FATO: The Food Allergen Traceability Ontology

Semantic Web
XX(X):1–12

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

George Baryannis[1], Lili Jia[2] and Emmanuel Papadakis[1]

Abstract

With the increase of food allergic population worldwide, food allergen traceability has become an imperative food safety concern. Food businesses, however, have difficulty ensuring food allergen traceability because it is time-consuming and costly to obtain accurate food allergen data along the supply chain. Semantic Web technologies have great potential to improve efficiency and accuracy of food allergen traceability through automating food data exchange along the supply chain. In this paper, we present the Food Allergen Traceability Ontology (FATO), the first ontology that focuses on food allergen management and traceability processes. To overcome the overspecification problem in the development of ontologies, we propose the integration of a range of knowledge sources on improving food allergen management, in addition to domain experts, to inform the development of FATO. The ontology builds on and is compatible with existing food and product ontologies and models and captures knowledge on food allergen declarations, food allergen management processes and traceability. Application examples are provided to illustrate how FATO can be employed to address long-standing issues in food allergen management as well as drive innovation in food businesses.

Keywords

food allergens, food traceability, modular ontology, EPCIS

Introduction

Food allergen traceability, which involves tracking food allergens throughout food production and management processes along supply chains, has become a high global concern due to its important role in food safety (Jia and Evans 2021a; Lack 2008). A recent study has shown that undeclared food allergens accounted for 46% of total food safety incidents and recalls from 2008 to 2018, becoming the top food hazard in the world (Jia and Evans 2021a; Soon et al. 2020). This represents a great public health threat to people with food allergies as there is no cure for food allergies to date and the only way to prevent allergic reactions is to avoid triggering allergens (Jia and Evans 2021a; Begen et al. 2018). According to statistics of the World Allergy Organization (http://worldallergy.org/), about 2.5% of world population has food allergies. In the UK, about 4 people in every 100,000 per year went to hospitals due to food allergies in 2018 (Conrado et al. 2021). Among them, hospital admissions of children under 15 years have jumped from 2.1 to 9.2 per 100,000 population per year from 1998 to 2018.

The key challenge of food allergen traceability is that it is time-consuming and costly for food businesses manufacturers, suppliers and retailers to track food allergens along the supply chain (Jia and Evans 2021a). Food businesses have to collect and validate their food product information from suppliers manually via spreadsheets and emails. This manual process is labour-intensive and bears a high risk of human errors. For example, Jia and Evans (2021a) reported that errors in label check, ingredient data update and package check processes are the top three causes of food allergy recalls, accounting for 48.4% of food allergy recalls. It is hence imperative to help food businesses reduce the burdens and risks of food allergen declaration by improving food allergen traceability.

Semantic Web technologies have great potential to reduce time and cost of food traceability through automating food data exchange between different computer systems and improve accuracy of food allergen data through supporting food data verification (Jia and Evans 2021a; Pearson et al. 2021). Existing work on food traceability ontologies, such as FoodOn (Dooley et al. 2018), Food Track & Trace Ontology (FTTO) (Pizzuti et al. 2014) and Electronic Product Code Information Services (EPCIS) (GS1 2022), does not consider food allergen management. Food allergen management software, such as Vital (Taylor et al. 2019) used in Australia and New Zealand, considers food allergen management, but is unable to support automated data exchange. In this work, we address this gap by introducing the Food Allergen Traceability Ontology (FATO). FATO is an application ontology that builds on concepts from FoodOn, EPCIS, and the Global Product Classification related to food, businesses and products, respectively, with further expansion on tracking and tracing food allergens.

[1]Department of Computer Science, School of Computing and Engineering, University of Huddersfield, UK
[2]Department of Engineering, University of Cambridge, UK

Corresponding author:
George Baryannis, Department of Computer Science, School of Computing and Engineering, University of Huddersfield, Huddersfield, West Yorkshire, HD1 3DH, UK.
Email: g.bargiannis@hud.ac.uk

FATO is intended to provide a standardised representation of food allergen management concepts and relationships to facilitate and enhance food traceability approaches and technologies. The realisation of this potential across businesses in the food and drink sector, however, depends on the scope of the developed ontology. Commonly, specification of ontologies relies on the knowledge of a group of domain experts. In some cases, these experts may not be enough to capture all necessary requirements for ontology development, potentially leading to an overspecification problem (Shimizu et al. 2022). As the challenge of food allergen traceability is prevalent across food businesses, solely relying on particular domain experts is not sufficient to meet sector-wide needs of food allergen traceability.

To tackle this challenge, we provide a systemic framework which allows to develop a scalable ontology, for the entire food and drink sector rather than a few food companies, by integrating both scientific studies and domain experts in food allergen management into FATO. This systemic approach to FATO can be used in developing ontologies for other domains that have overspecification issues through certain adaptations to align with relevant scientific disciplines. Version 1.0 of FATO is available at `https://w3id.org/FATO` as an OWL file following RDF/XML syntax and a request has been submitted for its inclusion in the OBO Foundry repository. Human readable documentation was created with the assistance of OnToology (Alobaid et al. 2019).

