

Title: The Ontology for Nutritional Studies (ONS) and its future directions

Authors: Francesco Vitali^{1*}, Agnese Gori², Paola Zinno³, Emily Schifano³, Chiara Devirgiliis³, Duccio Cavalieri^{2*}

1. Institute of Agricultural Biology and Biotechnology (IBBA) - National Research Council (CNR), Via Moruzzi 1, 56124, Pisa, Italy **2.** Department of Biology, University of Florence, Via Madonna del Piano 6, 50019, Sesto Fiorentino, Italy **3.** Research Centre for Food and Nutrition, CREA (Council for Agricultural Research and Economics), Via Ardeatina 546, 00178, Rome, Italy

*corresponding authors.

ONS: from nutritional studies to nutritional sciences

The ontology for nutritional studies (ONS) was originally developed as part of the European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI) project under the JPI-HDHL initiative. The ENPADASI project had the objective of developing an one-stop infrastructure for data/metadata sharing in nutritional studies, and for its distributed analysis. In that context, it served both as a mean of standardizing variables names in the metadata, and as a mean of harmonizing and facilitating the dialogue between different databases and different instances. The ONS was published in 2018¹ and is available for consultation and contribution on github at <https://github.com/enpadasi/Ontology-for-Nutritional-Studies> an on NCBO BioPortal (<http://bioportal.bioontology.org/ontologies/ONS>). The ontology was developed in compliance with the OBO Foundry principles (to which we recently joined), by the harmonization of multiple pre-existing ontologies with novel classes definition dealing with health and nutritional domain. The ONS includes classes and relations that are commonly encountered while conducting, storing, harmonizing, integrating, describing, and searching nutritional studies.

Widening from this initial intended application scenario of the nutritional studies to the broader scenario of nutritional sciences and practice, we included in ONS a core set of classes dealing and formalizing the “diet”, “food”, “food component” concepts. As a note to the reader, the words in quotation marks indicated classes actually formalized in ONS. Briefly, we modeled the “food” as a complex matrix bearing different constituents (i.e. “food component” and “nutrient”), entering the human nutrition with the “eating” process. The “diet” concept was modeled as an “information content entity” as in the domain of nutritional studies, it would be something deriving from surveys of from researchers/nutritionists prescription. The diet, is ultimately defined by the course of “eating” and “drinking” habit, which is recaptured by the “dietary pattern” class. In this vision, the diet is not something precisely defined, but rather a set of eating habits. For example, in this model we would not define the Mediterranean diet with precise nutritional values, but rather we could think of a more comprehensive definition including the traditional eating habits (i.e. “dietary pattern”) of the populations bordering the Mediterranean basin.

Integrating a more comprehensive “food” concept

Recent and upcoming development steps in ONS are increasingly dealing with the modeling of a more comprehensive view of food that expands from nutrients to recipes. Our idea is

¹ Vitali, F., Lombardo, R., Rivero, D., Mattivi, F., Franceschi, P., Bordoni, A., & Cavalieri, D. (2018). ONS: an ontology for a standardized description of interventions and observational studies in nutrition. *Genes & nutrition*, 13(1), 12.

that food exert benefits to our health that are greater than the benefits of the sums of its ingredients (as well as the nutrients, elements, and other molecules therein included). The vast majority of food is consumed after some modification processes (i.e. cooking or fermentation). During these processes, some molecules and metabolites originally present in the raw ingredients are transformed in the bioactive molecules of the final food product by a set of chemical, chemico-physical (i.e. the maillard reaction during meat cooking), or biological (i.e. fermentation by bacteria or yeasts) reactions. Moreover, following the set of instructions contained in a recipe to prepare a certain food/dish, especially if it originates in traditional cuisine, the order of actions and particular operations done can additionally modify a subset of those molecules (i.e. proven lycopene isomers formation from tomatoes thanks to “sofrito” operation²). It is also important to note that the same molecule can exist in food both in a biologically inactive or bio-available active form; a distinction that actually is not well recaptured by molecule ontologies like CHEBI (i.e. resveratrol). This complex set of bioactive molecules in food ultimately exert health effects and sustains our life.

