Constructing a knowledge graph for food health claims

Remzi Celebi[1], Ilse van Lier[2], Alie de Boer[2] and Michel Dumontier[1]

[1]*Institute of Data Science, Faculty of Science and Engineering, Maastricht University, the Netherlands*
[2]*c*

Abstract

Authorised health claims and their scientific opinions offer valuable insights into health effects of foods and food ingredients. However, these texts are highly technical and not easily accessible for those interested in the development and use of healthy food products and diets. In this paper, we present the development of a knowledge graph that makes the information of 260 authorised health claims more accessible. The knowledge graph, based on data from scientific opinions, is subdivided into four categories: the food (ingredient); the health effect; the target group; and the scientific evidence underlying the cause-and-effect relationship. Various differences were found between authorised claims and their underlying scientific opinions. These findings underline the need for further structuring the approach to substantiating and assessing health claims. Most importantly however, the development of this knowledge graph allows consumers, food producers and health care professionals to make personalised decisions in selecting healthy nutrition.

Keywords

personalised nutrition, food health claim, knowledge graphs, FAIR data

1. Introduction

Structuring and formalising data about food is important not only for humans (i.e., individuals, farmers, food producers and policy makers), but also for computers to make sense of data and make it re-usable for other purposes such as personalized nutrition. One particular important knowledge resource about food are the claims that currently can be made about health effects of foods or food products. These claims are in the European Union known as health claims. Only claims that have been scientifically substantiated are allowed to be placed on products that are produced and/or sold within the EU, to ensure the highest level of consumer protection. For transparency reasons, all opinions of the European Food Safety Authority (EFSA) (both negative and positive) about these claims are made available online. In its current database, EFSA has listed information on 2,338 health-related claims that deal with 1,202 unique food items. These claims are either function claims (Article 13(1) or 13(5) claims), disease risk reduction claims (Article 14(1)(a) claims) or claims on children's growth and development (Article

IFOW'21: Integrated Food Ontology Workshop (IFOW), September 15-18, 2021, Bolzano, Italy
remzi.celebi@maastrichtuniversity.nl (R. Celebi); i.vanlier@maastrichtuniversity.nl (I. v. Lier); a.deboer@maastrichtuniversity.nl (A. d. Boer); michel.dumontier@maastrichtuniversity.nl (M. Dumontier)
0000-0001-7769-4272 (R. Celebi); 0000-0001-8381-1252 (I. v. Lier); 0000-0002-6500-4649 (A. d. Boer); 0000-0003-4727-9435 (M. Dumontier)

CEUR Workshop Proceedings (CEUR-WS.org)

14(1)(b) claims). The putative claims in the database deal with various health relationships such as "reduction of post-prandial glycaemic responses" or "maintenance of normal blood LDL-cholesterol concentrations." The database includes both authorised and non-authorised claims, for which a review has been completed. However, the scientific opinion for a health claim made by EFSA is published in free text format, which requires considerable expertise and effort for someone to understand and interpret the content of the reviews.

In our research, we have therefore aimed to develop a knowledge graph based on a structured vocabulary for the terminology used in scientific opinions, to extract relevant information from the scientific opinions on health claims. Knowledge graphs are a type of knowledge structure where data is represented in the form of interlinked facts where each fact is represented as a directed labelled link between two concepts. Knowledge graphs not only provide a model data to define and formalise the concepts and relations in a domain, but also use a (machine and human readable) standard vocabulary and format, and unique identifiers for concepts and relationships, making data more findable, accessible, interoperable and reusable (FAIR) [1, 2]. In this paper, we describe the development of the data model and the knowledge graph for food health claims. The knowledge graph is accessible through the Web APIs [1], which were generated from some query templates. We show that we can query the knowledge graph on health claims related to the four sub-dimensions of claims: the food ingredient, the type of health benefit, the target population and the scientific evidence substantiating the causal relationship between (food) ingredient and health effect.

2. Method

For the knowledge graph, we have developed a data model to reflect all available information from EFSA's scientific opinions on health claims, including the specific nutrients or bioactive ingredients, the health relationships and biomarkers, the conditions of use for such a claim (e.g. the population that is referred to in the claim) and the supportive evidence underlying these claims.

