

Refinement of the COHESIVE Information System towards a Unified Ontology of Food Terms for the Public Health Organizations

Iolanda Mangone¹, Nicolas Radomski^{1*}, Adriano Di Pasquale¹, Andrea Santurbano², Paolo Calistri¹, Cesare Cammà¹ and Kitty Maassen³

¹ National Reference Centre (NRC) for Whole Genome Sequencing of microbial pathogens: data-base and bioinformatics analysis (GENPAT), Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "Giuseppe Caporale" (IZSAM), via Campo Boario, Teramo, 64100 (TE), Italy (www.izs.it)

² LARUS Business Automation, via Bruno Maderna 7, Mestre, – 30174 (VE), Italy (www.larus-ba.it)

³ National Institute for Public Health and the Environment (RIVM), P.O. Box 1, Bilthoven, 3720 (BA), The Netherlands (www.rivm.nl)

* corresponding author

Abstract

Background. The task 4.1 of the One Health European joint programme (OHEJP) “One Health Structure In Europe” (COHESIVE) focuses on integrating pathogen information from public health, animal health and food safety surveillance at Member State level. Considered information are metadata associated to each sample (i.e. isolation date, origin, matrix) and whole genome sequencing (WGS) data from official laboratories (e.g. next generation sequencing data and bioinformatics-based analytical outcomes).

Methods. A WEB-based platform called the COHESIVE Information System (CIS) has been created with separate instances for three Member States, in order to provide a proof of concept showing the advantages for surveillance and investigation of outbreaks at the genomic scale, considering food as a source of human pathogens. Currently, a CIS Version 2 (CISv2) is under development to integrate a unified food ontology at Member State level, taking into account as a first step organizations from Italy, Norway and The Netherlands: countries involved in the feasibility study foreseen in the project. More precisely, the last developments focused on the harmonization of the foodborn disease biosample contextual data collected over the past few decades (i.e. contextual metadata of foodborne samples sent in by labs for sequencing) based on the rule-based text mining tool LexMapr, and the implementation of the FoodOn ontology into the CIS based on the graph-database Neo4j to allow future records of harmonized food terms in the CISv2.

Results. The successful harmonization of the past food terms and implementation of the FoodOn ontology into the CIS were mandatory steps allowing food ontology harmonization between organizations and improvement of queries from the CISv2 based on relational- and graph-databases.

Keywords

COHESIVE information system, food ontology, relational-database, graph-database, genomics-based surveillance

IFOW 2021: 2nd Integrated Food Ontology Workshop, held at JOWO 2021: Episode VII The Bolzano Summer of Knowledge, September 11-18, 2021, Bolzano, Italy

EMAIL: i.mangone@izs.it (I. Mangone); n.radomski@izs.it (N. Radomski); a.dipasquale@izs.it (A. Dipasquale); andrea.santurbano@larus-ba.it (A. Santurbano); p.calistri@izs.it (P. Calistri); c.camma@izs.it (C. Cammà); kitty.maassen@rivm.nl (K. Maassen)

ORCID: <https://orcid.org/0000-0002-6716-091X> (I. Mangone); <https://orcid.org/0000-0002-7480-4197> (N. Radomski);

<https://orcid.org/0000-0002-9328-3972> (A. Dipasquale); <https://orcid.org/0000-0001-5066-8971> (A. Santurbano); <https://orcid.org/0000-0001-6004-9373> (P. Calistri); <https://orcid.org/0000-0002-7547-1195> (C. Cammà); <https://orcid.org/0000-0002-0864-464X> (K. Maassen)

© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

1. Introduction

The microbiological surveillance and outbreak investigation are today supported by public health organizations through genomics-based information systems integrating multiple metadata related to foodborne disease biosample collected over the past few decades for sequencing by veterinarians, biologists, researchers and medical doctors [1]. Unfortunately, these metadata associated to samples are, neither organized nor harmonized between public health organizations in charge of food, veterinary and environmental sectors [2]. Consequently, several projects of the food [3], veterinary [4] and environmental [5] sectors, aim currently at organizing and harmonizing these metadata based on developments and implementations of ontologies.

Both in and between European countries, the European Joint Programme (EJP) “One Health Structure In Europe” (COHESIVE) is of paramount importance for organizations of food production systems, as well as the veterinary and human health domains, in dealing with (re-)emerging zoonoses, including antimicrobial resistance and food-borne zoonoses ². Because of current implementation of omics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms across European Member States [6], the EJP COHESIVE, initially developed to collect data related to the area of risk-analysis, aims today at integrating also genomics data from human and veterinary domains involved in genomics-based surveillance (**Figure 1**).

