PO^2 - A Process and Observation Ontology in Food Science. Application to Dairy Gels

Liliana Ibanescu[1(✉)], Juliette Dibie[1], Stéphane Dervaux[1], Elisabeth Guichard[2], and Joe Raad[1]

[1] UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005 Paris, France
{liliana.ibanescu,juliette.dibie,stephane.dervaux, joe.raad}@agroparistech.fr
[2] UMR 1324 INRA, UMR 6265 CNRS, Université de Bourgogne, 21000 Dijon, France

Abstract. This paper focuses on the knowledge representation task for an interdisciplinary project called Delicious concerning the production and transformation processes in food science. The originality of this project is to combine data from different disciplines like food composition, food structure, sensorial perception and nutrition. Available data sets are described using different vocabularies and are stored in different formats. Therefore there is a need to define an ontology, called PO^2 (Process and Observation Ontology), as a common and standardized vocabulary for this project. The scenario 6 of the NeON methodology was used for building PO^2 and the core component is implemented in OWL. By making use of PO^2, data from the project were structured and an use case is presented here. PO^2 aims to play a key role as the representation layer of the querying and simulation systems of Delicious project.

Keywords: Process and Observation Ontology · Domain ontology

1 Introduction

Recently, Europe faces two societal challenges: the increasing of overweight and obesity and the population aging. These problems, while having a tremendous impact on population life quality (e.g. poor health, social exclusion, increase in the need of assistance), are challenging the food industry to develop new strategies to produce well-balanced products in terms of nutritional requirements (e.g. less fat, sugar and salt) while using sustainable transformation processes. It is therefore crucial to better understand the food production system and a very interesting issue is to combine data and knowledge from different disciplines, like food composition in terms of nutrition, food digestion as a physiological process and sensorial perception of food.

Delicious project addresses the problem of analyzing the production and transformation processes of dairy gels using information available from different

E. Garoufallou et al. (Eds.): MTSR 2016, CCIS 672, pp. 155–165, 2016.
DOI: 10.1007/978-3-319-49157-8_13

collaborative projects concerning the food composition, food structure, mobility/bioavailability of flavor compounds and nutrients, sensory perception and digestibility. It involves domain experts and computer scientists researchers from INRA, the French National Institute for Agricultural Research. The expected result of Delicious project is to collect and structure the available data and knowledge into a data warehouse in order to enhance the analysis of the production process according to different cross-domain criteria. However, it is very difficult to take advantage of all the available data and knowledge from Delicious project. The main difficulty comes from the heterogeneity of their sources, the different inter-domain or cross-domain vocabularies, the different formalisms used according to the involved domain. A second challenge concerning the data integration task is the uncertainty quantification such as randomness, incompleteness, imprecision, vagueness, resulting from the natural variability of the domain and the lack of information. In order to address the question of the integration of knowledge and data, a relevant solution is the use of an ontology [4]. An ontology can be defined as a formal common vocabulary of a given domain, shared by the domain experts [7].

This paper present the Process and Observation Ontology, called PO^2, designed for Delicious project. The scenario 6 of the NeON methodology [2], i.e. reusing, merging and re-engineering ontological ressources, was used for building PO^2. The core component is implemented in OWL[1] and the domain component is under development.

By making use of the PO^2 vocabulary, the data sets available for the project were well-structured for the integration task. An use case is presented in order to show the complexity of this task.

This first step of building the PO^2 ontology allows to structure and organize the knowledge into a meaningful model at the knowledge level. This will lead to the possibility of designing more complex decision support systems allowing to compare different production scenarios and therefore suggesting improvements concerning the product quality while reducing the environmental impact. It may also help the field by giving hints about what data should be collected in order to perform an analysis concerning a target population (e.g. children or old people) or an cause and effect analysis. It may also provide the French food industry with the necessary tools to anticipate and develop future food products.

The paper is organized as follows. In Sect. 2, we present the ontology specification. In Sect. 3, the conceptualisation of PO^2 is detailed. In Sect. 4, we illustrate PO^2 through a use case. Finally, we conclude in Sect. 5 and present our further work.