The information used to develop this knowledge graph originates from EU authorised health claims and their underlying scientific opinion, in which the details are provided regarding the active substance (food or food ingredient), the beneficial effect as well as the relevant evidence to support the causal relationship between ingredient and effect. We have extracted data related to the food (ingredient) and the suggested consumed amount, the physiological effect and the target group for this effect, as well as the submitted scientific evidence from all scientific opinions that were in the Register described to support the authorised claim. This resulted in the inclusion of 260 claims and their scientific opinions: the scientific substantiation and conditions of use of 235 authorised function claims (229 based on generally accepted scientific evidence and six based on newly developed evidence), 13 disease risk reduction claims and 12 claims on children's development and health.

[1]http://grlc.io/api-git/MaastrichtU-IDS/food-claims-kg

2.1. Dimensions of health claims

By developing a data model and the related knowledge graphs about food health claims, we have gained insight into the scientific basis for all authorised claims. For example, we show how to find information on which authorised health claims they can use on folic acid containing foods, but also what sources would support the causal relationship between calcium and normal blood clotting, one specific authorised claim. And whereas the first aim was to make this data FAIR to support the use of authorised claims (and ingredients bearing authorised claims), this scientific basis showed to be a highly interesting result in itself. We divided our data model into four sub-structures in order to better explain different dimensions of health claims, as shown in Figure 1. These relate to the three steps taken in assessing scientific dossiers for health claim authorisations [3, 4] the food or active ingredient itself (here labelled as Food), the beneficial physiological effect (Health Effect), the potential target group for this beneficial physiological effect (Target Population) and finally, the scientific evidence that is supports the causal relationship between consuming the ingredient and the suggested beneficial physiological effect (Scientific Evidence)

The proposed data model for food health claim has four dimensions: Scientific Evidence, Food, Health Effect and Target Population. Each dimension is represented by an existing/proposed vocabulary. To define the scientific evidence on food health claims, we used a micropublication data model (shown as prefix mp:) which captures relationships between supporting statements (mp:Statement), literature references (mp:Reference) and claims (mp:Claim). To characterize foods, we created our own class definition for the type and source of a food (fhck:Food and fhck:FoodSource). We also used the schema.org vocabulary to describe the recommended intake for the food (sc:DoseSchedule). We used main phenotype vocabularies including GO, MEDDRA and HPO to describe the health effect mentioned (fhck:Phenotype) in a claim and classified the health effect into 4 types: (i) maintenance of a function, (ii) enhancing a function, (iii) supporting 'normal' homeostasis, and (iv) reducing a risk factor in disease development. Finally, we used the existing classes and properties from the PICO ontology to define the target population for a claim.

2.2. Food

To characterise a food or food ingredient, the category of the food must first be determined. The categories for foods are defined as follows: Food, Nutrient and Substance. Next, it is necessary to describe the food item using a common vocabulary/ontology to normalise and disambiguate a food item from others. We choose FooDB vocabulary as a common vocabulary because of its rich and extensive content, providing the cross references to other databases. Although some claims only specify a specific food or nutrient, these must be consumed under certain conditions (e.g., the source of the food and the minimum amount that should be consumed) in order to show its effect. If the food should be taken from a specific source, this source also needs to be defined. Finally, the minimum quantity and unit recommended to be consumed need to be defined.