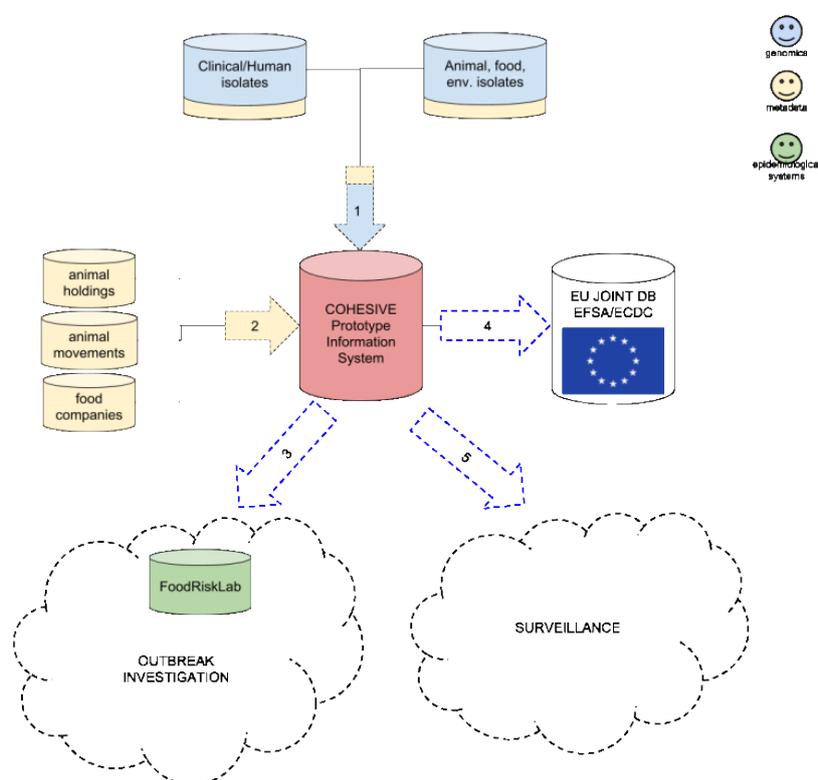


Figure 1: Actors and data types of the “COHESIVE information system” (CIS) supported by the European Joint Programme (EJP) “One Health Structure In Europe” (COHESIVE)

²EJP COHESIVE
Home: <https://onehealthjep.eu/jip-cohesive/>

In the framework of the EJP COHESIVE, the “COHESIVE information system” (CIS) has been developed by IZSAM and three demo versions have been provided to organizations from Italy, Norway and The Netherlands (Task 4.1) to integrate pathogen information from public health, animal health and food safety surveillance at Member State level, integrating metadata related to samples (i.e. isolation date, origin, matrix) and genomics analyses (i.e. genome assembly, mapping of reads, species identification, mutations of interest) [7–9].

The first challenge of the CIS is to harmonize the past food terms which have been accumulated in different languages from different European organizations over the past few decades, though free-systems of recording independently of food term ontology. The second challenge of the CIS is to allow recording by European organizations of future food terms following a common ontology of food terms. A common and unified ontology of food terms into the CIS would allow queries from past and future food terms recorded by different European organizations (e.g. Which samples related to the cheese factory sector were isolated during 2008 in Italy with a clonal complex CC8?).

In parallel, the EJP “One health suRveillance Initiative on harmOnization of data collection and interpretationN” (ORION) (WP3) focused on the development of a “Health Surveillance Ontology” (HSO) at the European level [10], while other international consortia developed “Open Biological and Biomedical Ontology” (OBO) foundries [11], such like the “Genomic Epidemiology Ontology” (GenEpiO) [12] and the “food ontology” (FoodOn) [13]. Among these ongoing projects, the FoodOn ontology fit particularly well requirements of the CIS concerning the need of unified food terms (**Figure 2**), while HSO, OBO, GenEpiO focus mainly on surveillance system level data (e.g. number of samples collected, cases observed, ...), development of interoperable ontologies for the biological sciences (e.g. chemical entities, human disease, gene ontology, phenotype and trait, ...), as well as vocabulary necessary to identify, document and research foodborne pathogens (e.g. genomic laboratory testing, specimen and isolate metadata), respectively.

