

# FoodOn: A Semantic Ontology Approach for mapping Foodborne Disease Metadata

Dalia A. Alghamdi, Damion M. Dooley, Gurinder Gosal, Emma J. Griffiths, Fiona S.L. Brinkman and William W.L. Hsiao

<sup>1</sup> BC Center for Disease Control, 655 W 12th Ave, Vancouver, BC V5Z 4R4, Canada

<sup>2</sup> University of British Columbia, 2329 West Mall, Vancouver, BC V6T 1Z4, Canada

<sup>3</sup> Simon Fraser university, 8888 University, Burnaby, BC V5A 1S6, Canada

## ABSTRACT

The FoodOn Food Ontology contains standardized terms and a facet-based classification scheme for describing food products, processing and environments. Mapping of foodborne pathogen isolate source information (descriptors of the contaminated materials and locations) to the FoodOn standard can facilitate data sharing and integration between multi-jurisdictional health and regulatory agencies utilizing disparate software platforms and data dictionaries. Faster and more efficient sharing of information is critical for tracking and controlling outbreaks of foodborne disease at local, national and international levels. This work describes mapping procedures which can be utilized by organizations and software developers to better enable interoperability between foodborne pathogen surveillance and outbreak management systems.

## INTRODUCTION

Globalization of food networks increases opportunities for the spread of foodborne pathogens beyond borders and jurisdictions, with major impacts on global health and economies (Altekruse & Swerdlow *et al.*, 1996; World Health Organization, 2008). Whole genome sequencing (WGS) provides the highest resolution evidence for identifying, typing and matching foodborne pathogen isolates from different sources. WGS results must be combined with source information to be meaningfully interpreted for regulatory and health interventions, outbreak investigation, and risk assessment. Isolate metadata (source of a pathogen) is critical for determining mode of disease transmission, sources of exposure and risk, susceptible populations, geographical distribution and more. Public health and regulatory agencies not only use different analytical platforms to track and resolve outbreaks, but implement different data dictionaries and free text descriptions for describing isolates and exposures. The most important factor in reducing the number of preventable cases of disease is timeliness of investigations and responses, which is negatively impacted by the time-consuming re-coding and manual curation required for translating non-standardized information between systems and agencies. To address the interoperability problem, it is important to relate similar concepts or relations from one agency, information management system or jurisdiction, to another. Mapping terms using an ontology represents a very powerful solution for standardizing and integrating heterogeneous data.

FoodOn (<http://foodon.org>) is an ontology resource that aims to model the food domain, which includes knowledge about food and food-related human activities, such as agriculture, medicine, food safety inspection, shopping patterns and sustainable development (Griffiths *et al.*, 2016). Mapping of foodborne pathogen isolate metadata to FoodOn can provide a means for standardizing, translating, and communicating this critical contextual information between health agencies and platforms in a timely fashion. Here we describe a semi-automated method derived from mapping metadata from the widely used online microbial MLST typing platform Enterobase, which can be broadly applied to other use cases.

## FOODON DESIGN PRINCIPLES

Although there are several existing indexing systems directly or indirectly related to food and food-borne illness, including those maintained by Health Canada, the US Department of Agriculture, and the UN's Food and Agriculture Organization, they have been built for different purposes and so differences in their architecture hinder interoperability. To provide a more comprehensive view of food safety, data from these various sources must be integrated. In a concerted effort to solve this semantic interoperability problem, the OBOFoundry.org family of ontologies was established in 2007 in order to provide a comprehensive set of vocabularies in the biomedical domain. FoodOn, built largely on a longstanding American and European facet-based food indexing system called LanguaL (<http://langual.org>), provides a list of over 2,000 plant and animal food ingredient terms, as well as a supplemental list of over 9,000 indexed food products. Facets include fields for describing food processing, cooking and preservation, as well as source ingredient anatomy, taxonomy, geography and cultural heritage. The aim of FoodOn is to develop an international standard for describing properties of food related to agriculture, animal husbandry, collection, distribution, preservation, culinary use, consumption and food safety. FoodOn was accepted into the OBOFoundry in 2017.

## FOODON MAPPING AND DATA HARMONIZATION

Microbial Multilocus Sequence Typing (MLST) is a technique used to classify and identify pathogenic strains for outbreak investigation and surveillance of contamination.

\* To whom correspondence should be addressed: Dalia.alghamdi@bccdc.ca

Enterobase is a widely used online platform enabling MLST analysis of enteric pathogens such as *E. coli*, *Salmonella*, *Shigella*, *Yersinia* and *Moraxella*. Enterobase contains >100,000 genomes, along with their source metadata, encompassing food, anatomical and clinical, as well as environmental domains.

Based on curation and mapping of Enterobase isolate metadata, we have developed a semi-automated ontology mapping system that will enable mapping of food safety metadata to FoodOn food products and processing environments according to the following steps: (1) Syntactic analysis, where each categorical term in a single free text entry will be separated. (2) Semantic mapping of each categorical term according one of the following rules: (a) mapping to similar concept; (b) mapping to similar ancestors; (c) mapping to similar relations; (d) combining several matching techniques. (3) structural mapping, where the items are mapped to a corresponding subclass in the reference hierarchy.

Non-interactive ontology matching tool can be evaluated using recall and precision (J. Euzenat and P. Shvaiko., 2010). They are measured based on comparing the expected results with the results of the evaluated system. Precision measure the ration of the correct matched terms over the total number of the matched terms, on the other hand, recall measures the ratio of the correct matched of the total number of expected terms to be matched. Logically, one can say that the precision can evaluate the correctness of the evaluated system and recall can evaluate the completion of the evaluated system (Euzenat., 2007). Here, we will evaluate our mapping accuracy by sub-sampling randomly a set of 500 records from the Enterobase and evaluate the suitability of the terms assigned using recall and precision methods. We will further manually review all the terms that cannot be manually mapped to an ontological term and add the terms to appropriate ontologies. The goal of our exercise is to minimize manually intervention when annotating food-related sample sources using FoodOn.

Integrating genomic profiling of foodborne pathogens with a food descriptor framework will help reduce barriers for knowledge exchange among research communities, government risk managers and health providers.

## ACKNOWLEDGEMENTS

We thank Mark Achtman, Nabil-Fareed Alikhan, Mark Pallen, Martin Sergeant, Zheming Zhou for sharing the Enterobase.

## REFERENCES

Altekruse S, Swerdlow D. The changing epidemiology of foodborne disease. *Am J Med Sci.* 1996, 311: 23-29.

World Health Organization, Foodborne disease outbreaks: guidelines for investigation and control, WHO Press, Geneva (2008).

Griffiths, E., Dooley, D., Buttigieg, P.L., Hoehndorf, R., Brinkman, F., and Hsiao, W. (2016). FoodOn: A global farm-to-fork food ontology. ICBO Conference. Corvallis, OR, USA

J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer, 2007.

Jérôme Euzenat. 2007. Semantic precision and recall for ontology alignment evaluation. In *Proceedings of the 20th international joint conference on Artificial intelligence (IJCAI'07)*, Rajeev Sangal, Harish Mehta, and R. K. Bagga (Eds.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 348-353.