2 Ontology Specification

Ontology specification was done during an iterative process. The ontology developers and the domain experts had a lot of meetings in order to identify (1) why

[1] https://www.w3.org/2001/sw/wiki/OWL.

the domain experts want to build an ontology (i.e. for what purpose), (2) what its intended users will be and (3) what are the main entities.

First, the purpose of building an ontology is to provide a consensual model of the production and transformation of dairy gels and to solve the lack of communication between domain experts. Available data were gathered for many different purposes by different experts with their own experimental itineraries, vocabularies and technical materiel and methods. There is an obvious need to build a common and shared structured vocabulary.

Second, the intended users are researchers in several distinct domains: nutrition, microbiology, biochemistry, physico-chemistry, chemistry, process engineering, food science and sensory analysis. Reaching a consensus about a common vocabulary was therefore a hard task. The ontology developers and the 15 domain experts involved in Delicious project spent about 20 h using CMap Tool[2] to identify a vocabulary common to all the involved experts. The resulting vocabulary was unstructured and composed of approximately 500 entities dealing with composition, structure, technical and physiological transformation processes, mobility and bioavailability of small molecules in relation with sensory perception and nutritional value. It proposes a first representation of the explicite and implicite knowledge of all the involved domain experts.

Third and finally, in order to investigate how to structure the vocabulary, we focused on a small representative subset of data and knowledge concerning the *In the mouth* process. Taking into account the previously identified entities, relying on available documents [1,5] and data and in close collaboration with domain experts of the target domain, entities were grouped into three main parts (see Fig. 1):

- the part concerning the production and transformation process which contains the concepts: `process`, `itinerary` and `step`;
- the part concerning the participant which contains the concepts: `product`, `mixture`, `material` and `sensing device`;
- the part concerning the observation which contains the concepts: `observation`, `scale`, `sensor output`, `computed observation`, `method` and `measure`.

We therefore reached a consensus about a common structured vocabulary with the following specifications. An `itinerary` is an execution of a production or transformation process, i.e. a set of interrelated steps. A `step` is characterized by its participants and its temporal duration/interval. A `participant` may be a `mixture`, a `material` or a `sensing device`. Each participant is characterized by its experimental conditions. Moreover a `mixture` is characterized by its composition. An `observation` observes a `participant` at a certain `scale` during a `step`. It is characterized by some `participants` such as a given `material` or a `sensing device` and implements a `method`. It has for result a `sensor output` and/or `a computed observation`, each of them can have for value a function or a simple measure. A `measure` is characterized by either a quantity and a unit of measure or a symbolic concept and a measurement scale.

[2] http://cmap.ihmc.us.

Fig. 1. The three main parts of the ontology for Delicious

3 Ontology Conceptualisation

The ontology conceptualization follows the Scenario 6 of the NeON methodology [2], i.e. reusing, merging and re-engineering ontological ressources. A number of existing ontologies have been analyzed: the supply chain ontology [6], the bussiness process ontology [9], the ontology for wine production [8], SSN[3], BFO[4], IAO[5] and $[MS]^2O$ (Multi Scales and Multi Steps Ontology)[6].

Based on our experience and after a careful analysis, it was decided that the best method to adopt for building the PO^2, Process and Observation Ontology, is to re-engineer the core component of $[MS]^2O$, an ontology designed for a project concerning the representation of the production of stabilized micro-organisms (see [3] for more details). This re-engineering task of $[MS]^2O$ was done with the two following main concerns:

- establish a clear distinction between a process and its participants which was achieved by reusing BFO;
- link all together the observations with the step where they occur, their participants, their materials and methods and their measures reusing IAO (Information Artifact Ontology) an ontology of information entities.

The PO^2 core component is given in Fig. 2. The concepts identified in Sect. 2 during the ontology specification are represented as nodes and the relations between the concepts are represented as arrows.

The PO^2 core component is implemented in OWL and it is available at http://agroportal.lirmm.fr/ontologies/PO2. The domain component is under development.

