



HAL
open science

Besoins ontologiques pour la transformation des aliments

Damion Dooley, Magalie Weber, Liliana Ibanescu, M. Lange, L. Chan, L. Soldatova, C. Yang, R. Warren, C. Shimizu, H. McGinty, et al.

► To cite this version:

Damion Dooley, Magalie Weber, Liliana Ibanescu, M. Lange, L. Chan, et al.. Besoins ontologiques pour la transformation des aliments. IC2023 : 34es Journées francophones d'Ingénierie des Connaissances, Jul 2023, Strasbourg, France. hal-04167547

HAL Id: hal-04167547

<https://hal.science/hal-04167547>

Submitted on 20 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Besoins ontologiques pour la transformation des aliments

D. Dooley¹, M. Weber², L. Ibanescu³, M. Lange⁴, L. Chan⁵, L. Soldatova⁶, C. Yang⁷, R. Warren⁸,
C. Shimizu⁹, H. McGinty¹⁰, W. Hsiao^{1,11}

¹ Centre for Infectious Disease Genomics and One Health, Simon Fraser University, Burnaby, BC, Canada

² INRAE, UR BIA, Biopolymères Interactions Assemblages, Nantes, France

³ Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay, Palaiseau, France

⁴ International Center for Food Ontology Operability Data and Semantics, Davis, California, USA

⁵ College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA

⁶ Goldsmiths College, the University of London, UK

⁷ Department of Food Technology, Safety and Health, Ghent University, Ghent, Belgium

⁸ Glengarry Agriculture and Forestry, Guelph, On, Canada

⁹ Kansas State University, Manhattan, KS, USA

¹⁰ Department of Chemistry and Biochemistry, Ohio University, Athens, Ohio, USA

¹¹ Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

liliana.ibanescu@agroparistech.fr

Résumé

La nourriture est souvent appréciée pour ses aspects sensoriels, festifs et pour sa contribution à la santé, mais derrière se cachent de nombreux autres processus de production agricole, de distribution, de fabrication et des processus physiologiques qui peuvent maintenir ou miner la santé de la population. La complexité de ces processus est évidente à la fois dans la préparation quotidienne des repas et dans la fabrication industrielle, l'emballage et le stockage des aliments. Dans ce voyage de la ferme à la fourchette il n'existe pas encore d'ontologie pour couvrir toutes les étapes et il semble nécessaire de construire une telle vision en réutilisant des ontologies existantes pour des sous-domaines, organisés par des experts. Le défi est que cette fusion soit, par analogie, une seule langue, plutôt que des noms et des verbes d'une douzaine de dialectes. Ce travail se concentre sur les composants de l'ontologie – propriétés des objets et des données et annotations - nécessaires pour modéliser les processus alimentaires ou, plus général, la modélisation de processus, dans le contexte de l'Open Biological and Biomedical Ontology Foundry et des ontologies congruentes.

Mots-clés

Ontologie, transformation des aliments, recette, modélisation des processus, OBO Foundry

Abstract

People often value the sensual, celebratory, and health aspects of food, but behind this experience exists many other value-laden agricultural production, distribution, manufacturing, and physiological processes that support or undermine a healthy population. The complexity of such pro-

cesses is evident in both every-day food preparation of recipes and in industrial food manufacturing, packaging and storage. An integrated ontology landscape does not yet exist to cover all the entities at work in this farm to fork journey. It seems necessary to construct such a vision by reusing expert-curated fit-to-purpose ontology subdomains. The challenge is to make this merger be, by analogy, one language, rather than nouns and verbs from a dozen or more dialects. This work focuses on the ontology components – object and data properties and annotations – needed to model food processes or more general process modelling within the context of the Open Biological and Biomedical Ontology Foundry and congruent ontologies.

Keywords

Ontology, food processing, recipe, process modelling, OBO Foundry

1 Introduction

La transformation des aliments a, comme Janus, deux visages opposés, offrant d'une part des progrès significatifs qui rendent les aliments plus résistants à la détérioration, et donc plus faciles à conserver, plus nutritifs, plus savoureux, plus pratiques et souvent peu coûteux [17]. D'autre part, nombre de ces mêmes caractéristiques ont encouragé la surconsommation, les régimes alimentaires malsains et des produits alimentaires qui perturbent la dynamique du microbiome intestinal humain et la santé en général [13]. La recherche en génie des procédés est apparue dans les années 1950 pour comprendre les implications économiques et de santé humaine d'un système de fabrication alimentaire de plus en plus industrialisé et mondialisé, en prenant en compte la transformation, la production, la manutention,

le stockage, la conservation, le contrôle, l'emballage et la distribution des produits alimentaires. L'ingénierie alimentaire est un domaine multidisciplinaire qui combine la microbiologie, les sciences physiques appliquées et le génie chimique pour comprendre et concevoir des produits et des opérations dans les industries de l'alimentation.

