prevotella%20sp%20CAG520	"10.12896"
4	http://localhost:8080/source/ SAMD00114775	http://localhost:8080/source/ Faecalibacterium%20prausnitzii	"10.13096"
5	http://localhost:8080/source/ SAMD00164867	http://localhost:8080/source/ Bacteroides%20uniformis	"10.13915"

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"^^xsd:float	"2.26875"^^xsd:float	"0.9556599855422974"^^xsd:float	"1.2582162618637085"^^xsd:float	"0.996005892753601"^^xsd:float
http://localhost:8080/source/SAMD00114719	"control"	"100"	"male"	"9.1E-4"^^xsd:float	"0.07518"^^xsd:float	"0.07427000254392624"^^xsd:float	"2.5703961849212646"^^xsd:float	"2.308185815811152"^^xsd:float
http://localhost:8080/source/SAMD00114720	"control"	"1800"	"male"	"0.00525"^^xsd:float	"4.14455"^^xsd:float	"4.1392998695373535"^^xsd:float	"2.566056251525879"^^xsd:float	"2.303845882415775"^^xsd:float
http://localhost:8080/source/SAMD00114721	"control"	"300"	"male"	"16.3262"^^xsd:float	"0.09351"^^xsd:float	"16.232690811157227"^^xsd:float	"13.754894256591797"^^xsd:float	"14.017105102539002"^^xsd:float
http://localhost:8080/source/SAMD00114722	"CRC"	"800"	"male"	"11.08987"^^xsd:float	"8.04042"^^xsd:float	"3.0494508743286133"^^xsd:float	"8.518564224243164"^^xsd:float	"8.780775070190437"^^xsd:float
http://localhost:8080/source/SAMD00114723	"CRC"	"0"	"male"	"0.02001"^^xsd:float	"1.48201"^^xsd:float	"1.4620000123977661"^^xsd:float	"2.5512962341308594"^^xsd:float	"2.289085865020752"^^xsd:float
http://localhost:8080/source/SAMD00114724	"CRC"	"0"	"female"	"4.41078"^^xsd:float	"14.22575"^^xsd:float	"9.814970016479492"^^xsd:float	"1.8394737243652344"^^xsd:float	"2.101684093475342"^^xsd:float

QUERY 3:

CQS about asscoiation 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