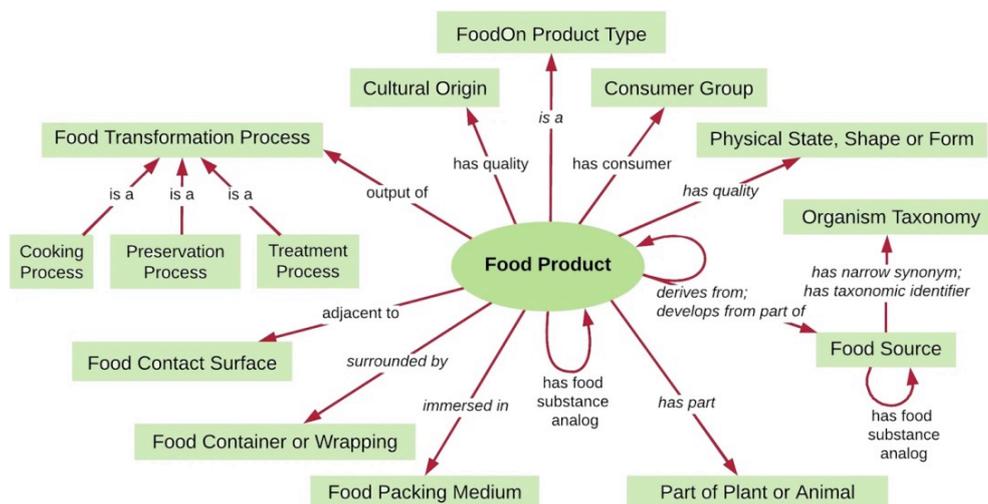


Figure 2: FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration [13]

The specific objectives of the presented CIS Version 2 (CISv2) are (**Figure 3**):

- 1 the harmonization of the past food terms into the CISv2 which we hope to achieve using text mining tool LexMapr and FoodOn ontology from the CIS (i.e. Action 1),
- 2 the recording of the future food terms into the CISv2 which we hope to achieve implementing FoodOn ontology into the CIS through the graph-database Neo4j (i.e. Action 2),
- 3 the discussion of future actions related to food ontology harmonization between organizations which we hope to achieve implementing the CISv2 in different organizations (i.e. Action 3),
- 4 and the discussion of future improvements related to queries from the CISv2 which we hope to achieve combining relational- (CIS) and graph-database (CISv2) relationships (i.e. Action 4).

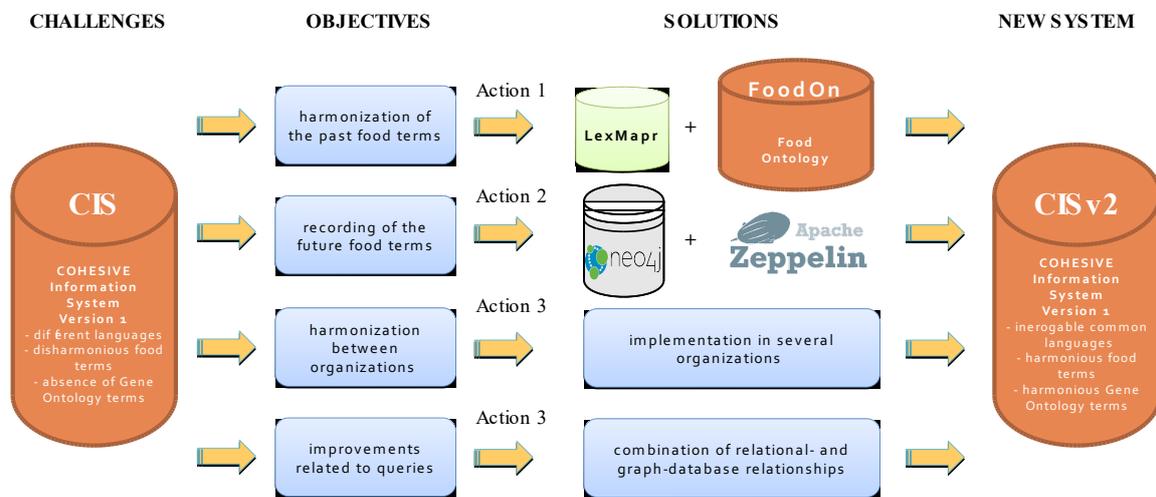


Figure 3: Challenges, objectives, solutions and future new version (CISv2) of the “COHESIVE information system” (CIS) supported by the European Joint Programme (EJP) “One Health Structure In Europe” (COHESIVE)

2. Material and Methods

The harmonization of the past food terms from the CIS (i.e. Action 1) and implementation of the FoodOn ontology into the CIS (i.e. Action 2) are required before performing any actions related to food ontology harmonization between organizations (i.e. Action 3) and improvement of queries from the CISv2 (i.e. Action 4).

2.1. Action 1: Harmonization of Past Food Terms from the CIS based on the Food Ontology FoodOn

The CIS terms come from different providers and were manually curated by each involved organization from Italy, Norway and The Netherlands. The lists of food terms from several organizations written in different languages (i.e. Italy, Norway, Netherlands) were translated by each organization into lists of food terms in English without independent verification of the translation to avoid bias or error (**Figure 4**). This English translation of multilingual food terms was done with Google translate. Then, these lists of English food terms were mapped against the food ontology FoodOn [13] with the rule-based text mining tool LexMapr [14]. More precisely, LexMapr uses a rule-engine for handling synonyms, prefixes and suffixes to automatically map a matrix of English food terms with FoodOn codes. The resulting harmonized food terms were finally imported in the CISv2 (**Figure 4**). The accuracy of LexMapr mapping and missing food terms from FoodOn were not assessed in the present study.