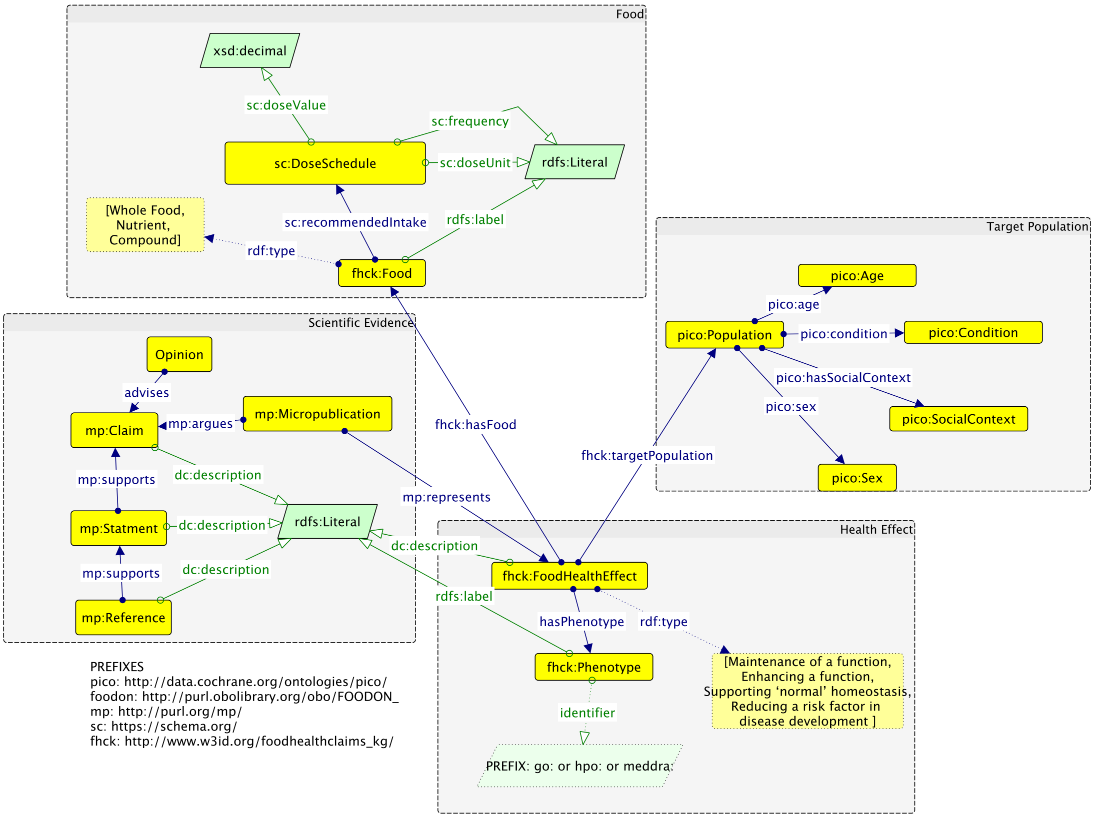

Figure 1: Data model for health claims made on foods.

The proposed data model for food health claim has four dimensions: Scientific Evidence, Food, Health Effect and Target Population. Each dimension is represented by an existing/proposed vocabulary. To define the scientific evidence on food health claims, we used a micropublication data model (shown as mp: prefix) which captures relationships between supporting statements (mp:Statement), literature references (mp:Reference) and claims (mp:Claim). To characterize foods, we created our own class definition for the type and source of a food (fhck:Food and fhck:FoodSource). We also used the schema.org vocabulary to describe the recommended intake for the food (sc:DoseSchedule). We used main phenotype vocabularies including GO, MEDDRA and HPO to describe the health effect mentioned (fhck:Phenotype) in a claim and classified the health effect into 4 types: (i) maintenance of a function, (ii) enhancing a function, (iii) supporting ‘normal’ homeostasis, and (iv) reducing a risk factor in disease development. Finally, we used the existing classes and properties from the PICO ontology to define the target population for a claim.

2.3. Health Effect

Scientific dossiers need to contain information on the putative beneficial effect that is caused by consuming the specific food or food ingredient. This is the second step of the risk assessment procedure conducted by EFSA [3, 4]. An important question is how a health effect within a claim is defined. In the medical domain, the health effect is very well defined and can be represented with many ontologies or terminologies like SNOMED-CT and ICD-10. However, these mostly focus on diseases or adverse health effects. Health claims on foods that refer to treating, curing or preventing a disease are not allowed [5] these are properties that can only be associated with medicinal products. Health claims can refer to health maintenance and improvement effects only [6]. We have used various vocabularies including UMLS, MEDDRA, HPO and GO to represent health effects of foods as the matching concepts cannot be found in a single ontology.

2.4. Target Population

With expanding knowledge about different biological responses to foods and food ingredients, the basis for personalised nutrition strategies [7], it becomes increasingly important to identify who will benefit more from a specific food or food ingredient. This affects the definition of the beneficial physiological effect, the second step for assessment of health claim dossiers [3, 4]. Although the health claims discussed in the EFSA opinions are usually defined for the general population, the characteristics (demographics and pre-existing conditions) of the targeted population for some claims are described in these opinions. We used the existing classes and properties from the PICO ontology [8, 9] to define the target population for a claim. This ontology models the population using information on age, gender, health condition and social context.