In recent years, the term “microbiota” has raised in popularity. This word indicates the complex consortium of the microbial populations which lives in association with our body. It is now known that those community co-evolved with us, and exerts fundamental tasks for our health. A further level of complexity we aim to add to the food conceptual model reside in the microbiota and in its role in the inextricable flows between food, environment, and human host. Food, in fact, is not only providing nourishment directly available for our health, but is also providing building blocks for gut microbial community sustainment. As an example, fiber cannot be digested by the human metabolic machinery but are readily fermented by gut bacteria forming, among others, short chain fatty acids, which then can be absorbed by the human host. In addition, food, and especially fermented or traditional fermented foods, is also providing human host with live microorganisms through consumption. It is not easy to prove the effective transit and colonization of those microorganisms in the gut, but a growing body of evidences demonstrate that this field is far from extinction. We are developing a branch of ONS ontology dealing with standardized information on fermented food consumption. In modeling this branch, we aim at considering that a fermented food is providing bioactive molecules formed during the fermentation process, but is also providing a biochemical/genetic nutritional potential brought by the gene functions and genome functional pathways of live microorganisms. Due to its peculiar nature (i.e. presence of a complex fermentative microbiota and bioactive compounds) and the emerging interest in its implications on human health, we chose Kefir as an example to showcase the potentials and the requirements of our approach.

Kefir as use case of fermented food ontology

Kefir, originated in the Caucasus Mountains, is a traditional fermented milk containing over 50 species of microorganisms, including lactic acid bacteria, yeasts and acetic bacteria, identified from a spontaneous microbial starter culture known as kefir grains. Many beneficial effects have been studied following the consumption of kefir, such as cholesterol lowering, regulation of blood pressure, antitumor effects, immunomodulatory effects, antiobesity and

² de Alvarenga, J. F. R., Tran, C., Hurtado-Barroso, S., Martinez-Huélamo, M., Illan, M., & Lamuela-Raventos, R. M. (2017). Home cooking and ingredient synergism improve lycopene isomer production in Sofrito. *Food Research International*, 99, 851-861.

antidiabetic properties. Furthermore, kefir and its microorganisms can modulate gut microbiota and mycobiota exerting significant health benefits based on colonization abilities and production of bioactive compounds by foodborne microbes. One of the first concept we had to disentangle in developing this semantic model, was that of the difference between inoculum and starter culture concepts, which was of particular interest for the kefir use case. Kefir fermentation, in fact, is traditionally and industrially started by addition of kefir grains in milk. Kefir grains represent a complex matrix composed of a polysaccharide in association with bacteria and yeasts, which are employed as a starter culture for milk fermentation. From a semantic point of view, kefir grains are not independent continuant, but rather generically dependent continuant, as they are formed (i.e. depends on) as a consequence of kefir fermentation, and a chicken-or-egg like problem arose. In ONS we modeled this by distinguishing between “starter culture” and “microbial inoculum”. A starter culture is defined as a preparation of living microorganisms, at different degree of complexity, employed to start the fermentation process. Starter cultures are further classified into: “natural starter cultures” (undefined microbial cultures resulting from the development of the microbiota naturally present in the raw fermentable substrate and its environment) and “selected/commercial starter cultures” (defined microbial cultures derived from the selection of best performing natural starters reproduced under controlled conditions and characterized by highly reproducible performance). On the other hand, the “microbial inoculum” is defined as the vehicle for introduction of one or more microorganisms (representing starter culture) into a suitable substrate where microorganisms can grow activating their metabolic functions, providing a high level of viable biomass as well as several metabolites. Starter cultures are usually inoculated directly into food matrices in order to obtain desired and suitable changes in the finished product. Traditional fermented foods, including kefir, have long been manufactured with the use of natural starter cultures, by “backslopping”, defined as the use of a small amount of the finished or intermediate product to inoculate a new batch. The fermentation process itself is always performed by a “starter culture”, but the “microbial inoculum” (i.e. the vehicle of such a community) can be different. The modeling we propose was therefore useful from one side to distinguish the traditional fermentations from the industrial ones, and from the other to solve the chicken-and-egg like problem above described for kefir grains. For kefir, the fermentation process is initially started by the addition of a “starter culture”, which can be natural or industrially selected. As a result of this fermentation kefir grains are produced which, once recovered, can be used as a “microbial inoculum” for subsequent fermentation processes, through a “backslopping” procedure.