Les citoyens sont de plus en plus préoccupés par les questions de santé et l'impact sur l'environnement, et ils recherchent également des aliments savoureux, naturels et pratiques. Aujourd'hui, l'ingénierie alimentaire s'appuie sur les données que les systèmes d'information apportent pour répondre aux exigences de santé grâce à une nutrition de précision s'appuyant sur des connaissances scientifiques afin de répondre aux besoins nutritionnels des individus en fonction de différents facteurs, tels que l'âge, l'état de santé ou une pathologie, les facteurs liés à l'activité ou au mode de vie et d'éviter des aliments ou ingrédients qui produisent des réponses allergiques. Les fabricants de produits alimentaires trouvent également une valeur ajoutée à travers l'adaptation de leurs produits aux demandes des consommateurs qui basent aujourd'hui leurs choix alimentaires sur certains critères éthiques (par exemple, le bien-être des animaux), qui vont au-delà des courants de consommation particuliers que sont le casher, le halal et le végétarisme.

La prise de décision complexe et interdépendante au sein d'un paysage d'acteurs multiples —de la production, de la distribution, de la consommation et de la réglementation— nécessite un registre normalisé des informations collectées tout au long des cycles de vie des produits alimentaires. Les systèmes de jumeaux numériques construits à partir de normes d'identification des produits comme le GS1 Digital Link, combinés à une infrastructure de graphe de connaissances comme Origin Trail [21], fournissent aux parties prenantes un mécanisme de saisie des données contextuelles sur les produits, de la ferme à l'assiette, mais ce système sera peu utilisable s'il ne s'agit que d'un pot-pourri de langages restant chacun cloisonné et propre à son contributeur. Les participants à cette chaîne d'information bénéficieront de la mise au point d'une ontologie — un langage commun, standardisé, exploitable par les machines — qui recouvre la taxonomie et l'anatomie des organismes sources des matières premières, les informations sur le contexte de la production agricole et du transport, ainsi que les méthodes de transformation et de préparation des aliments.

L'un des principaux objectifs de l'analyse présentée dans notre article [11] est de formuler des recommandations pour un cadre générique de modélisation des processus qui soit basé sur l'initiative et les recommandations d'OBO Foundry (Open Biological and Biomedical Ontology Foundry) [15]. Notre objectif principal est de passer en revue des ontologies OWL qui possèdent des classes et des propriétés nécessaires à un modèle de processus OBO. Cette analyse nous a permis, d'une part, d'identifier les entités nécessaires pour modéliser la transformation des aliments et, de manière plus générale, les protocoles de laboratoire ou les recettes de fabrication, et, d'autre part, de comparer les différentes manières de représenter les dépendances tempo-

relles, les enchaînements, les entrées et les sorties des processus, les participants ou agents impliqués dans un processus ou une partie du processus. Nous avons examiné l'ontologie Exact2 (EXperimental ACTions) [23], l'ontologie PO2 (Process and Observation Ontology) [10], le modèle de processus planifié OBI (Ontology for Biomedical Investigations)[9], qui présentent divers degrés de compatibilité avec OBO, ainsi que l'ontologie PROV-O [20], l'ontologie OWL Time [18] et l'ontologie SOSA (Sensor, Observation, Sample, and Actuator) [16], dont l'adoption a connu un essor considérable. Enfin, nous avons analysé le modèle de recette de Schema.org, qui comporte un certain nombre d'éléments liés aux processus.

Après cet examen, nous constatons que certaines propriétés objets peuvent être réutilisées telles quelles (comme OWL-Time "hasTime"), tandis que d'autres nécessiteraient la création d'entités OBO équivalentes, pour respecter les principes OBO. Ces entités peuvent ensuite être ajoutées à l'ontologie alimentaire FoodOn [12], qui est un pivot au sein de l'initiative OBO permettant de relier les caractéristiques des produits alimentaires et les aspects liés à la transformation des aliments, ce qui est essentiel pour la traçabilité des aliments et d'autres applications de la ferme à la fourchette. Enfin nous proposons un modèle pour représenter un processus de transformation alimentaire se basant sur la comparaison avec les autres ontologies et nous concluons en présentant un cas d'utilisation sous la forme d'une simple recette de purée de carotte ou "carottes bouillies". Nous décomposons la recette en étapes, avec des informations qui pourraient satisfaire le consommateur qui a simplement besoin d'instructions pour réaliser la recette à la maison, l'industriel qui recherche une formulation pour cibler des caractéristiques précises pour le produit final et traduire cette formulation en processus industriel. Le modèle de recette que nous proposons est déjà utilisé dans un article récent [24] qui propose un nouveau modèle de conception ontologique pour étudier les substitutions possibles d'ingrédients au sein d'une recette [4].

Dans cet article en français nous avons fait le choix de ne présenter que le cadre général et le modèle de FoodOn détaillé dans article [11].