The remainder of this paper is organised as follows. Section provides an overview of existing ontologies related to food traceability. Section details the modelling and development process followed. FATO is presented in Section and is evaluated in Section . Section explains how the ontology can be integrated within existing food allergen management processes and Section concludes by summarising prospective applications of FATO and future steps in further developing the ontology.

Related Work

In this section, we provide a focused summary of existing ontologies that are directly related to food allergen management and traceability.

Valarakos et al. (2006) represents an early effort to model allergen knowledge semantically through the development of an allergen ontology grounded in the WHO/IUIS Allergen Nomenclature (Pomponi et al. 2020). While this work established the value of semantic modelling for allergens, FATO expands upon this foundation by integrating allergen management directly into the broader supply chain traceability standards (such as EPCIS) that have emerged since.

Arguably the most substantial effort in representing knowledge about food is the FoodOn ontology (Dooley et al. 2018) (`https://foodon.org/`), developed by a consortium of universities, research institutes, health agencies and non-profit organisations. The main aim is to build a farm-to-fork ontology that is capable of comprehensively describing information about food from around the world in an easily accessible manner. FoodOn is built on top of the Basic Formal Ontology (BFO, `https://basic-formal-ontology.org/`) and includes elements from the Environment Ontology (ENVO, `https://environmentontology.org/`). The end result is a quite extensive ontology that covers knowledge ranging from food categories and products to animal and plant food sources and food transformation and preservation processes. With regard to food allergens, FoodOn considers only information for food allergen labelling, while the management of food allergens in production processes is not included.

In the domain of food traceability, another attempt is the Food Track & Trace Ontology (FTTO) (Pizzuti et al. 2014). FTTO aims to integrate all information relevant to traceability in the food supply chain and comprises four main modules that model food products, service products (e.g. for packaging and treatment), processes and stakeholders along the food supply chain. The main purpose for developing FTTO is to provide terms and concepts to a wider global food traceability framework that involves modelling the complete food supply chain as a business process. FTTO contains several modules related to food allergen traceability, such as packaging, labelling and quality control. However, it does not contain the details related to food allergen management, such as cross-contact and packaging checks. Also, the ontology is not publicly accessible in any format, which prevents any reuse and extensions.

A more recent attempt to advance food traceability technology is the ontology for the EPCIS standard (GS1 2022). EPCIS enables information sharing about the physical movement of products through the supply chain by providing standardised definitions for terms and processes. EPCIS is an event-based model, recording events of individual products, products that are a result of aggregation or transformation and business transactions. While EPCIS was first encoded in XML, EPCIS v2.0 is also available as an OWL ontology (GS1 2022). Although EPCIS is not specifically focused on food allergen traceability, it is important to consider it for compatibility purposes to increase likelihood of adoption of any new ontology such as FATO.

Methodology

Ontology engineering methodologies need to involve stakeholders related to all aspects of the ontology life cycle (Kotis et al. 2020). In practice, there is no defined boundary regarding the stakeholders who should be involved. A common practice is that ontology engineers collaborate with domain experts and users to develop the ontology specification (Kotis et al. 2020; Vrandečić et al. 2005). Nevertheless, such a practice may be affected by overspecification problems (Shimizu et al. 2022) in cases where the available domain experts and users are not sufficient to capture all knowledge required for the ontology to be generally applicable to more users. In developing FATO, we sought to tackle this challenge by collecting domain knowledge not only from experts, but also from scientific literature into ontology modelling (Karpatne et al. 2022). Our framework is illustrated in Figure 1. At its core, our approach integrates the Modular Ontology Modelling (MOMo) methodology (Shimizu et al. 2022), with two additions: (1) the aforementioned broadened knowledge gathering approach going beyond domain experts; and (2) a post-deployment feedback loop to drive future versions of the ontology. In this section, we describe in detail all steps in Figure 1, explaining how we applied MOMo for the design and development of FATO.

Step 0: Forming the modelling team The modelling team was structured according to the guidelines of Shimizu et al. (2022), ensuring that the particularities of the food allergen management domain are considered. It comprises domain experts (including both food safety regulators and food companies), food allergen management researchers, ontology engineers and data scientists. A consortium was also set up, including food safety regulators and food companies, to act as a sounding board throughout the development of FATO. Monthly meetings with food safety regulators (10-20 participants/meeting) were held and in-depth interviews took place with two food companies that were selected as testbeds. Food allergen management researchers, led by the second author, developed a standardised Food Product Information Form (FPIF) by harmonising the FPIFs in the UK, Australia and New Zealand while considering root cause analyses of food allergy recalls and suggestions of domain experts in the UK. Ontology engineers, led by the first author, created FATO based on the standardised FPIF.