[3] http://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628.
[4] https://bioportal.bioontology.org/ontologies/BFO.
[5] https://bioportal.bioontology.org/ontologies/IAO.
[6] http://lovinra.inra.fr/2015/12/16/multi-scale-multi-step-ontology/.

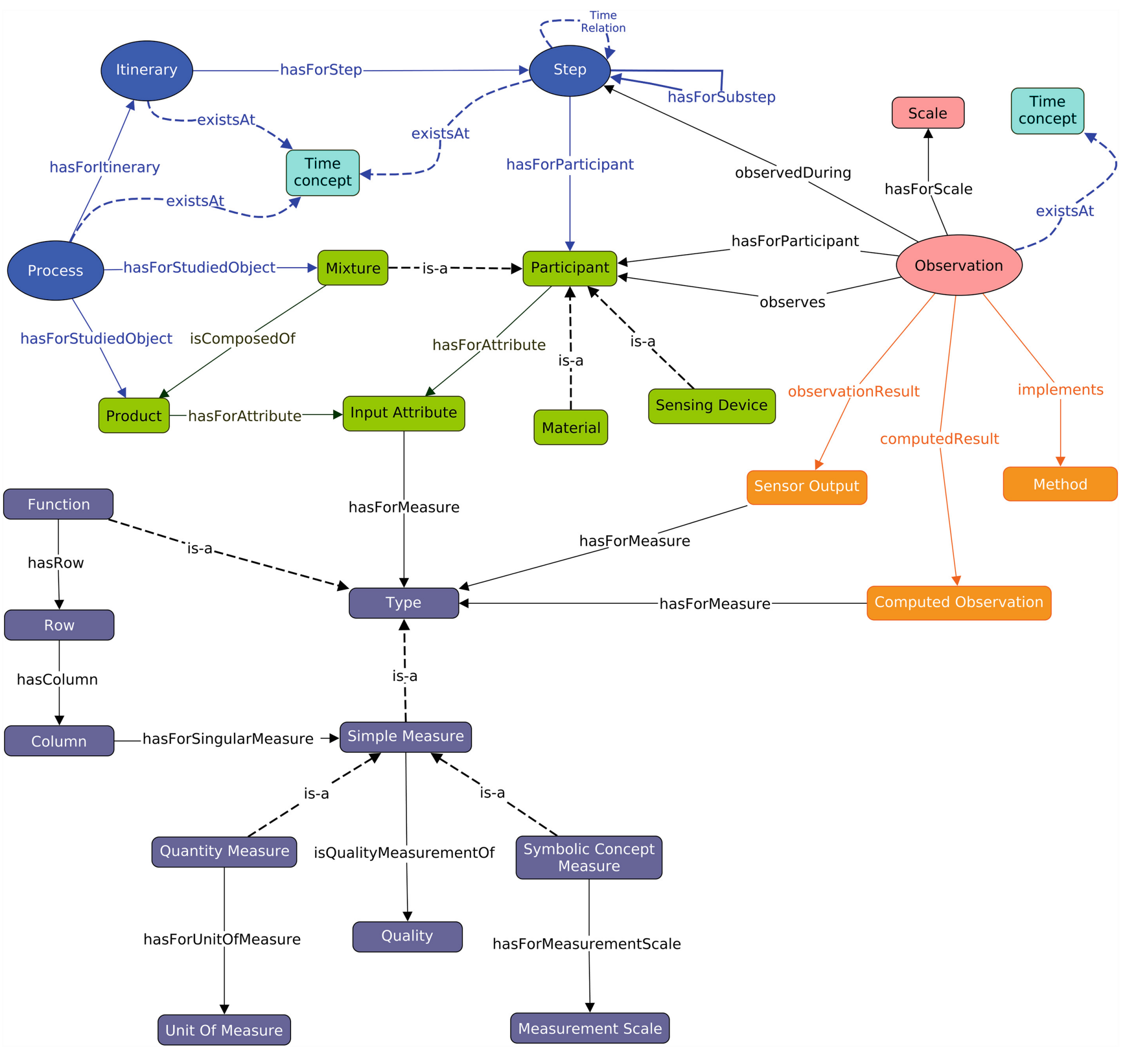

Fig. 2. PO^2 core component

4 PO^2 Use Case

This section presents an use case concerning the *In the mouth* process, in order to show the complexity of the representation task.