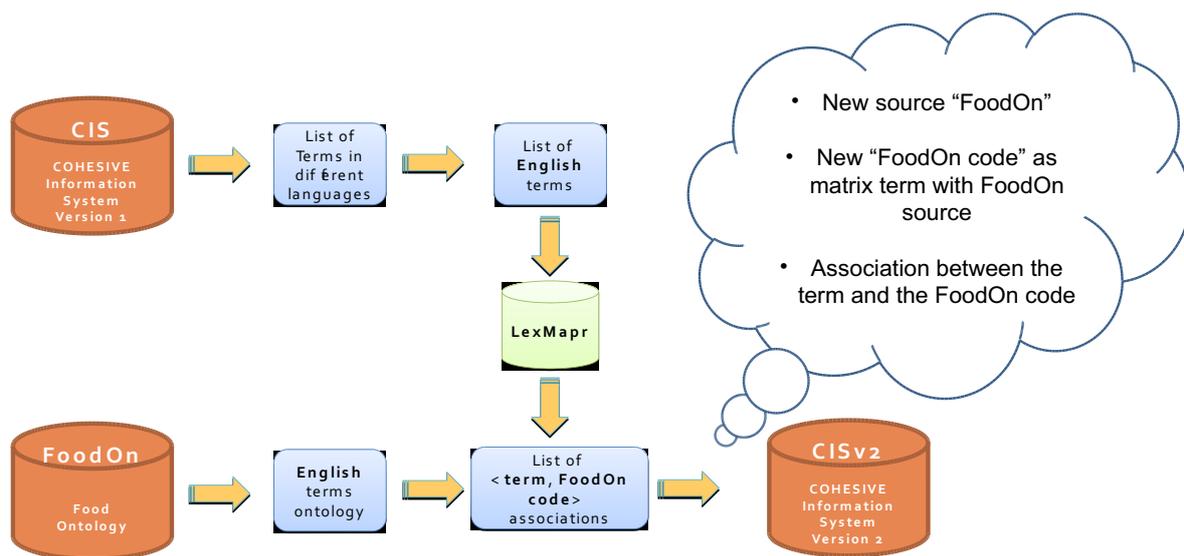


Figure 4: Harmonization of food terms from the CIS based on the food ontology FoodOn with the objective to harmonize the past food terms of samples accumulated over the past few decades into the CISv2

2.2. Action 2: Implementation of the FoodOn Ontology into the CIS to Allow Recording of Future Food Terms from the CISv2

The implementation of the FoodOn ontology into the CIS was performed with the graph-database Neo4j [15] (**Figure 5**). The graph-database Neo4j is able to reveal invisible contexts and hidden relationships, storing and traversing networks of highly connected data. In the present context, Neo4J transforms xml specifications of the FoodOn ontology [13] into a graph-database of food terms. Using open source technologies ³, the FoodOn ontology was imported into Neo4j in order to build the first iteration of the knowledge graph. From the resulting CISv2, the Apache Zeppelin notebook [16] was used to create a data pipeline that covers from the ingestion, the enrichment, to the visualization of queries performed on the FoodOn ontology through Neo4J [15] (**Figure 5**). The Apache Zeppelin notebook is a Web-based notebook that enables data-driven, interactive data analytics and collaborative documents with SQL and other languages.

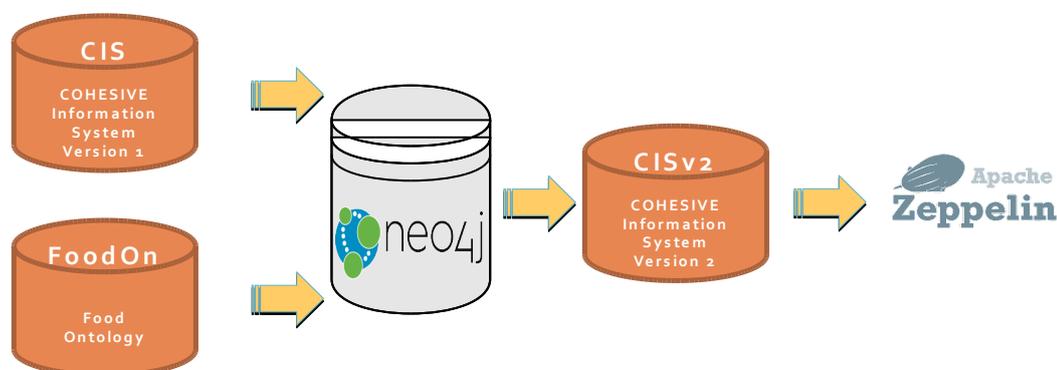


Figure 5: Implementation of FoodOn ontology into the CIS with the objective to allow future records of food terms into the CISv2 based on the food ontology FoodOn

³Open source technologies

GitHub: <https://github.com/neo4j-contrib/neo4j-apoc-procedures> and <https://github.com/neo4j-labs/neo4j-semantic>

3. Results

The tools for the harmonization of the past food terms into the CIS (i.e. Action 1) and implementation of the FoodOn ontology into the CIS (i.e. Action 2) were carefully selected according to future other actions related to food ontology harmonization between organizations (i.e. Action 3) and improvement of queries from the CISv2 (i.e. Action 4).