2 Modélisation de processus basée sur une ontologie

2.1 OBO Foundry : le cadre

Nous proposons de faire la modélisation des processus dans le cadre d'OBO Foundry parce que nous sommes favorables à une stratégie qui consiste à avoir un ensemble minimal de propriétés de données, un ensemble réduit de propriétés d'objets et à mettre davantage l'accent sur la définition des classes d'entités reliées par une propriété d'objet. Une deuxième motivation pour utiliser OBO - bien qu'elle ne soit pas propre à son cadre - est d'encourager la normalisation en réduisant le nombre de termes sémantiquement dupliqués dans les ontologies membres, promouvant ainsi une compréhension encyclopédique. Les ontologies membres

de OBO sont encouragées à réutiliser une grammaire de base des relations, fournie principalement par l'ontologie des relation RO (Relation Ontology) [1]. Cela permet de respecter le deuxième engagement de la communauté OBO, à savoir que les ontologies membres soient logiquement compatibles les unes avec les autres. En ce qui concerne la cohérence logique, OBO a tacitement encouragé le raisonnement sur la structure logique d'une ontologie donnée fusionnée avec l'ontologie formelle de base (BFO) [8] pour mettre en évidence les incohérences internes. Plus récemment, un autre point de départ de la compatibilité est le COB (OBO Core Ontology for Biology and Biomedicine) [22], une ontologie en cours de développement qui comprend un ensemble réduit de classes et de relations de niveau supérieur couramment utilisées - entité matérielle, processus, caractéristique et information - combinées en une seule ressource avec pour objectif principal de prendre en charge les contraintes de domaine et de co-domaine des relations de RO.

Pour s'intégrer dans la communauté OBO, une ontologie doit mettre en œuvre les principes OBO. Si des composants utiles ne sont pas réutilisables tels quels dans l'OBO, ils doivent être remplacés par des termes comparables répondant aux critères de l'OBO [3], dont voici quelques exemples :

- Gestion des URL permanentes (PURL) : Chaque terme de l'ontologie se voit attribuer une URL et est rattaché à un service qui renvoie des informations lisibles par l'homme et par l'ordinateur sur le terme. L'URL du terme est censée exister en permanence ; il existe un système de référence pour la dépréciation et le remplacement des termes qui facilite les mises à jour de la base de données en fonction de l'évolution des ontologies.

- Normes de curation : Les termes sont désignés au singulier, en anglais et en minuscules, sauf pour les noms propres. Chaque terme est accompagné d'une définition aristotélicienne qui fait référence à la classe d'origine et la différencie de ses frères et sœurs. Les auteurs des termes et les sources des définitions sont mentionnés.

- Axiomatisation : Les termes sont, dans une certaine mesure, logiquement reliés par des relations avec d'autres entités.

- Collaboration : Une ontologie importe le terme d'une autre ontologie plutôt que de reproduire la même sémantique dans l'un de ses propres termes conformément au principe de l'information minimale pour référencer un terme d'ontologie externe (MIREOT) [14]. Au sein d'OBO Foundry, des ontologies "de référence" définissent des domaines qui sont les ressources de référence pour les autres ontologies qui en ont besoin, donc idéalement une ontologie pour la taxonomie, une pour la chimie, une pour l'anatomie, etc., mais il faut admettre que ce niveau de qualité est encore loin d'être atteint.

2.2 Capacités attendues du modèle de processus

Il existe différents types de processus physiques, chimiques, biologiques ou de transformation des données, ainsi que

des processus intentionnels qui les exploitent. Nous pouvons modéliser des processus non planifiés présents dans le monde physique, qui peuvent ensuite être exploités par des processus planifiés impliquant un ou plusieurs objectifs. Par exemple, le mûrissement des fruits, un processus biologique, peut être manipulé par un mûrissement artificiel et/ou une récolte et un transport planifiés qui, ensemble, répondent à l'objectif de livrer des fruits mûrs aux clients. Les processus planifiés ont un plan ou un protocole détaillé à réaliser, et agissent sur des entrées et des sorties qui sont des entités matérielles ou des données. Un processus planifié peut impliquer des transformations intrinsèques ou extrinsèques :

- Transformer la composition d'une entité par des processus de mécaniques, chimiques, biologiques ou autres processus de production physiques.

- Caractériser (générer des informations sur) une entité en utilisant des processus d'observation qui peuvent être invasifs ou non invasifs.

- Modifier le contexte relatif ou les relations extrinsèques d'une entité, comme le transport d'objets d'un endroit à un autre.

- Affecter passivement une entité, par exemple un objet stocké dans un objectif de conservation.

Les processus peuvent être organisés de manière linéaire ou combinés dans des réseaux plus complexes qui se comportent collectivement comme des opérations par lots ou continues. Ils peuvent être limités par des dépendances et la disponibilité des intrants, et nécessiter diverses ressources et exécutants tels que des appareils et des personnes. Ils peuvent avoir des durées minimales nécessaires, des taux de changement, des sous-processus et des effets secondaires.

L'un des objectifs d'une ontologie de processus générale est de couvrir des niches d'ontologie de processus plus spécifiques - de la traçabilité des aliments, qui nécessite une granularité grossière du processus - récolte, stockage, transport, division, combinaison - à une modélisation plus spécifique, telle que la manière d'élaborer une recette. Nous cherchons à répondre à la question suivante : les mêmes relations génériques du modèle de processus peuvent-elles couvrir la modélisation de niches spécifiques, de sorte que les relations dans ces niches se révèlent équivalentes aux relations génériques ?