At the beginning of the Delicious project, data and knowledge concerning this process were available in different vocabularies and formats. By making use of the vocabulary from the PO^2 core component presented in Sect. 3, the data concerning the studied use case were structured into 20 EXCEL files:

– 2 files describe the *In the mouth* process (e.g. Fig. 3),
– 11 files describe the mixture composition (e.g. Fig. 4),
– 6 files describe experimental observations (e.g. Fig. 6), and
– 1 file describes the materials and methods with 29 methods and 16 materials (e.g. Figs. 8 and 9).

Date	2012
Project name	CARREDAS PhD BOISARD
Sample name	cheese model
Sample code number	L20P28

Step	Sub-step	Measure	Scale	Files numbre	Data type
Before putting in the mouth	NA	Composition	Mixture	1	Raw
Before putting in the mouth	NA	Itinerary	Mixture	1	Raw
Before putting in the mouth	NA	Characteristics	Mixture	10	Raw and computed
In the mouth	Chewing	Study of the sodium release in the saliva	Mixture	1	Raw
In the mouth	Chewing	Chewing activity	Step	1	Raw and computed
In the mouth	Chewing	Study of the flavour components release in the air	Mixture	1	Computed
In the mouth	Chewing	Sensory properties	Mixture	1	Raw and computed
In the mouth	Swallowing	Swallowing number	Step	1	Raw
In the mouth	Swallowing	Study of the flavour components release in the air	Mixture	1	Raw and computed

Fig. 3. The EXCEL file which describes the *In the Mouth* process

Let us notice that these EXCEL files allow the domain experts to collect and re-structure the available data using the PO^2 vocabulary. Moreover, these files can be automatically translated into instances of PO^2 (see e.g. Figs. 5 and 7).

In Fig. 3 the description of *In the mouth* `process` is given: it contains one `itinerary` which is composed of two steps: the *Before putting in the mouth* `step` and the *In the Mouth* `step`. The last step is composed of two sub-steps: *Chewing* and *Swallowing*.

This process has for studied object a sample of the `mixture` *cheese model* identified by the code number *L20P28*. This mixture is composed of ten products as described in Fig. 4, each `product` being characterized by the `input attribute` *Weight*.

Figure 5 gives an example of an instance extracted from Fig. 4: the `mixture` *L20P28* is composed of the `product` *Rennet casein* where its `input attribute` *Weight* has for `simple measure` the value 238.3 of `unit of measure` g/kg *of cheese model*.

Product description		
Sample code number	L20P28	
Name of the description file	2012-CAREDAS-001-FicheDescriptif	
Raw data		
Characteristics	*Value*	*Unit*
Rennet casein	238.3	g/kg of cheese model
Acide casein	59.6	g/kg of cheese model
AMF	200.2	g/kg of cheese model
Deionized water	464.4	g/kg of cheese model
Melting salt	25	g/kg of cheese model
Added Nacl	0	g/kg of cheese model
Citric acid	2.5	g/kg of cheese model
Aromatic solution + water	10	g/kg of cheese model
lipids	20	g/kg of cheese model
proteins	28	g/kg of cheese model