As stated, we aim at developing a semantic model for a more comprehensive view of food. For this reason, the future activities will be directed to include foodborne microorganisms and their functions in this growing model. In the case of fermented foods, microorganisms would be the main players of the food preparation process in terms of biological fermentation, and could also reach live and vital the human host gut through consumption. We are reviewing scientific literature to identify the main microbial species involved in kefir fermentation and to identify the main biochemical reactions involved (i.e. lactose fermentation) and their substrates (i.e. lactose) and products (i.e. lactic acid). In ONS, the principal microbial species in kefir will be imported from the NCBI taxonomy ontology, and connected to the fermentation process. We are evaluating strategies to import the gene annotation and pathways of available genomes of this microbial species catalogue. Once completed, this would represent the nexus among molecules in the raw ingredients and their

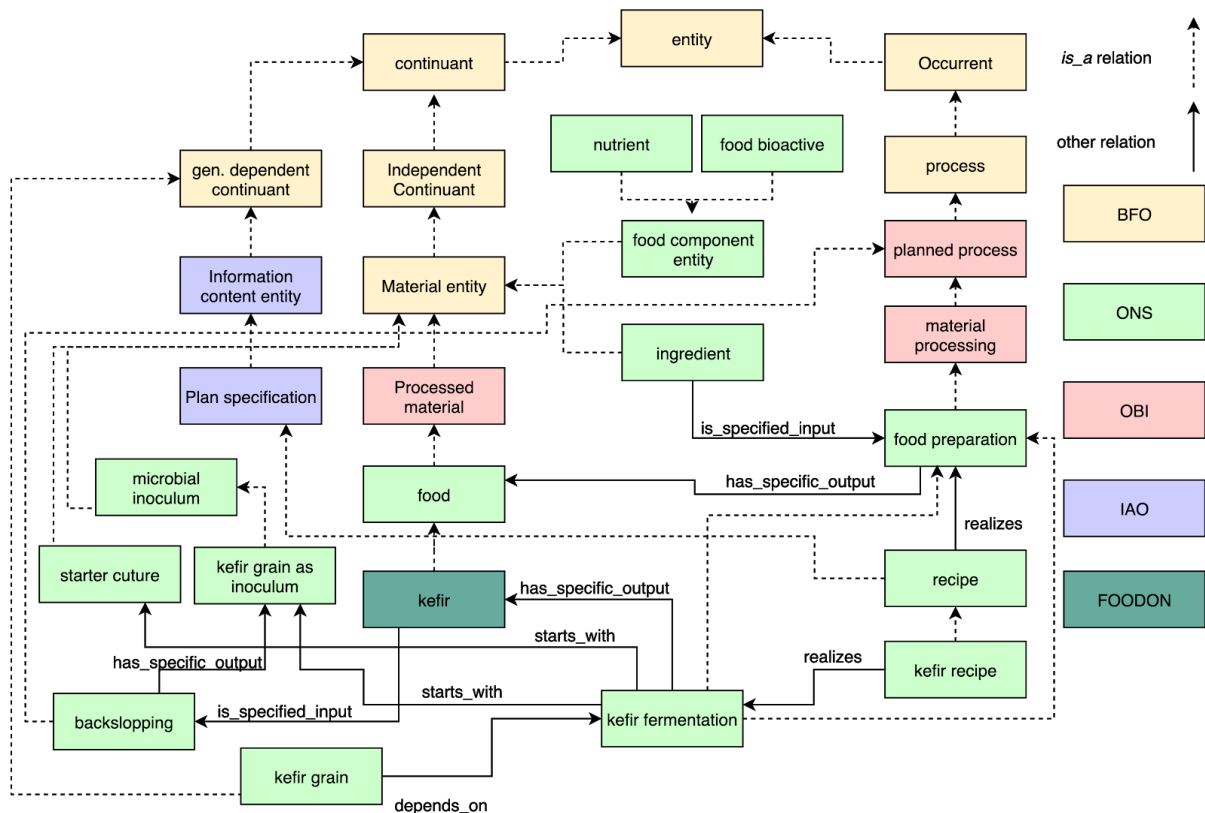


Figure 1. ONS model and schema with classes for modelling of kefir use case.

transformation into bioactive metabolites in the food by the action of microorganisms, as well as a complete catalogue of microbial species and molecular functions which are taken up by the human host following consumption of a fermented food. The use of precisely defined ontological terms will facilitate cross reference, cross link and discover information on Kefir present in the web or in the public domain to further support the claims associated to this product.

Conclusions

Food and nutrition are mediated by the processes that make food ingredients available and transformable. Among the players of this transformation, the microorganisms have a plethora of functions that render food components available to the human body. From the nutritional and computational standpoint, ontologies are the best and more appropriate way to categorize and make searchable this information in a coherent way.