2.3 Objectifs du processus

L'objectif d'un processus peut être exprimé simplement en se référant à une entité de sortie, un produit désiré. Divers procédés peuvent être connus pour obtenir un tel résultat, mais le choix de l'un d'entre eux peut être limité par les dispositifs ou les intrants disponibles, ou par des contraintes de temps, ou par un manque d'information (par exemple, en l'absence de protocole précis). Une modélisation plus générique est obtenue en décrivant ce qu'est l'entité de sortie de telle sorte que le processus puisse être reconnu comme un moyen d'atteindre cet objectif.

Les objectifs du processus peuvent également être exprimés en faisant référence aux capacités fonctionnelles des appareils. Dans BFO, une fonction est présente dans un dis-

positif ou un matériel et un processus "réalise" ou effectue la transformation qu'une fonction caractérise. Un protocole ou plan contenant des instructions détaillées pas-à-pas peut être impliqué—il "exécute" le processus.

2.3.1 Variations de la modularité des processus

Il existe quelques types d'approches pour modéliser des processus et les capacités qu'un modèle basé sur une ontologie peut idéalement satisfaire, allant de modèles holistiques permettant d'appréhender à la fois le comportement autonome d'agents libres de se déplacer dans un environnement, recherchant ou attendant activement des informations contextuelles auxquelles ils sont censés réagir, que des modèles limités à la description des étapes d'un flux d'opérations, implicitement contrôlés par une couche plus abstraite. Les modèles de processus orientés vers l'autonomie peuvent exprimer les conditions dans lesquelles un processus est activé - un ensemble de critères concernant le contexte environnemental requis, les matériaux d'entrée, l'énergie, les contraintes de temps, le(s) dispositif(s) et les opérateurs. Les modèles de processus orientés vers le contrôle peuvent comporter une couche parallèle de processus de contrôle qui fournissent des signaux d'entrée tels que "démarrage", "pause" et "arrêt", parallèlement aux conditions d'activation orientées vers l'agent. Par exemple, la Data Documentation Initiative (DDI) [6] et sa variante récente DDI Cross-Domain Integration (DDI CDI) [2] fournissent un modèle de processus qui présente cette approche basée sur la spécification de provenance PROV du W3C, dont PROV-O fait partie [5].

Les deux types de modèles peuvent être connectés de manière « plug and play » pour créer des dépendances de processus et une transformation globale du matériau/phénotype d'un état initial à un état final. Les moteurs de règles mettent en œuvre l'approche autonome en surveillant l'état de l'environnement. Un processus autonome commence à ressembler à un processus orienté vers le contrôle quand son environnement est réduit à un ensemble étroit de stimuli/entrées (les entrées qui provoquent des réactions deviennent plus évidentes).

Les spécifications du flux d'opérations et leurs cadres informatiques permettant la définition d'un réseau de processus constituent l'ossature du traitement des données et de la reproductibilité expérimentale. La configuration du flux d'opérations est souvent un processus de configuration manuelle, mais de nombreux systèmes tels que CyVerse [19] et Galaxy [7] fournissent des interfaces utilisateur pour le développement du flux d'opérations qui indiquent les entrées requises et les options contextuelles. La modélisation des processus guidée par l'ontologie devrait pouvoir répondre à cette approche plus orientée vers le contrôle au moyen de classes de "contrôle des processus" dédiées à l'ajout d'un contrôle informatif des processus.

2.3.2 Modèles centrés sur le processus et modèles centrés sur l'objet

Il est possible d'établir une distinction de représentation selon que les données sont modélisées du point de vue du processus ou du point de vue de l'objet, ou des deux. Une

perspective processus-et-objet permet de garder le chemin de provenance - l'information sur la façon dont une entité ou son contexte a été modifié dans l'environnement ou dans une chaîne d'approvisionnement. Les processus de transformation qu'il s'agisse d'analyses, de méthodes de traitement des données ou de méthodes de production agricole, de stockage, d'expédition ou de fabrication, relie les états passés, présents et futurs d'une entité, qui peuvent tous être capturés dans un graphe de connaissances. Cela correspond bien aux registres de traçabilité des produits qui documentent la façon dont les produits sont créés ou transformés au cours de leur cycle de vie par divers agents.

Une perspective objet des entités et de leurs caractéristiques à un moment donné peut couvrir le contexte qui / quoi / quand / pourquoi / où d'une situation mais manque d'un cadre pour décrire le "comment".

Une perspective objet est dérivée d'un modèle processus-et-objet en remplaçant tout processus donné par un lien généralisé et donc moins informatif "dérive de" ou "concerne" (entre l'information et une entité) pour relier les entrées et les sorties (i/o). De même, un modèle de processus qui détermine les entités d'entrée et de sortie peut être réduit à un lien de dépendance entre les processus dans une "perspective de processus uniquement" en remplaçant les entités i/o par une relation RO "fournit directement des entrées pour". On perd alors tous les détails concernant les caractéristiques qui auraient pu être associées aux objets auxquels les processus ont été appliqués.