Fig. 4. The EXCEL file which describes the composition of the mixture *L20P28*

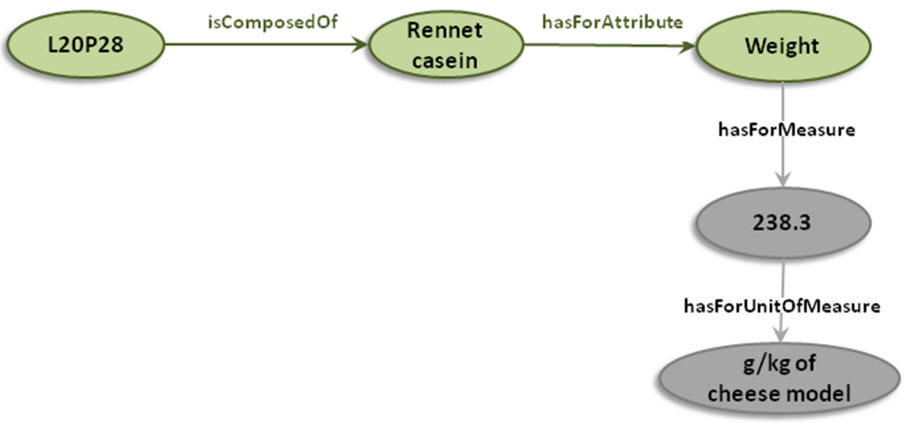

Fig. 5. An example of instance concerning a mixture and its composition

Let us now focused on an experimental observation of the *In the Mouth* process as described in the EXCEL file of Fig. 6. This instance of observation, called in the following *Observation1*, has the following properties (see Fig. 7):

– is observed during the `sub-step` *Chewing* of the `step` *In the mouth*;
– observes the `mixture` *L20P28*;
– has for `participants` the two `materials`: *Material 1* and *Material 14* as described in Fig. 8;
– has for scale the molecular `scale`;
– has for `date` 10/09/2012;
– implements the two `methods`: *Method 22* and *Method 23*, both described in Fig. 9;
– has for observation result the `sensor output` *Sodium concentration in the saliva* which is function of the sodium concentration during time;
– has for computed result the `computed observation` *yield curve of the release* which has for `measure` the value 2.75 of `unit` mM.

General informations				
Date of the measure	10/09/2012			
Time				
Duration of the manipulation				
Step	In the mouth			
Sub-step	Chewing			
Scale	Molecular			
Repetitions	1			
Product description				
Sample code number	L20P28			
Name of the description file	2012-CAREDAS-001-FicheDescriptif			
Material used				
Identifier	Material 1			
Identifier	Material 14			
File	2012-carredas etapesenbouche-001-M&M			
Method used				
Identifier	Method 22			
Identifier	Method 23			
File	2012-carredas etapesenbouche-001-M&M			
Raw data				
Characteristics	*value*	*Unit*	*Incertainty*	*Description*
Sodium concentration in the saliva at T0	0	mM		
Sodium concentration in the saliva at T5	10	mM		
Sodium concentration in the saliva at T15	35	mM		
Sodium concentration in the saliva at T30	55	mM		
Sodium concentration in the saliva at Td	70	mM		
Computed data				
Characteristics	*value*	*Unité*	*Incertainty*	*Description*
Yield curve of the release = sodium concentration by tooth bite	2.75	mM		

Fig. 6. The EXCEL file which describes an experimental observation during the sub-step *chewing* of the step *In the mouth* for the mixture *L20P28*

What it is interesting to report about our experience with this use case is that the process of building the ontology is an iterative one. Notice that the EXCEL files of Figs. 3, 4, 6 and 8 contain well structured data and knowledge, but the EXCEL file of Fig. 9 describing the methods with many textual informations, is currently unusable for automatic querying. Domain experts were not able up to now to express their needs about the querying concerning the different methods they used in the different domains. The lessons they learned while they organized and structured their data and knowledge according to the concepts from PO^2 give them the understanding that allow to refine the specification concerning the methods. This is an ongoing process.

To conclude, we would like to stress on the fact that the complexity of the knowledge representation task of this use case allows us to identify a common and shared structured vocabulary that encompasses almost all the domains involved in the Delicious project.