3.1. Action 1: Harmonization of Past Food Terms from the CIS based on the Food Ontology FoodOn

Compared to nutritional ontologies designed to annotate and describe intervention trials [17–19], the FoodOn ontology [13] was selected because of its strong representation of food nutrients and processing [20]. Compared to other text mining tools in Literature [21], finance [22] and medicine [23], the selection of the rule-based text mining tool LexMapr to map food terms translated in English against the FoodOn ontology (Table 1) was driven by its interoperability across sectors [14]. Initially developed to fulfil objectives of public health surveillance networks like the US FDA’s GenomeTrakr system and the US National Antimicrobial Resistance Monitoring System (NARMS), LexMapr describes indeed food pathogen source for reporting of transmission dynamics in public health foodborne pathogen surveillance and investigation 4. Without speaking about issues related to English translations, the proposed mockup (Table 1) shows that there is a need to improve FoodOn's curation because FoodOn does not have a term for a generic pizza with or without meat or cheese 5.

Table 1

Mockup of food terms from the CIS harmonized through LexMapr-based mapping against the FoodOn ontology with the objective to harmonize the past food terms of samples accumulated over the past few decades into the CISv2

Code from providers	Matrix	FoodOn Code*	Source [#]
Code-1	Pizz ka a muzzarel e a pummarol ngopp	FOODON_000102	Naples coding system
Code-2.1	Pizza Margherita	FOODON_000102	ISO
Code-2	Pizza	FOODON_000001	ISO

* FoodOn codes are arbitrary examples [#] ISO means that samples from the present example follow the requirements of the International Organization for Standardization

⁴Open source text mining tools

LexMapr: <https://www.cineca-project.eu/blog-all/lexmapr-a-rule-based-text-mining-tool-for-ontology-term-mapping-and-classification>

⁵FoodOn term "pizza food product"

URL: https://urlsand.esvalabs.com/?u=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FFOODON_03310775&e=2f9e67d3&h=88f8ff14&f=n&p=y

3.2. Action 2: Implementation of the FoodOn Ontology into the CIS to Allow Recording of Future Food Terms from the CISv2

The implementation of the FoodOn ontology into the CIS was successfully performed through Neo4J [15], as exemplified with the Apache Zeppelin notebook (Figure 6) [16]. The graph-database Neo4j was selected for its capacity to query easily ontologies [15].