2.3.3 Étapes de processus, parties, dépendances et abstraction

Divers modèles de processus ont une entité "étape", qui, du point de vue du contrôle du processus, est une convention permettant de nommer une vue de processus abstraite et de la classer dans un flux d'opération. Une étape porte la sémantique d'un processus qui peut être isolé comme un événement contrôlable et qui échoue ou réussit, et qui peut avoir des états de pause ou redémarrage. Une étape capture un certain niveau de modularité du processus et de granularité de contrôle. Elle doit être complétée avant de passer à une étape ultérieure qui en est dépendante. Un modèle de processus élégant serait en mesure d'offrir différents niveaux de granularité. Les connexions entre les entités matérielles, les caractéristiques, les processus et les informations à un instant donné sont souvent prises ensemble pour décrire les « états » possibles d'un « système ». Alors que certains modèles n'expriment que des actions à effectuer sur des entités (comme "ébullition"), l'expression des qualités d'entrée et de sortie permet d'exprimer les changements d'états d'un système et les conditions qui doivent être remplies, comme "attendre jusqu'à ce que l'eau soit en ébullition ». Certains modèles permettent aux processus eux-mêmes d'avoir ou d'influencer des caractéristiques qui peuvent être observées ou contrôlées.

Associer une qualité contrôlée ou surveillée à un processus est un raccourci pour modéliser les qualités qui identifient l'entrée ou la sortie du processus - un niveau de détail requis pour l'automatisation industrielle. Les concep-

tions des études expérimentales font une distinction quant à l'entité matérielle ou aux caractéristiques de processus qui sont traitées comme des contrôles (ou paramètres), des variables indépendantes ou des variables dépendantes. Comme condition préalable au succès, le contrôle réel d'une expérience ou les niveaux de variables indépendantes doivent correspondre à la conception ou au plan expérimental, de sorte qu'une étape de contrôle de la qualité implique l'étalonnage précis (mesure) des niveaux de réglage du dispositif de contrôle.

2.3.4 Limites du modèle de processus OWL

L'un des problèmes que pose la logique OWL par rapport au sens commun est que, techniquement, il n'y a pas de moyen facile d'exprimer qu'un processus a directement modifié la qualité d'un matériau d'entrée (ou d'une entité numérique), étant donné qu'il n'est pas possible d'affirmer qu'une instance de matériau de sortie est identique au matériau d'entrée (par exemple, un processus de blanchiment transforme les cheveux foncés d'une personne en cheveux blonds, de sorte que ni les cheveux ni la personne ne peuvent techniquement être considérés comme la "même" entité). Une approche consiste à raisonner sur des observations horodatées concernant des objets, plutôt que sur les propriétés d'un objet directement. On peut aussi créer une instance d'une entité qui est utilisée (manuellement ou par logiciel) pour désigner une entité matérielle donnée n'importe où dans une matrice d'entrée/sortie de processus, marquant effectivement un ensemble d'entités d'instance comme étant à peu près la même entité tout en évitant la logique d'équivalence d'OWL. Cela ne facilitera pas les prouesses en matière de classification avec OWL, mais permettra d'assurer la provenance et la traçabilité en fonction des identifiants attachés aux ressources du processus.

Compte tenu des limites des capacités actuelles des raisonneurs OWL, une ontologie de processus devrait principalement être considérée comme le véhicule permettant de fournir la grammaire des catégories d'entités et des relations qui sont utilisées pour créer des phrases décrivant les processus. Les raisonneurs OWL ne peuvent que déduire l'appartenance à une classe, de sorte que l'ordre des étapes du processus ne peut probablement pas être déduit des dépendances du processus, et des algorithmes extérieurs à la logique OWL sont nécessaires. Certains raisonneurs OWL peuvent traiter des comparaisons simples de propriétés de données numériques ($x \geq y$), mais l'automatisation dans les laboratoires et les chaînes de montage nécessite de nombreuses opérations de non-classification, telles que l'optimisation des matériaux en lots, le recalibrage des appareils et le calcul des coûts des procédures. Les technologies telles que SWRL [58] ou SPARQL [52] sont adaptées aux calculs de flux d'opérations des graphes de connaissances pilotés par l'ontologie. En créant une séparation entre la validité structurelle et le calcul du flux d'opération, OBO Foundry (et d'autres communautés) peut alors se concentrer sur la mission d'interopérabilité consistant à aligner diverses ontologies de domaine OWL afin d'utiliser un vocabulaire relationnel partagé capable d'exprimer et de valider des struc-

tures de données aux niveaux des classes et des instances.

3 Modèle de recette FoodOn

FoodOn intégrera un modèle de processus général lié à l'alimentation afin que les méthodes de transformation des aliments puissent être détaillées dans une série de micro-modèles spécifiques à chaque type d'aliment ou de processus allant de l'acte apparemment simple de cuisiner ou de faire bouillir, à suivre un processus de recette consommateur ou industriel. En raison des limites du raisonnement, l'accent est davantage mis ici sur la recherche de valeur dans des structures de données conformes aux distinctions ontologiques sur le monde, plutôt que sur la réalisation de prouesses de raisonnement sur les données elles-mêmes.