Step 1a: Data sources Collection of relevant data began with the extraction of scientific knowledge on improving food allergen management for food manufacturing through a comprehensive review of root causes from literature and historical food recalls, as well as meetings and interviews with domain experts, food safety regulators and food businesses (Jia and Evans 2021a,b). This included 24 common errors identified through a scoping review of 198 journal articles between 2010 and 2020, 453 food allergy alerts between 2016 and 2019 in the UK, and 60 root cause analyses during 2018 and 2019 in England (Jia and Evans 2021b).

As mentioned earlier, we collected and analysed various FPIFs, including the PIF 6.0 (2018) in Australia and New Zealand developed by the Australian Food and Grocery Council, and 24 FPIFs from diverse food businesses in the UK. In the process of harmonising the different FPIFs, we considered only the information in PIF 6.0 when it is suitable for and consistent with the FPIFs in the UK to optimise the shared knowledge in the consideration of food regulatory differences between the UK and Australia and New Zealand.

This collective knowledge then led to a standardised FPIF that can capture a broad range of food products and processes, regardless of the country of origin. The standardised FPIF acts as the main source of data and knowledge for initiating a modular ontology modelling process. Additional data sources include the ontologies summarised in Section .

Step 1b: Use cases The overall motivation for the development of FATO is to help food businesses improve food allergen management. This motivation is primarily linked to three main needs and associated use cases. These were identified by tapping onto all three sources of scientific knowledge identified in Fig. 1: review of food allergen management literature, interviews with food safety regulators and food businesses as domain experts and an analysis of historical food allergy recall incidents.

The first use case is associated with the need to facilitate capturing essential food product information required by food safety regulations, such as manufacturers, ingredients, food allergens and storage, as well as information related to food allergen management. Capturing this information is a legal requirement, however there is no standardised way of representing this information across the sector. The goal here is to design an ontology that is capable of capturing all such essential food product information that allows a food business to adhere to food safety regulations.

A second use case involves the identification of common operational errors in food allergen management which has the potential of leading to a reduction in food allergy recalls. These common operational errors can be considered in the collection of food product information to assure accurate and timely declaration of food allergens. The designed ontology should then allow both capturing information about food allergen management operations but also pinpoint areas that are commonly the source for errors that can lead to food allergy alerts and recalls, for instance a product with a label claim that does not match its recipe or allergen declarations.

A third use case relates to facilitating the automation of food data exchange between businesses. It is a common practice that a food manufacturer provides a FPIF to their clients. Such a form contains all the essential information needed for the food product. These FPIFs are different from one food business to another, which prevents food businesses from automating data exchange between each other. In this context, the developed ontology should provide a standardised data schema that can be used for automating food data exchange between businesses along the food supply chain leading to a reduction of time and cost in exchanging and verifying food allergen data.

Step 2: Competency questions The process of gathering competency questions began with the previously mentioned review of literature in food allergen management which identified 24 common operational errors in food allergen management. These operational errors were used as a basis for competency questions that can be used to retrieve information to prevent such errors. For example, a verification failure in ingredient information means that we could retrieve information on which product failed to pass its data verification. The verification failures need to be avoided because they could result in mislabelling of food allergens and food allergy recalls. Below we provide an example competency question from each of the main target user groups for FATO; a full list of competency questions is available in Appendix .

The target users of FATO are food businesses and their competency questions mainly come from three user groups within a food business. The first group is internal auditors who are responsible for auditing the compliance with food safety requirements at a production site. The ontology could support them to improve the efficiency of data management and auditing process. An example competency question for this group is: what are the common operational errors that need to be addressed in food allergen management? The second group is food production operators. FATO can help them avoid common errors in food allergen management

