# 7 Metadata Definition

In this section, the report collects the definitions of all the metadata defined for the different resources produced throughout the entire process. The metadata defined at this stage describes both the final result of the project and the intermediate results from each phase (language, schema, and standardized values of data sources). Defining metadata is crucial to enabling the distribution (sharing) of the produced resources through data catalogs. Therefore, it is important to describe where these metadata will be published to distribute the resources they describe (e.g., DataScientia catalogs).

In particular, the structure of this section is organized as follows, aiming to describe the metadata related to all types of resources produced by the project:

- **Project Metadata Description**
- **Language Resource Metadata Description**
- **Knowledge Resource Metadata Description**
- **Data Resource Metadata Description**
- **People Metadata Description**

## Project Metadata Description

The project metadata includes essential details about the overall project, such as the title, URLs, and descriptions. This metadata will be published and made available through repositories on GitHub.

| prjURL | prjKeywords | prjType | prjDescription | prjStartD | prjEndD | prjFundingAger | prjInput | prjOutput | prjCoordinator | prjObservations |
|---|---|---|---|---|---|---|---|---|---|---|
| https://github.com/Virginia-hub-del/Knowledge-Graph-Engineering | microbiome,risk factors,cancer,interactions | Knowledge Resource Generation | This Knowledge Graph is designed to highlight correlations between the composition of the microbiota in a collection of individuals, specific risk factors, and colorectal cancer diagnoses. It has been developed using the iTelos methodology. | set-24 | feb-25 | Datascientia foundation | The project utilizes a dataset sourced from the R package CuratedMetagenomicData (CMD). Knowledge resources were derived from BioPortal. | A Knowledge Graph was developed, complete with a full ontology, teleology, and teleontology. Additionally, a GitHub repository, a presentation and a report, including all the details of the full process, were created. | Simone Bocca | |

Figure 22: Project metadata

## Language Metadata Description

Language resource metadata refers to the language, detailing concepts, creators, descriptions, versions, domains, sizes and file formats.

| DatLicense | DatURL | DatKeyword | DatPublisher | DatCreator | DatOwner | DatLanguage | DatSize | DatName | DatPublication | DatDescription | DatVersion | DatDomain | DatFileFormat |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MIT License | https://git | microbioma, cancer, individual, risk factors | Virginia Leombruni | Virginia Leombruni | Virginia Leombruni | english | concepts | Language.tsv | 29/01/2025 | Description of all languages used in the Human-Microbial Interaction Knowledge Graph | 1.0 | Human-Microbial Interaction | tsv |

Figure 23: Language metadata

## Knowledge Metadata Description

Knowledge resources metadata contain detailed descriptions of the knowledge resources used in the project, including their description, creator, language, size ect.

| DatLicense | DatURL | DatKeyword | DatPublisher | DatCreator | DatOwner | DatLanguage | DatName | DatPublication | DatDescription | DatVersion | DatDomain |
|---|---|---|---|---|---|---|---|---|---|---|---|
| http://creativecommons.org/licenses/by/4.0/ | https://www.ebi.ac.uk/ols4/ontologies/ohmi | Biological process, interaction, microbioma-cancer interaction | European Bioinformatics Institute (EMBL-EBI) | Alexander Alekseyenko | European Bioinformatics Institute (EMBL-EBI) | english | Ontology for Host-Microbiome Interactions (OHMI) | 09-ott-23 | Ontology that represents the entities and relations in the domain of host-microbiome interactions | 17/09/2019 | Host-microbiome interactions |
| https://creativecommons.org/publicdomain/zero/1.0/ | https://www.ebi.ac.uk/ols4/ontologies/doid | Disease, cancer | Institute for Genome Sciences, University of Maryland School of Medicine | Northwestern University | Institute for Genome Sciences, University of Maryland School of Medicine | english | Disease Ontology (DOID) | 2003 | Ontology that models the hierarchy of various diseases | 18/12/2024 | Diseases |

Figure 24: Knowledge metadata

## Data Metadata Description

Data resources metadata describes the datasets used or produced during the project. This section includes information about data formats, size, and availability.

| DatLicense | DatURL | DatKeyword | DatPublisher | DatCreator | DatOwner | DatLangua | DatSize | DatName | DatPublicatio | DatDescription | DatVersion | DatDomain | DatFileFormat |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| curatedMetagenomicDa | https://github.com/Virginia-hub-c | person_id, age, gender, BMI, ajcc, brinkman_index, alcohol_numeric | 2021-10-14 YachidaS_201 | YachidaS | YachidaS | english | 55,5 KB | Person_Metadata.csv | 29012025 | It contains all the relevant metadata associated with each sample, providing detailed information about their characteristics | 14102021 | Person metadata | csv |
| curatedMetagenomicDa | https://github.com/Virginia-hub-c | relative_abundance, sample_id | 2021-10-14 YachidaS_201 | YachidaS | YachidaS | english | 1,24 MB | Relative_Abundance.csv | 29012025 | It provides data on the specific quantities of various microbial species found in each sample | 14102021 | Relative abundance | csv |
| ### | https://github.com/Virginia-hub-c | species_name, person_id, relative_abundance, taxonomy, interaction_id | Virginia Leombruni | Virginia Leombruni | Virginia Leombruni | english | 68,3 MB | Transformed_Data_with_Taxonomy. | 29012025 | It contains data on the relative abundance of various microbial species in different individuals, each identified by a unique person_id | 1.0 | Relative abundance | csv |

Figure 25: Dataset Metadata

## People Metadata Description

We have created an additional file, named *People_Metadata*, which compiles essential information about the project's contributors. This document provides a clear reference to the participation and affiliation of each team member.

| comIdentifier | firstName | lastName | email | nationality | gender | affiliation | personalWebpage |
|---|---|---|---|---|---|---|---|
| eleonora-giuliani | Eleonora | Giuliani | eleonora.giuliani@studenti.unitn.it | Italian | F | Datascientia, Knowdive group, Universita degli Studi di Trento | https://github.com/Ele91463 |
| virginia-leombruni | Virginia | Leombruni | virginia.leombruni@studenti.unitn.it | Italian | F | Datascientia, Knowdive group, Universita degli Studi di Trento | https://github.com/Virginia-hub-del |
| marc-shebaby | Marc | Shebaby | marc.shebaby@studenti.unitn.it | Lebanese | M | Datascientia, Knowdive group, Universita degli Studi di Trento | https://github.com/Marc-shebaby |

Figure 26: People metadata