Les patrons de conception de la figure 1 sont une proposition, à discuter, mais une partie est déjà ajoutée à FoodOn.

Actuellement, FoodOn dispose d'un vocabulaire étendu pour décrire la composition d'ingrédients à source unique - la taxonomie et l'anatomie des ingrédients de l'organisme. Pour décrire la composition des aliments à plusieurs composants, FoodOn permet aux classes et instances de produits alimentaires de se référer aux ingrédients (eux-mêmes des produits alimentaires, y compris les additifs) au moyen des relations «has ingredient» et «has defining ingredient».

Ces relations peuvent suffire pour décrire une liste de courses, mais les ingrédients de la recette sont généralement associés à des étapes et doivent apparaître dans l'étiquetage du produit : il faut une liste d'ingrédients, ordonnée en fonction des quantités, données en proportion ou valeur absolue. Souvent, une recette a également besoin de dispositif - même une table ou une cuillère doivent être considérées comme une ressource nécessaire. En même temps, la réutilisation du vocabulaire à différentes échelles de modélisation - de la cuisine au laboratoire en passant par la robotique - tombe, espérons-le, dans des modèles de réutilisation, de sorte que moins de relations sont nécessaires à mesure qu'un modèle de processus générique se développe, l'accent étant davantage mis sur la définition de nouvelles sous-classes de processus, entité, dispositif et information. Le modèle de recette FoodOn proposé dans la figure 1 comporte un certain nombre de nouveaux composants (représentés par des contours en pointillés), notamment un « ensemble d'ingrédients », un « ensemble d'appareils » et un « ensemble d'instructions » de spécifications d'étape qui permettent respectivement de se référer aux matériaux alimentaires, dispositifs et les processus planifiés qu'ils alimentent. Ces ensembles seront des sous-classes de l'ensemble de données IAO (IAO :0000100) qui est un ensemble de choses du même genre. Une certaine quantité de récursivité du jeu d'instructions est nécessaire lorsque les recettes ont des étapes ou des étapes globales, chaque étape ayant elle-même potentiellement un composant de jeu d'instructions. Chaque spécification d'ingrédient fait référence à une matière alimentaire (produit, produit intermédiaire, ou additif, ou au choix) mais aussi à une mesure proportionnelle, de comptage ou scalaire. Les capacités du dispositif sont indiquées par ses spécifications (taille de la

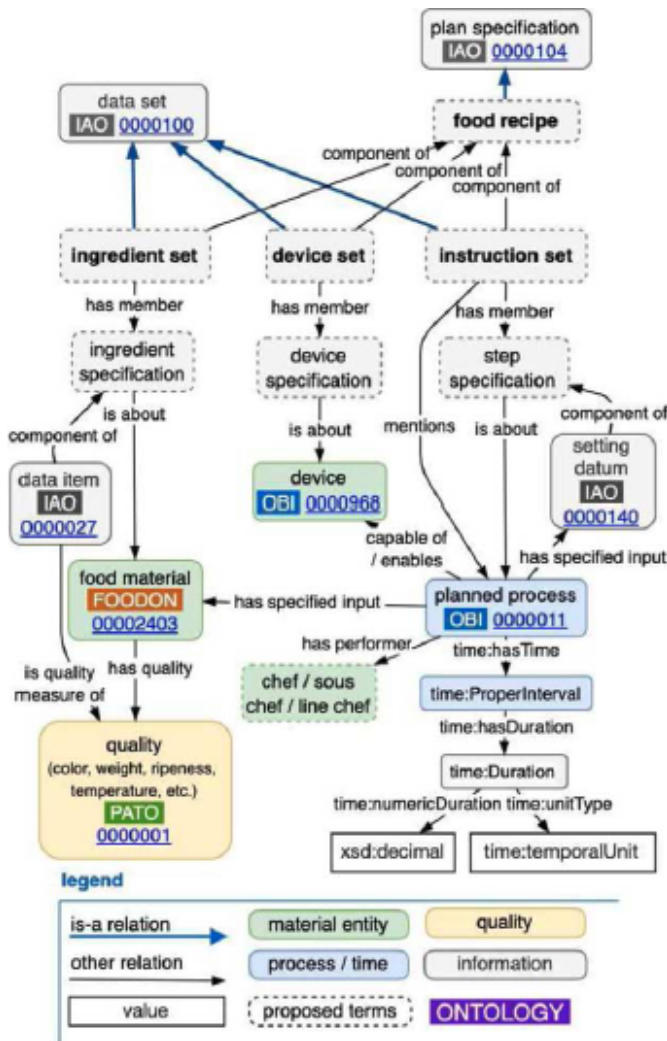


FIGURE 1 – Prototype d'une recette dans FoodOn.

casserole, de la poêle ou du bol, la température nominale du four à vaisselle, etc.). Un commentaire ou une instruction en texte libre peut potentiellement être lié à l'une des entités de la classe en utilisant des annotations telles que "dc :description" ou "rdfs :comment". Le modèle actuel de recette n'inclut pas les procédures de service ou les portions, ni les informations dérivées sur la nutrition ou la durée de cuisson.