2.5. Scientific Evidence

The third and most important aspect of a scientific dossier is the data supporting the causal relationship between consuming a food (ingredient) and the suggested beneficial physiological effect. Whereas before the first batch of health claims had been assessed by EFSA's NDA Panel, it was insufficiently clear what type of evidence was needed to support putative claims [3], currently guidance documents have been developed to provide detailed insights into what type of evidence is needed to in health claim dossiers for all three steps, including supporting the causal relationship (ref to general guidance document EFSA). The scientific opinions analysed however were partially published prior to the development of these detailed guidance documents. These opinions are not always clear how the evidence for a health claim was assessed. We hypothesize that this uncertainty is due to the fact that this information is not sufficiently formalised. Regarding health claims, a claim and its evidence can be modelled using a micropublication data model [10]. In this model, the claim is defined as a natural-language sentence (e.g., "Calcium contributes to normal blood clotting") and its evidence (statements, data and references to scientific articles) can be modelled as shown in Figure 1.

3. Conclusion

With a large amount of scientific information available on the relationship between food (ingredients) and health, the knowledge graph described in this paper supports making such information more accessible and insightful. Increased accessibility and findability of such information may assist food business operators and consumers in developing and selecting functional foods for the general healthy population, as well as targeted nutrition interventions for specific population subgroups. Our knowledge graph has provided new insights into the causal evidence underlying various health claims, specifically related to the types of health relationships that can be expected from nutrition, as well as the supporting evidence for authorised claims. These findings, related to scientific assessments and the scientific basis for health claims and the finally authorised statements, provide relevant insights to business operators who want to make use of claims, and to food businesses and scientists who want to know what type of scientific information is needed to substantiate (specific types of) claims. Our results additionally show that the full health claim assessment has generated increasing insights into the differences that may be expected from and attributed to nutrition and medicines (and their relevant biomarkers of effects). Even though this knowledge graph is merely the first step in making information on food health claims more accessible and insightful, it is shown to be an essential step in making nutrition and health related information more user-accessible.

Acknowledgments

Research reported in this publication was partially supported by Limburg University Fund / SWOL. The work of AdB is supported by the Dutch Province of Limburg. Both funders had no role in the design nor reporting of the study.

References

[1] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, ACM Computing Surveys (CSUR) 54 (2021) 1–37.

[2] M. Krötzsch, Ontologies for knowledge graphs?, in: A. Artale, B. Glimm, R. Kontchakov (Eds.), Proceedings of the 30th International Workshop on Description Logics (DL 2017), volume 1879 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017.

[3] A. de Boer, E. Vos, A. Bast, Implementation of the nutrition and health claim regulation – The case of antioxidants, Regulatory Toxicology and Pharmacology 68 (2014) 475–487. URL: https://linkinghub.elsevier.com/retrieve/pii/S0273230014000221. doi:10.1016/j.yrtph.2014.01.014.

[4] K. G. Lenssen, A. Bast, A. de Boer, Clarifying the health claim assessment procedure of EFSA will benefit functional food innovation, Journal of Functional Foods 47 (2018) 386–396. URL: https://linkinghub.elsevier.com/retrieve/pii/S1756464618302640. doi:10.1016/j.jff.2018.05.047.

[5] H. Verhagen, E. Vos, S. Francl, M. Heinonen, H. van Loveren, Status of nutrition and health claims in Europe, Archives of Biochemistry and Biophysics 501 (2010) 6–15. URL: https://linkinghub.elsevier.com/retrieve/pii/S0003986110001451. doi:10.1016/j.abb.2010.04.012.

[6] B. van der Meulen, Medicinal claims: Prohibition, enforcement and delineation: Food in fact but medicine in law, Eur. Food & Feed L. Rev. 12 (2017) 392.

[7] J. M. Ordovas, L. R. Ferguson, E. S. Tai, J. C. Mathers, Personalised nutrition and health,

BMJ (2018) bmj.k2173. URL: https://www.bmj.com/lookup/doi/10.1136/bmj.k2173. doi:10.1136/bmj.k2173.

[8] C. Mavergames, S. Oliver, L. Becker, Systematic reviews as an interface to the web of (trial) data: using pico as an ontology for knowledge synthesis in evidence-based healthcare research., SePublica 994 (2013) 22–6.

[9] R. Leonardo, Pico: Model for clinical questions, Evidence Based Medicine and Practice 3 (2018) 2.

[10] T. Clark, P. N. Ciccarese, C. A. Goble, Micropublications: a semantic model for claims, evidence, arguments and annotations in biomedical communications, Journal of Biomedical Semantics 5 (2014) 28. URL: http://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-5-28. doi:10.1186/2041-1480-5-28.