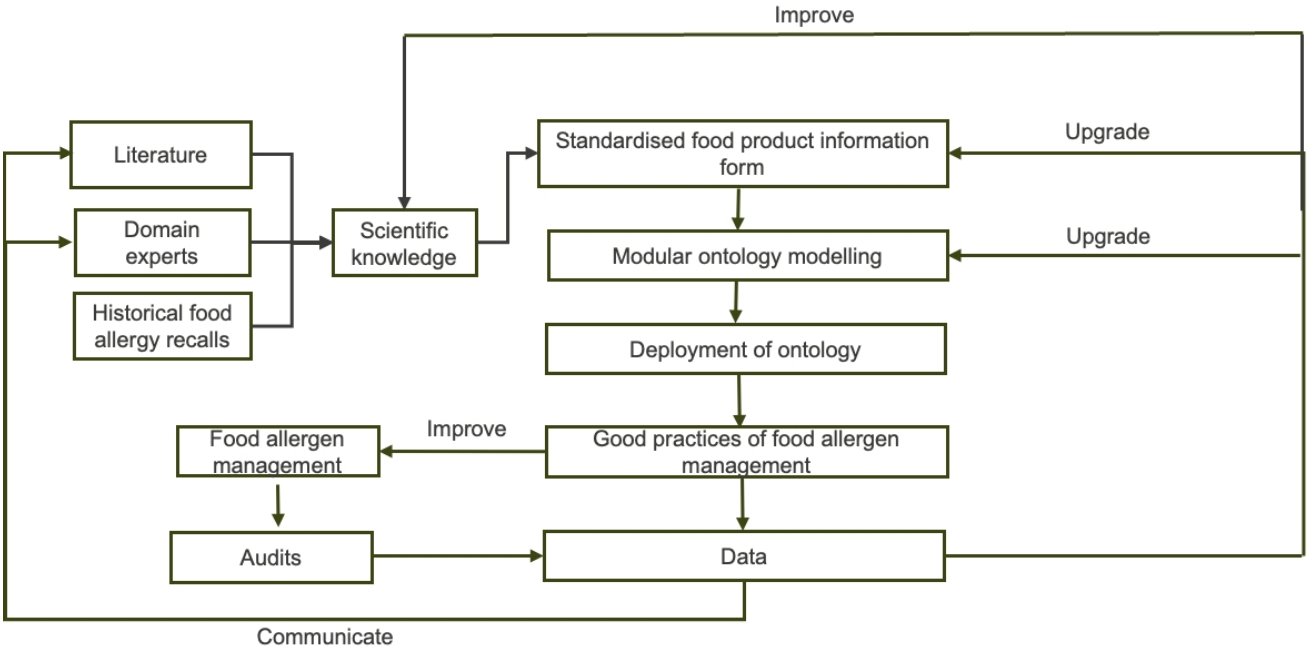

Figure 1. Integrating scientific knowledge for developing and maintaining FATO

and improve food product quality. An example competency question for this user group is: which operations are needed to improve food allergen management? The third group is food product developers. FATO can help them develop food products focusing on consumers with a particular food allergy while considering the business' capacity of food allergen management in the production process. An example competency question in this case is: can a food product, that is not labelled as suitable for a particular consumer group, be identified as such? For instance, a food manufacturer can only safely claim a nuts-free product when they have the capacity to ensure no trace of nuts throughout its production and handling processes.

Step 3: Key notions Through the identified competency questions, key notions were extracted, also relying on the data sources mentioned in step 1, including existing ontologies such as EPCIS and FoodOn. The aforementioned standardised FPIF which contains all the essential information needed for food businesses to exchange food product data with other companies was the primary driver in extracting key notions. By relying on key identifiers included in the standardised form to identify key notions, the likelihood of overspecification is reduced, given that the form is not tied to a particular food company or food product. Additionally, the information and questions captured in the form are already instance free, so generalising into instance free statements (Blomqvist et al. 2016) was only needed in the case of some of the competency questions gathered in step 2. For example, the competency question referring to particular nuts-free products mentioned at the end of Step 2 was generalised to refer to any nut-free product. Based on statements drawn from both competency questions and the standardised FPIF, we identified the following key notions: company, manufacturer, product, ingredient, food allergen management.

Further key notions were elicited based on existing related ontologies, due to the need of maintaining a compatibility with existing standards. This involved conducting an exploration of these ontologies to determine which concepts included in them were relevant to FATO. From EPCIS v2.0, this primarily included concepts around organisation, place, product and transaction. In the case of FoodOn, several key notions around food products were deemed relevant, such as dietary uses and label claims, food allergen labelling, food classification, food product and quality, as well as processes such as packaging, production, storage and transformation. Specific key notions such as product and company were identified through multiple sources, leading to selecting these as the first candidates for modules that would act as nexuses between different data sources.

Step 4: Ontology design patterns The process of selecting ontology design patterns was divided into two parts, depending on whether key notions were derived directly from existing ontologies or from competency questions. In the former case, to ensure compatibility with existing ontologies, we reused patterns or sub-patterns as defined within these ontologies. These include the patterns for EPCIS places and organisations, GPC product classification and FoodOn labelling information. Note that these patterns are essentially combinations and specialisations of more fundamental patterns and we opted for their direct inclusion to avoid duplication of work.

For the remainder of key notions, we explored the patterns included in the MODL library. Patterns that were reused include: the Tree pattern for modelling food classification; the AgentRole pattern for linking products with food businesses; the Provenance pattern to describe relationships between products and ingredients; and the Identifier pattern for the many cases where key notions are identifiers through alphanumeric strings, such as product codes, barcodes or Global Trade Identification Numbers (GTIN). The aforementioned patterns were also combined together to form more complex patterns, such as the combination of the Provenance and Identifier patterns, or the Tree and Identifier patterns, in order to associate products, ingredients and product classes with identifiers.