4 Besoins ontologiques pour la transformation des aliments

Le cas d'usage, la recette de purée de carottes, décrit amplement dans [11] donne lieu à plusieurs recommandations d'entités et de propriétés à ajouter dans FoodOn ou OBO :

- l'entité **Ingredient set** (Ensemble d'ingrédients) qui est un data set (ensemble de données) contenant les spécifications des ingrédients ;
- l'entité **device set** (Ensemble de dispositifs) qui est un data set (ensemble de données) contenant une ou plusieurs spécifications des dispositifs ;
- l'entité **instruction set** (Ensemble d'instructions)

qui est un data set (ensemble de données) contenant les spécifications d'une étape ;

- **Observation** : Une donnée qui est le résultat d'un processus d'observation et qui a des mesures comme composants ;
- l'entité **food recipe** (recette alimentaire) qui est un plan specification (document de spécification) et qui peut avoir des composants qui sont des ensembles d'ingrédients, des ensembles de dispositifs et des ensembles d'instructions ;
- l'entité **ingredient specification** qui spécifie un food material et sa quantité ou son taux ;
- l'entité **device specification** qui spécifie un dispositif et ses réglages ;
- l'entité **step specification** qui spécifie un processus planifié et les restrictions imposées à ses participants.
- la propriété **has ratio** pour représenter une fraction ou un pourcentage ;
- Idéalement, l'élément "has specified output" de OBI devrait être placé sous l'élément "has output" de RO et, de la même manière, l'élément "has specified input" devrait être placé sous l'élément "has input".

Les ustensiles et équipements de cuisine ainsi qu'une longue liste de processus alimentaires seront également nécessaires, mais ils pourront être ajoutés progressivement à FoodOn ou à l'ontologie EO (Environnement Ontology) [?] (qui contient les produits manufacturés) ou à une autre ontologie, selon les besoins. Ce modèle peut être étendu dans un certain nombre de directions - en le reliant à des informations sur la nutrition, le risque d'allergie et les substitutions d'ingrédients et l'évaluation de l'empreinte écologique. Les pratiques et représentations culturelles pourraient être incluses et les réglages des appareils pourraient être spécifiés. Une piste de recherche consiste à déterminer la logique appropriée ou les informations nécessaires pour déduire les étapes à partir d'un graphe de dépendance d'un modèle de recette. De même, l'abstraction des recettes pourrait être possible grâce à une transformation qui conduirait aux processus de niveau supérieur et à tous les ingrédients sous-jacents. La validation du modèle pourrait être réalisée en le testant sur les projets qui ont des bases de données de recettes.

5 Conclusion

La modélisation des processus permet d'obtenir le liant nécessaire pour expliquer l'histoire des choses matérielles et des informations dérivées. La transformation des aliments - depuis le point de récolte agricole et au-delà - est un sujet de préoccupation immédiate à l'heure de l'adaptation au climat, de la mondialisation de la transformation et de la distribution des aliments, des déserts alimentaires et des risques sanitaires liés à l'alimentation - obésité, malnutrition, contaminations et agents pathogènes d'origine alimentaire. Une ontologie de processus générique cohérente, avec des composants spécialisés, permettra plus d'interopérabilité entre la recherche en science alimentaire et les systèmes

opérant au sein de la chaîne agro-alimentaire, pour une transition vers un modèle plus sain et plus durable. Notre analyse des lacunes de la capacité de modélisation des processus basé sur le modèle OBO actuel, fournit une feuille de route pour l'ajout de nouveaux termes essentiels pour la construction de micromodèles permettant de représenter des recettes, ainsi que pour la recherche et l'automatisation dans le domaine de l'alimentation. Les principaux axes de recherche restent l'étude des qualités organoleptiques et physiques des aliments (kinesthésie, rhéologie et gastronomie moléculaire, par exemple), afin de pouvoir décrire, classer et prédire leur transformation à la suite de processus naturels et planifiés. Nous espérons également que notre travail sur les modèles de processus OBO sera pertinent au-delà du domaine alimentaire. Les commentaires sur les exigences en matière de modélisation de la transformation des aliments, sur la sémantique plus large des modèles de processus et la participation à cet effort sont les bienvenus.