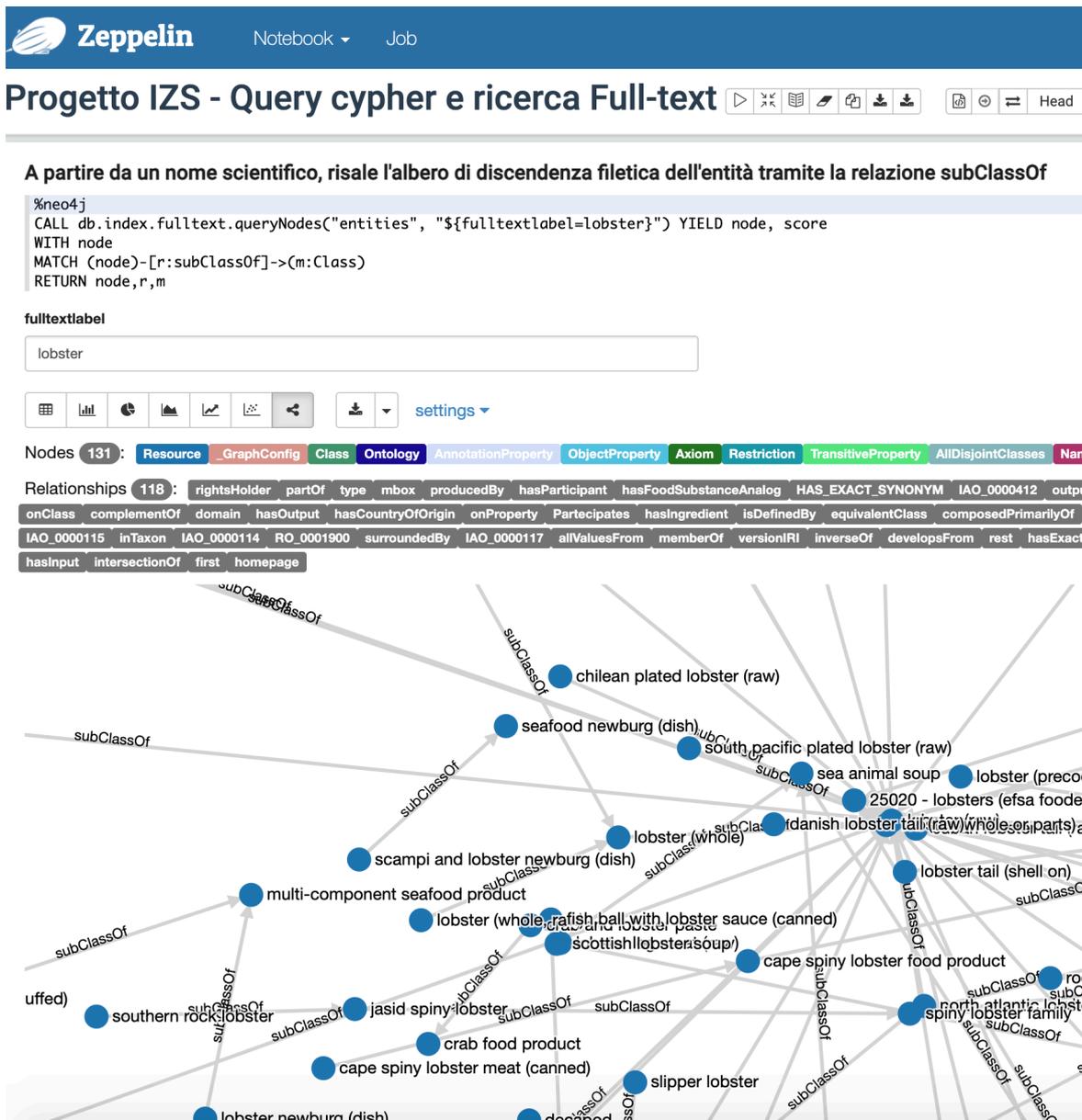


Figure 6: Zeppelin browser allowing to navigate through and query the FoodOn ontology implemented into the CIS through Neo4j with the objective to allow future records of food terms into the CISv2 based on the FoodOn ontology

4. Discussion

4.1. Action 3: Harmonization of Food Terms and Implementation of the FoodOn Ontology in Public Health Organizations

In a near future, we plan to harmonize the past food terms from the CIS (i.e. Action 1) and implement the FoodOn ontology into the CIS (i.e. Action 2) instances provided during the feasibility study for the organizations: IZSAM (Italy), NVI (Norway) and RIVM (The Netherlands) at the time, in order to use the same CISv2 implementing an identical food ontology. Indeed, different coding systems can be harmonized looking for different English translated terms that LexMapr [14] maps towards the same FoodOn codes [13] (Table 1). Even if LexMapr mapping of English translated terms against the FoodOn ontology may not be completely efficient, it allows sharing of common FoodOn terms derived from English translated terms between organizations using different languages.

4.2. Improvement of Queries from the CISv2

Today, the CIS without Neo4j implementation (**Figure 4**) can be interrogated through relational-database queries for data related to isolation date, sampling origin, food matrix and/or genomics data, while the CISv2 implementing Neo4j [15] can be interrogated through queries of graph-database relationships for food sectors organized inside the FoodOn ontology [13]. We plan in a near future to improve queries combining queries of relational- and graph-database relationships in order to interrogate the CISv2 for isolation date, sampling origin, food matrix, genomics data and/or food sectors organized inside the FoodOn ontology. Just using the CIS, we can query samples via standard SQL statements ().

```
----- SQL QUERY on sampling and genomics metadata -----  
select samples where  
  sampling_date is 2018 and  
  sampling_place is Italy and  
  CC=CC8
```

Draft code 1: SQL query using relational-database relationships of the CIS (response: samples isolated during 2008 in Italy with a clonal complex CC8)

Neo4J implemented into the CISv2 allows query of FoodOn terms using graph-database relationships (). If the queried FoodOn term do not exist, it would become a new Environment Ontology (ENVO) term useful for FoodOn curation.

```
----- Neo4J QUERY on FoodOn Terms -----  
select foodoncode  
  with a relationship to foodoncode of "cheese factory"
```

Draft code 2: Neo4J query using graph-database relationships of the CISv2 (response: FoodOn codes related to the FoodOn term "cheese factory")

Using the combination of SQL and Neo4J queries, an innovative query could use relational- and graph-database relationships ().

```
(
  ----- SQL QUERY on sampling and genomics metadata -----
  select samples where
    sampling_date is 2018 and
    sampling_place is Italy and
    CC=CC8
) AND foodoncode in (
  ----- NEO4J QUERY on FoodOn Terms -----
  select foodoncode
    with a relationship to foodoncode of "cheese factory"
)
```

Draft code 3: SQL and Neo4J queries using relational- and graph-database relationships of the CISv2, respectively (response: samples isolated during 2008 in Italy with a clonal complex CC8 and a FoodOn code related to the FoodOn term "cheese factory")