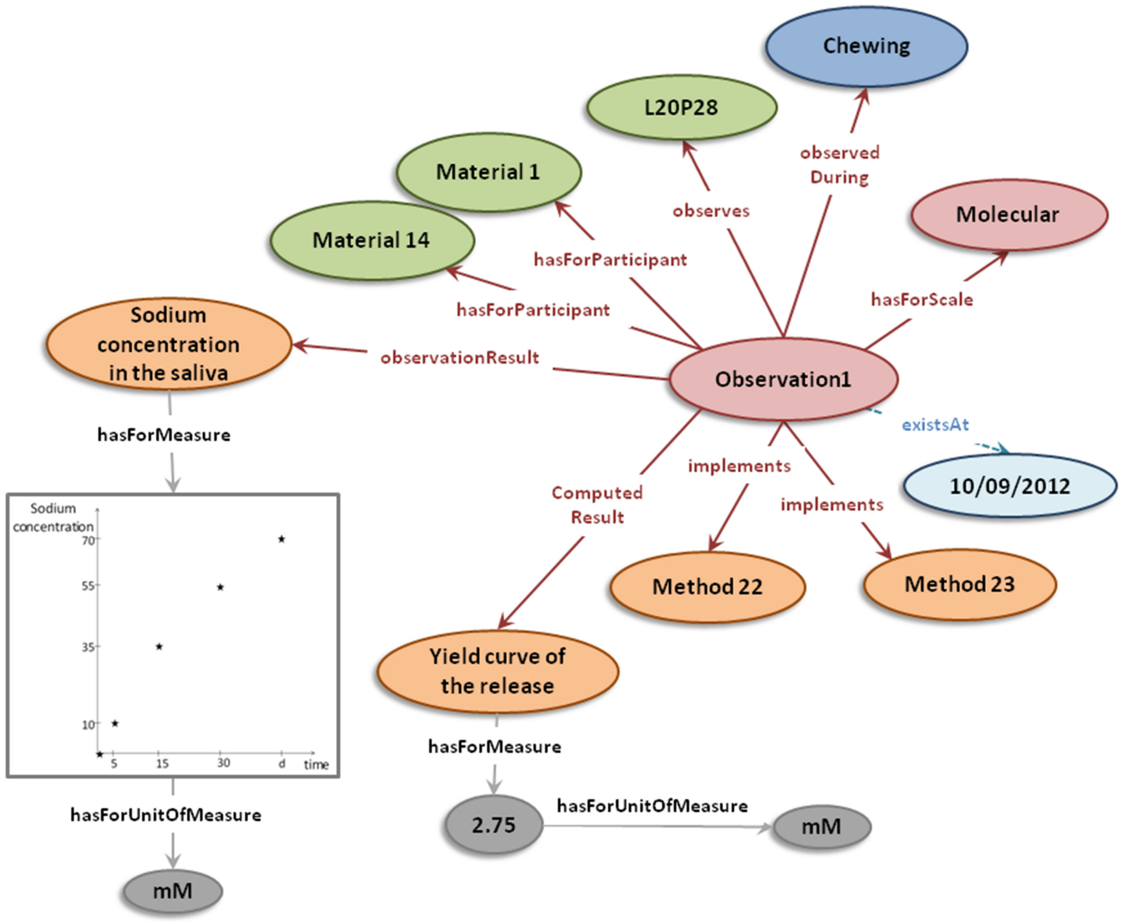

Fig. 7. An example of instance representing an experimental observation *Observation1* during the sub-step *Chewing* for the mixture *L20P28*

Material 14		
Sensing device	x	
Transformation device		
Device used	HPLC (dionex ICS 3000)	
Column	IonPac CS12A 5µm	
Column temperature	25	°C
Eluent	Sulphuric acid 11mM	
Eluent flow	0,5	mL/min
Detector	conductometer	
Suppressor	CSRS 300 in 2mm (courrent 33 mA)	
Software	UCI 100 Chromeleon	
Additional information	The sodium concentrations were obtained from a calibration curve	

Fig. 8. The EXCEL file which describes the *Material 14* used in the experimental observation *Observation1* of Fig. 7

Method 22	
Method name	Itinerary sample cut and saliva gathering
Method description at step 1	The sample of 5g of the cheese model L20P28 is proposed at 13°C
Method description at step 2	The saliva was spat during the chewing at T5 (5 tooth bites), T10, T15, T30 and Td (before swallowing) in 5 tubeS eppendorfs put after in a basin of crushed ice
Method description at step 3	The tubs were centrifuged during 15min at 15300g at 4°C
Method description at step 4	50µL were levyed from thes du supernatant et dropped again in a tube eppendorf with 950µL of water Milli-Q
Method description at step 5	2mL were levyed with a syringe in order to be filtered (filter in nylon 13mm, pore size 0.45 µL)
Method description at step 6	then put in a flat base flask of 2mL.
Method 23	
Method name	Determination of the sodium concentration in the saliva by HPLC
Method description at step 1	used of HPLC