Références

- [1] Github.
- [2] Cross-domain integration.
- [3] Obo foundry principles : Overview.
- [4] Submissions :food recipe ingredient substitution ontology design pattern - odp.
- [5] Using the process pattern ? ddi 4.0 dev documentation.
- [6] Welcome to the data documentation initiative.
- [7] E. Afgan, D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. Cech, J. Chilton, D. Clements, N. Coarar, B. A. Grüning, A. Guerler, J. Hillman-Jackson, S. Hiltmann, V. Jalili, H. Rasche, N. Soranzo, J. Goecks, J. Taylor, A. Nekrutenko, and D. Blankenberg. The galaxy platform for accessible, reproducible and collaborative biomedical analyses : 2018 update. *Nucleic acids research*, 46(W1) :W537?W544, Jul 2018.
- [8] R. Arp, B. Smith, and A. D. Spear. *Building Ontologies with Basic Formal Ontology*. MIT Press, Jul 2015.
- [9] A. Bandrowski, R. Brinkman, M. Brochhausen, M. H. Brush, B. Bug, M. C. Chibucos, K. Clancy, M. Courtot, D. Derom, M. Dumontier, L. Fan, J. Fostel, G. Frago, F. Gibson, A. Gonzalez-Beltran, M. A. Haendel, Y. He, M. Heiskanen, T. Hernandez-Boussard, M. Jensen, Y. Lin, A. L. Lister, P. Lord, J. Malone, E. Manduchi, M. McGee, N. Morrison, J. A. Overton, H. Parkinson, B. Peters, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. H. Scheuermann, D. Schober, B. Smith, L. N. Soldatova, J. Stoeckert, Christian J., C. F. Taylor, C. Torniai, J. A. Turner, R. Vita, P. L. Whetzel, and J. Zheng. The ontology for biomedical investigations. *PLoS one*, 11(4) :e0154556, Apr 2016.
- [10] S. Dervaux, J. Dibie, L. Ibanescu, and J. Raad. Po2 process and observation ontology, 2021.
- [11] D. Dooley, M. Weber, L. Ibanescu, M. Lange, L. Chan, L. Soldatova, C. Yang, R. Warren, C. Shimizu, H. K. McGinty, and W. Hsiao. Food process ontology requirements. *Semantic Web, Preprint(Preprint) :1–32*, 2022. Publisher : IOS Press.
- [12] D. M. Dooley, E. J. Griffiths, G. S. Gosal, P. L. Buttigieg, R. Hoehndorf, M. C. Lange, L. M. Schriml, F. S. L. Brinkman, and W. W. L. Hsiao. Foodon : a harmonized food ontology to increase global food traceability, quality control and data integration. *NPJ science of food*, 2 :23, Dec 2018.
- [13] L. Elizabeth, P. Machado, M. Zinöcker, P. Baker, and M. Lawrence. Ultra-processed foods and health outcomes : A narrative review. *Nutrients*, 12(7), 2020.
- [14] Y. He, Z. Xiang, J. Zheng, Y. Lin, J. A. Overton, and E. Ong. The extensible ontology development (xod) principles and tool implementation to support ontology interoperability. *Journal of biomedical semantics*, 9(1) :3, Jan 2018.
- [15] R. C. Jackson, N. Matentzoglou, J. A. Overton, R. Vita, J. P. Balhoff, P. L. Buttigieg, S. Carbon, M. Courtot, A. D. Diehl, D. Dooley, W. Duncan, N. L. Harris, M. A. Haendel, S. E. Lewis, D. A. Natale, D. Osumi-Sutherland, A. Ruttenberg, L. M. Schriml, B. Smith, C. J. Stoeckert, N. A. Vasilevsky, R. L. Walls, J. Zheng, C. J. Mungall, and B. Peters. Obo foundry in 2021 : Operationalizing open data principles to evaluate ontologies. Jun 2021.
- [16] K. Janowicz, A. Haller, S. J. Cox, D. Le Phuoc, and M. Lefrançois. Sosa : A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics*, 56 :1–10, 2019.
- [17] D. Knorr and M. Augustin. Food processing needs, advantages and misconceptions. *Trends in Food Science Technology*, 108 :103–110, 2021.
- [18] S. Little and C. Cox. Extensions to the owl-time ontology - entity relations, Jul 2020.
- [19] N. Merchant, E. Lyons, S. Goff, M. Vaughn, D. Ware, D. Micklos, and P. Antin. The iplant collaborative : Cyberinfrastructure for enabling data to discovery for the life sciences. *PLoS biology*, 14(1) :e1002342, Jan 2016.
- [20] P. Missier, K. Belhajjame, and J. Cheney. The w3c prov family of specifications for modelling provenance metadata. In *Proceedings of the 16th International Conference on Extending Database Technology*, EDBT 13, pages 773–776, New York, NY, USA, Mar 2013. Association for Computing Machinery.
- [21] OriginTrail. Gs1 digital link : A gateway towards trillions of digital twins?, Aug 2020.
- [22] O. X. Service. Core ontology for biology and biomedicine < ontology lookup service < embl-ebi.
- [23] L. N. Soldatova, D. Nadis, R. D. King, P. S. Basu, E. Haddi, V. Baumlé, N. J. Saunders, W. Marwan, and B. B. Rudkin. Exact2 : the semantics of biomedical

protocols. *BMC bioinformatics*, 15 Suppl 14 :S5, Nov 2014.

- [24] A. Ławrynowicz, A. Wróblewska, W. T. Adrian, B. Kulczyński, and A. Gramza-Michałowska. Food recipe ingredient substitution ontology design pattern. *Sensors (Basel, Switzerland)*, 22(3) :1095, Jan 2022.