Step 5: Module and ontology diagrams Using the ontology design patterns identified in the previous step, we developed diagrams for each module associated with one or more key notions. We first dealt with modules developed from existing ontologies, where the focus was to identify the most appropriate and relevant subset to retain, given that the scope of reused ontologies is broader. EPCIS v2.0, for instance, concerns itself with traceability of any product, not just food products, while FoodOn is a farm to fork ontology. In contrast, FATO has a particular focus on the traceability of food allergens in food manufacturing. We then developed diagrams for the patterns listed in Step 4 that comply with patterns in the MODL library. In this case, the focus was to ensure that diagrams conform to the corresponding diagrams in the MODL library of the pattern or combination of patterns they were developed from.

To develop the complete ontology diagram, we first identified links between different modules. This primarily concerned classes from existing ontologies and standards that needed to be connected to newly created classes, such as connecting products to GPC they belong to. The combined schema diagram for FATO was produced following the guidelines for understandability provided by Shimizu et al. (2022).

Step 6: Documentation and review The documentation process partly run in parallel to the previous steps, in terms of drafting axioms for each individual module as the diagrams were being developed. In this step, these axioms were refined and finalised, ensuring they conform to OWL syntax in preparation of the final step of developing the OWL file. Additional axioms were included to formalise the cross-module links identified in the previous step. Documentation was developed for the complete ontology, listing individual modules and their diagrams, associated axioms and explanations and detailing how these modules are combined to create FATO. Finally, we reviewed both the resulting ontology and documentation to ensure completeness and appropriateness with regard to the intended use case, also ensuring naming conventions listed in Shimizu et al. (2022).

Step 7: OWL file In the final step we created an OWL file for FATO using Protégé. While we opted not to automatically generate the OWL file through CoModIDE, we used CoModIDE to generate annotations for modules which were then incorporated into the OWL file. Existing ontologies were included through the direct import mechanism of Protégé. In the case of FoodOn, where only a relatively small subset of the broader ontology was included, we used the ROBOT tool (http://robot.obolibrary.org/) to extract parts of FoodOn as individual OWL files, before importing them. In the case of GPC, which is only provided as XML or JSON (available at gs1.org/standards/gpc), we used a Python script to generate an OWL file from the GPC subset that refers to food, beverage and tobacco product classification. Note that we opted for the GPC-based product classification, rather than an agency-specific classification contained in FoodOn (e.g. FoodEx2 (European Food Safety Authority 2015) or US eCFR (U.S. Food and Drug Administration 2025)) to align with established international trade standards, thereby facilitating broader interoperability across global supply chains.

Step 8: Deployment and iterative improvement Once the ontology is deployed as part of knowledge management systems within food businesses, it can guide the development of good practices on improving food allergen management to generate data. According to Jia and Evans (Jia and Evans 2021a), food allergen traceability involves both human-based learning and machine learning processes. In our context, the former refers to an iterative learning process of food allergen management for experts and operators in food allergen management. The latter concerns an iterative improvement process whereby food allergen management information is audited internally to generate data can then be processed to extract further concepts and relations and deliver new domain knowledge. The new domain knowledge feeds an improvement loop to deliver updated versions of FATO. In this way, a continuous learning process is followed leading to iterative improvement in both food allergen management and FATO, as shown in Figure 1.

The Food Allergen Traceability Ontology

The central concern of FATO is to facilitate the traceability of food products and allergens. This involves a series of data properties to describe key information of the product related to traceability, such as its identifier (e.g. its GTIN - Global Trade Item Number), name, packaging and description. Figure 2 provides a schema diagram for the complete FATO ontology. Due to space limitations, some data properties have been omitted in Figure 2, please consult `https://w3id.org/FATO` for the full schema diagram and ontology. FATO is built on top of the EPCIS v2.0 ontology (available at `https://github.com/gs1/EPCIS`) retaining its class structure for compatibility purposes and extending it to include food traceability and allergen management information. As shown in Figure 2, FATO is organised in five thematic categories, which are summarised next.

Traceability and Food Safety: For traceability purposes, each product may be linked to one or more product batches that are identified separately. These batches can inherit product-related information as provided for the parent product (through a subclass relationship) but can also contain modified information, in case a particular batch differs (e.g. due to variations in the standard production process). Additional information related to tracing products throughout the supply chain is captured with the Traceability class, while the Food Safety class records information related to food safety documents, such as a Hazard Analysis and Critical Control Plan (HACCP). This can range from a standard lot number to any other coding method that needs to be stated explicitly, along with the location of the code on the physical product packaging and the date of the code.