Adding additional ontologies, such as the Gene Ontology (GO) dedicated to GO terms [24, 25] describing the metabolic pathways 6, other innovative combinations of SQL and Neo4J queries could use relational- and graph-database relationships filtering samples via classical relational constraints, and adding ontology constraints on food matrix and genome annotations (). This kind of Gene Ontology-based query is typically useful to list GO-terms from a subset of genomes and a larger collection of genomes in order to perform a Genome Ontology Enrichment Analysis (GOEA) [26] identifying over-represented metabolic pathways across genomes of interest (e.g. genomes involved in an outbreak).

```
(
  ----- SQL QUERY on sampling and genomics metadata -----
  select samples where
    sampling_date is 2018 and
    sampling_place is Italy and
    CC=CC8

) AND foodoncode in (
  ----- NEO4J QUERY on FoodOn Terms -----
  select foodoncode
    with a relationship to ontology_node of "cheese factory"

) AND GOterm in (
  ----- NEO4J QUERY on GO Terms -----
  select GOterm
    with a relationship to ontology_node of "ATPase activity"
)
```

Draft code 4: SQL and Neo4J queries using relational- and graph-database relationships of the CISv2, respectively (response: samples isolated during 2008 in Italy with a clonal complex CC8, a FoodOn code related to the FoodOn term "cheese factory" and a Gene Ontology code related to the Gene Ontology term "ATPase activity")

⁶Open source ontology
GENE ONTOLOGY: <http://geneontology.org/docs/download-ontology/>

4.3. Significant Overlapping with Existing Efforts

The present development of the CISv2, dedicated to food ontology between organizations, overlaps significantly to existing efforts in the field of human nutrition, especially the food, health, nutrition domain ontologies (FHNDO) [17], the Ontology for Nutritional Studies (ONS) [18], and the Ontology for Nutritional Epidemiology (ONE) [19]. While these nutritional ontologies aim at identifying healthy diets based on interoperability of ontologies related to classifications of diets, diseases and food [17–19], the CISv2 is dedicated to surveillance and investigation of foodborne outbreaks in human at the genomic scale based on ontologies related to genes and food. Instead of identifying healthy diets (i.e. FHNDO, ONS and ONE), the immediate goal of the project is to explain genetically foodborne outbreaks in human (i.e. CISv2). Compared to the relational-databases dedicated to the nutritional ontologies FHNDO [17], ONS [18] and ONE [19], the CISv2 presents the advantage to be able to combine relational- (i.e. CIS) and graph-databases (i.e. Neo4j).

5. Conclusion

The harmonization of the past food terms into the CIS (i.e. Action 1) and implementation of the FoodOn ontology into the CIS (i.e. Action 2) will allow harmonization of the food ontology between organizations (i.e. Action 3) and improvement the interrogation of the CISv2 (i.e. Action 4) combining queries from relational- (i.e. CIS) and graph-databases (i.e. Neo4j). Thenceforth, the CISv2 need FoodOn curators from the ontology community to perform better biosample description, text mining and text mashing to ontology terms. In the longer term, we also plan to extend the CISv2 to other ontologies, like Gene Ontology in order to perform GOEA. Based on successful outcomes of actions 1 and 2, the CISv2 presents today harmonious English food terms and can be distributed for easy implementation in different European organizations with standard servers (Action 3) and used to perform combinations of queries from relational- (i.e. past food terms from CIS) and graph-databases (i.e. future food terms from CISv2).

6. Acknowledgements

The study was funded by the European Joint Programme (EJP) “One Health Structure In Europe” (COHESIVE). Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the IZSAM. The authors declare that they have no competing interests and thank especially the Italian Ministry of Health for supporting in the acquisition of high-performance computing resources. This manuscript was drafted by Nicolas Radomski based on an available Word template under a Creative Commons License Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).

References

- [1] Y. Jang, T. Choi, J. Kim, J. Park, J. Seo, S. Kim, et al. An integrated clinical and genomic information system for cancer precision medicine. *BMC Med Genomics*. 2018;11:34.
- [2] E. Griffiths, D. Dooley, M. Graham, G. Van Domselaar, F.S.L. Brinkman, W.W.L. Hsiao. Context Is Everything: Harmonization of Critical Food Microbiology Descriptors and Metadata for Improved Food Safety and Surveillance. *Front Microbiol*. 2017;8:1068.
- [3] T. Eftimov, G. Ispirova, D. Potočnik, et al. ISO-FOOD ontology: A formal representation of the knowledge within the domain of isotopes for food science. *Food Chem*. 2019;277:382–90.
- [4] F.C. Dórea, F. Vial, K. Hammar, A. Lindberg, P. Lambrix, et al. Drivers for the development of an Animal Health Surveillance Ontology (AHSO). *Prev Vet Med*. 2019;166:39–48.
- [5] M. Masmoudi, S.B. Abdallah Ben Lamine, H.B. Zghal, et al. An ontology-based monitoring system for multi-source environmental observations. *Procedia Comput Sci*. 2018;126:1865–74.
- [6] EFSA Panel on Biological Hazards (EFSA BIOHAZ Panel), K. Koutsoumanis, A. Allende, A. Alvarez-Ordóñez, D. Bolton, S. Bover-Cid, et al. Whole genome sequencing and metagenomics