Fig. 9. The EXCEL file which describes the *Method 22* and the *Method 23* used in the experimental observation *Observation1* of Fig. 7

5 Conclusion

In this paper we presented the building of PO^2, a Process and Observation Ontology, designed for a cross-domain project concerning the production and transformation of dairy gels. The core component of PO^2 is the result of re-engineering $[MS]^2O$, using BFO and IOA. A use case on an *In the Mouth* process was presented.

Further work is to express users requirements through competency questions and prioritizing those requirements. Then the domain component will be developed and the ontology will be validated against the competency questions.

PO^2 aims to play a key role as the representation layer of the querying and simulation system of Delicious project. This leads to the possibility of comparing different production systems and may also help to develop a decision support system taking into account the uncertainty of data.

The developed ontology could be further adapted to other types of food products, such as bakery, vegetable or meat products. This may provide to the French food industry tools in order to develop food products according to the nutritional recommendation for a healthy population while increasing efficiency and adopting an eco design approach.

References

1. Boisard, L., Andriot, I., Martin, C., Septier, C., Boissard, V., Salles, C., Guichard, E.: The salt and lipid composition of model cheeses modifies in-mouth flavour release and perception related to the free sodium ion content. Food Chem. **145**, 437–444 (2014)

2. Suárez-Figueroa, M.C., Gómez-Pérez, A., Fernández-López, M.: The NeOn methodology for ontology engineering. In: Suárez-Figueroa, M.C., Gómez-Pérez, A., Motta, E., Motta, A. (eds.) Ontology Engineering in a Networked World, pp. 9–34. Springer, Heidelberg (2012). doi:10.1007/978-3-642-24794-1_2
3. Dibie, J., Dervaux, S., Doriot, E., Ibanescu, L., Pénicaud, C.: $[MS]^2O$ – a multi-scale and multi-step ontology for transformation processes: application to micro-organisms. In: Haemmerlé, O., Stapleton, G., Faron Zucker, C. (eds.) ICCS 2016. LNCS (LNAI), vol. 9717, pp. 163–176. Springer, Heidelberg (2016). doi:10.1007/978-3-319-40985-6_13
4. Doan, A., Halevy, A.Y., Ives, Z.G.: Principles of Data Integration. Morgan Kaufmann (2012)
5. Feron, G., Ayed, C., Qannari, E.M., Courcoux, P., Laboure, H., Guichard, E.: Understanding aroma release from model cheeses by a statistical multiblock approach on oral processing. PLoS ONE **9**(4), 1–15 (2014)
6. Grubic, T., Fan, I.S.: Supply chain ontology: review, analysis and synthesis. Comput. Ind. **61**(8), 776–786 (2010)
7. Guarino, N., Oberle, D., Staab, S.: What is an ontology? In: Staab, S., Staab, R. (eds.) Handbook on Ontologies. International Handbooks on Information Systems, vol. 2009, pp. 1–17. Springer, Heidelberg (2009). doi:10.1007/978-3-540-92673-3_0
8. Muljarto, A.-R., Salmon, J.-M., Neveu, P., Charnomordic, B., Buche, P.: Ontology-based model for food transformation processes - application to winemaking. In: Closs, S., Studer, R., Garoufallou, E., Sicilia, M.-A. (eds.) MTSR 2014. CCIS, vol. 478, pp. 329–343. Springer, Heidelberg (2014). doi:10.1007/978-3-319-13674-5_30
9. Rospocher, M., Ghidini, C., Serafini, L.: An ontology for the business process modelling notation. In: Garbacz, P., Kutz, O. (eds.) Formal Ontology in Information Systems - Proceedings of the Eighth International Conference, FOIS , 22–25 September 2014, Rio de Janeiro. Frontiers in Artificial Intelligence and Applications, vol. 267, pp. 133–146. IOS Press (2014)