Location: Each product is associated with location-based information provided through object properties linked to two EPCIS classes. Place is used to model information

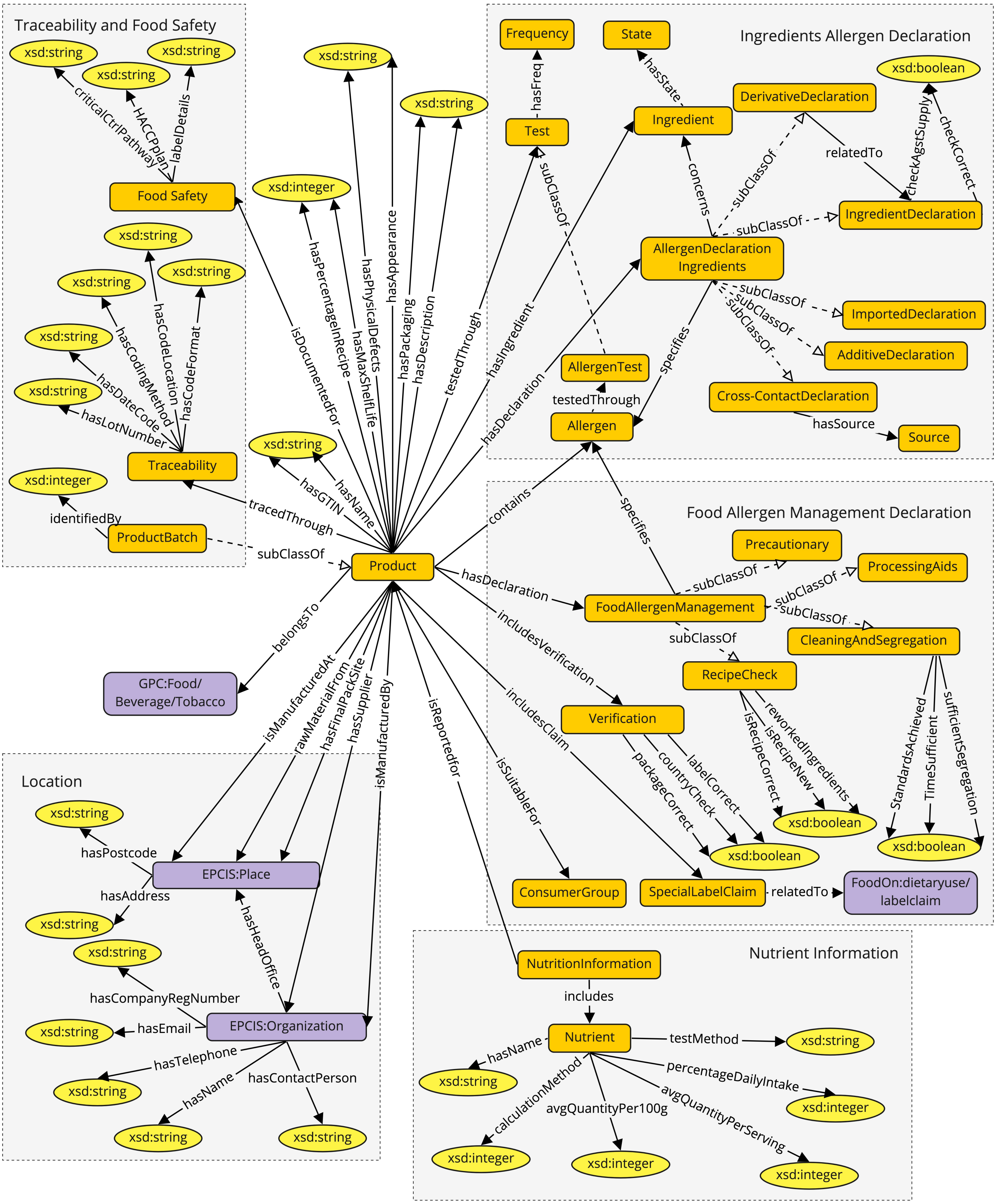

Figure 2. Schema diagram for FATO, with classes from external sources shown in purple.

related to where a product is sourced from, manufactured and packaged, in terms of physical addresses. Organization is used to refer to the business entities that manufacture and supply the product. Data properties linked to the Organization class include information on the name, registration number and contact person for the business entity, as well as contact information such as email and telephone. Organization individuals are also linked to the Place class, to provide location information for their head offices.

Ingredients Allergen Declaration: FATO includes two mechanisms for capturing declarations related to food allergens. The first concerns food allergen declarations at the ingredient level. An Ingredient class is used to capture information specific to an ingredient as part of a product, including its particular state (such as liquid or solid), modelled in the form of a State class. Then, an AllergenDeclarationIngredients class is used as a main focus point for an allergen declaration for a particular product either for the product as a whole, or for one or more of its ingredients. An allergen declaration specifies particular allergens, modelled through the Allergen class, as well as the means through which they are tested, modelled through AllergenTest. Note that AllergenTest is a subclass of a broader Test class which enables FATO to capture information for product tests that may or may not concern allergens. All types of tests require reporting frequency-related data through the Frequency class.