- for outbreak investigation, source attribution and risk assessment of food-borne microorganisms. *EFSA J.* 2019;17. doi:10.2903/j.efsa.2019.5898.
- [7] S. Srinivasan, S.K. Mandal, L. Kumar, C. Menaka, A.A. Menon. A Structured Protective Cohesive Health Care Information System Using Security And Storage Mechanism In Cloud. *Int J Eng Trends Technol.* 2021;69:29–33.
- [8] A. Di Pasquale and V. Caldarelli. COHESIVE Information System (CIS). 2019. doi:10.5281/ZENODO.3257358.
- [9] A. Di Pasquale and V. Caldarelli. COHESIVE Information System (CIS). 2019. doi:10.5281/ZENODO.3342572.
- [10] M. Filter, T. Buschhardt, F. Dórea, E. Lopez de Abechuco, T. Günther, E.M. Sundermann, et al. One Health Surveillance Codex: promoting the adoption of One Health solutions within and across European countries. *One Health Amst Neth.* 2021;12:100233.
- [11] The OBI Consortium, B. Smith, M. Ashburner, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol.* 2007;25:1251–5.
- [12] FAIRsharing Team. FAIRsharing record for: GenEpiO. 2018. doi:10.25504/FAIRSHARING.Y1MMBV.
- [13] D.M. Dooley, E.J. Griffiths, G.S. Gosal, P.L. Buttigieg, R. Hoehndorf, M.C. Lange, et al. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *Npj Sci Food.* 2018;2:23.
- [14] G. Gosal, E. Griffiths, D. Dooley, I. Gill, D. Fornika, H. Tate, et al. LexMapr: a rule-based text mining tool for ontology-driven harmonization of short biomedical specimen descriptions. 2019. doi:10.7490/F1000RESEARCH.1117323.1.
- [15] F.M. Santos López, E.G. Santos De La Cruz. Literature review about Neo4j graph database as a feasible alternative for replacing RDBMS. *Ind Data.* 2015;18:135.
- [16] D. Chrimes, H. Zamani. Using Distributed Data over HBase in Big Data Analytics Platform for Clinical Services. *Comput Math Methods Med.* 2017;2017:1–16.
- [17] T. Helmy, A. Al-Nazer, S. Al-Bukhitan, A. Iqbal. Health, Food and User’s Profile Ontologies for Personalized Information Retrieval. *Procedia Comput Sci.* 2015;52:1071–6.
- [18] The ENPADASI consortium, F. Vitali, R. Lombardo, D. Rivero, F. Mattivi, P. Franceschi, et al. ONS: an ontology for a standardized description of interventions and observational studies in nutrition. *Genes Nutr.* 2018;13:12.
- [19] C. Yang, H. Ambayo, B. De Baets, P. Kolsteren, N. Thanintorn, D. Hawwash, et al. An Ontology to Standardize Research Output of Nutritional Epidemiology: From Paper-Based Standards to Linked Content. *Nutrients.* 2019;11:1300.
- [20] L. Chan, N. Vasilevsky, A. Thessen, et al. The landscape of nutri-informatics: a review of current resources and challenges for integrative nutrition research. *Database.* 2021;2021:baab003.
- [21] L. Feng, Y.K. Chiam, S.K. Lo. Text-Mining Techniques and Tools for Systematic Literature Reviews: A Systematic Literature Review. In: 2017 24th Asia-Pacific Software Engineering Conference (APSEC). Nanjing: IEEE; 2017. p. 41–50. doi:10.1109/APSEC.2017.10.
- [22] A. Gupta, V. Dengre, H.A. Kheruwala, M. Shah. Comprehensive review of text-mining applications in finance. *Financ Innov.* 2020;6:39.
- [23] C. Luque, J.M. Luna, M. Luque, S. Ventura. An advanced review on text mining in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov.* 2019;9. doi:10.1002/widm.1302.
- [24] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, et al. Gene Ontology: tool for the unification of biology. *Nat Genet.* 2000;25:25–9.
- [25] E.I. Boyle, S. Weng, J. Gollub, H. Jin, D. Botstein, J.M. Cherry, et al. GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics.* 2004;20:3710–5.
- [26] M. Vila Nova, K. Durimel, K. La, A. Felten, P. Bessières, M.Y. Mistou, et al. Genetic and metabolic signatures of *Salmonella enterica* subsp. *enterica* associated with animal sources at the pangenomic scale. *BMC Genomics.* 2019;20:814.