The AllergenDeclarationIngredients class has several subclasses that assist in recording additional information as follows: (1) the IngredientDeclaration subclass involves information on ingredient validation such as checking ingredients against supply information or checking correctness of ingredients against the label; (2) the DerivativeDeclaration subclass is used to declare allergen derivatives for a particular ingredient; (3) the ImportedDeclaration subclass concerns imported ingredients only and allows declaring allergens that have not been specifically addressed in the product label; (4) the AdditiveDeclaration subclass captures food allergens introduced by additive or flavour ingredients; (5) the CrossContactDeclaration subclass allows declaring potential allergens introduced through a cross-contact source (captured through the Source class), such as through cleaning or shared containers, storage or production areas/lines.

Food Allergen Management Declaration and Nutrient Information: The second source for declaring food allergens comes from food allergen management. The FoodAllergenManagement class includes four subclasses that capture food allergen information in food production processes: (1) the Precautionary subclass refers to a warning on product labels that a potential risk of the food allergen may be present (e.g. “may contain” statements); (2) the ProcessingAids subclass models food allergens introduced through processing aids, such as antimicrobial agents or enzymes (Redan 2020); (3) the CleaningAndSegregation subclass includes a self-evaluation of cleaning processes in terms of adequate time and adherence to standards and sufficient segregation of food allergens; (4) the RecipeCheck subclass involves confirming whether a product recipe is new and whether its specification is correct, as well as whether it includes any reworked ingredients (removed and added back at a later stage). Any food allergens introduced by these processes should be recorded.

The Verification subclass captures whether and how verification has been conducted for labels, packaging and distribution, such as checking whether the correct packaging is used in the country of distribution. Allergen claims that are recorded on the product label are captured by the SpecialLabelClaim class, which is associated with the similar dietary use / label claim class in the FoodOn ontology. The ConsumerGroup class can be used to declare that a particular product is suitable for particular consumer groups, such as vegans or infants. Finally, the NutritionInformation class is used to record nutrients that are contained in a product, capturing information that is usually found on product labels, such as the nutrient name and its average quantity per serving and 100g, but also information such as test and calculation methods. Note that while we relied on FoodOn for nutrient-related information, other data sources such as FoodExplorer (EuroFIR AISBL 2025) can also be considered.

Table 1. Evaluation of FATO according to knowledge coverage, inverse popularity and structural metrics

Knowledge coverage	**Value**	**Structural**	**Value**
# of classes	1348	Max depth	9
# of properties	180	Min depth	1
- *Datatype properties*	91	Avg depth	3.95
- *Object properties*	89	Abs depth	5383
# of individuals	1094	Max breadth	50
Inverse popularity	**Value**	Min breadth	1
Ontology direct imports	3	Avg breadth	5.95
# of classes	1325	Abs breadth	1362
# of Datatype properties	27		
# of Object properties	49		

Evaluation

To evaluate FATO in terms of knowledge coverage, popularity and structure we utilised the metrics proposed by Fernández et al. (2009), replacing the direct popularity metric (which measures ontologies importing a given ontology and is equal to 0 for newly proposed ontologies) with the inverse popularity metrics used by Turchet et al. (2022) to measure the number of ontologies, classes and properties imported within FATO. These are reported in Table 1. It should be noted that the large number of imported classes is due to FATO importing the whole subset of food, beverage and tobacco product categories in GPC, which includes 1,046 distinct categories. In terms of terminological coverage, existing ontologies capture only four of the key notions determined by the competency questions (organisation, place, label claim and product category), with the rest introduced by FATO.

To validate FATO and evaluate adherence to good practices, we used the OntOlogy Pitfall Scanner (OOPS!) tool (Poveda-Villalón et al. 2014), available at `https://oops.linkeddata.es/`, which supports semi-automated evaluation of an ontology against structural, functional and usability dimensions, considering 41 pitfalls in total. The online version of OOPS! allows automated checks for 33 of these pitfalls and none of these were identified in FATO, with the exception of a minor pitfall on the URI containing a file extension. We also manually checked the remaining 8 and ensured that they are not found in FATO. Moreover, we used a locally implemented version of the validation tool Themis (Fernández-Izquierdo and García-Castro 2019) and developed tests linked to the competency questions discussed in Section . These tests confirmed that the developed ontology meets requirements related to capturing knowledge about products, allergens and declarations related to food allergen management and traceability.

Finally, we evaluated the potential of FATO to answer queries related to the competency questions discussed in Section . To facilitate this, we leveraged GPT-4o using context injection through prompts, verbally expressing a schema subset. We then developed SPARQL queries linked to the competency questions and confirmed the ability of FATO to provide answers. Appendix includes sample queries and discusses their results.

Applications

FATO was developed to facilitate solutions to improving food allergen management. In this section, we describe how FATO can help address these challenges and illustrate how food businesses can benefit from FATO. Intended applications for FATO, and areas where its adoption could bring significant benefits, include facilitating food data exchange and traceability, improving food allergen management and helping food businesses develop 'free-from' products.

In the UK, businesses often use their own FPIFs to exchange data and each food business may have to develop its own traceability technology along the food supply chain. This is quite costly and leads to low incentives for adoption. Drawing on a standardised FPIF developed in a previous Food Standards Agency project (Jia and Evans 2021b), FATO is applicable to many food businesses across a broad range of food products. For example, a test of the standardised FPIF carried out with a large food procurement company indicated that it can cover about 80% of food products (Jia and Evans 2021b). Our strategy for ensuring adoption in this context leverages the stakeholder network established during the aforementioned project. Specifically, FATO can be embedded in existing digital platforms or use it to develop a new digital platform for food data exchange. This could help food businesses reduce time and lower costs associated with food allergen traceability and facilitate automated data exchange across different businesses. Such a cost-effective solution could provide a stronger incentive for food businesses, especially Small and Medium Enterprises (SMEs), to adopt food traceability technologies.

FATO can be applied as a means for improving food allergen management practices in food manufacturing. As detailed in Section , during the development of FATO we included relevant good practices to tackle common operational errors that are more likely to lead to food allergy recalls across food businesses (Jia and Evans 2021a; Soon et al. 2020). By aligning the audit process with these good practices, food businesses that employ FATO as a model for capturing relevant knowledge can benefit from 1) avoiding common operational errors; 2) providing more effective training for food operators; and 3) reducing the risk of food allergy recalls.

A further application of FATO is linked to the development of "free-from" food products. It has been predicted that the 'free-from" food product market will be doubled by 2030 (FactMR 2022). To access this market, food businesses need to go beyond examining ingredient lists for their products. For instance, to safely claim that a food product is free from nuts, knowing that ingredients of a food product do not contain nuts is far from sufficient, as its manufacturing processes may introduce traces of nuts. FATO can help food businesses examine detailed ingredient and food allergen management information along the supply chain and explore the potential to develop "free-from" food products from their existing product portfolio. For example, FATO permits the detection of all the sources of a certain food allergen, such as gluten or milk, along the food supply chain to identify the strategies and operations required in order to develop a food product "free-from" the food allergen.

Food businesses often do not have sufficient incentives to adopt food traceability technologies (Jia and Evans 2021a). The development of FATO considered incentives of food businesses. The key is to increase co-benefits of using FATO while removing barriers for food businesses to improve food allergen management (Jia and Evans 2021a). More specifically, the incentives of applications include but are not limited to the assurance of data confidentiality without requiring a central data system, reducing time and cost of food allergen management, supporting more effective food allergen management and audits, increasing team moral through reducing operational errors and assisting the development of new products.

Beyond the immediate supply chain, FATO also enables applications in downstream sectors such as healthcare, food service and hospitality. Here, providers act as data consumers rather than generators. Adoption in these contexts relies on integrating FATO-based querying into catering management or patient safety systems. For example, a hospital system could cross-reference a patient's allergy profile against FATO-structured meal data, validating suitability more rigorously than simple label checking.

FATO's alignment with FoodOn ensures inherent compatibility with BFO-compliant ontologies, enabling integration into systems leveraging the BFO structure. Conversely, for non-BFO compliant systems (e.g., those relying solely on GS1/EPCIS), FATO supports domain-level alignment. Local concepts can be mapped directly to core FATO classes (e.g., Product) without adopting the full BFO hierarchy. FATO thus serves as a semantic bridge, retaining the rigour of a BFO-backed structure while permitting flexible industrial adoption.

Conclusion

In this paper, we have presented FATO, an ontology for food allergen traceability. By tackling the shared operational errors in food allergen management, our ontology is scalable across food businesses by design. Depending on existing infrastructure, OWL, XML or JSON formatting can be utilised to enable different food businesses to automate data exchange and track food allergens along the supply chain more effectively and efficiently. Our ontology brings a new vision for improving food allergen management through ontology-based data interoperability technologies that can address traceability issues for the majority of food businesses, especially Small and Medium Enterprises (SMEs). While many food businesses encounter various barriers in adopting traceability technologies (e.g. affordability and lacking knowledge and skills), FATO seeks to provide a more cost-effective and flexible solution.

Some limitations of FATO are included in Appendix . The application of FATO is not limited to businesses in the UK food and drink sector. The collected scientific knowledge of improving food allergen management used in its development encompasses both national and international studies, as well as practical insights drawn from historical food allergy recalls in the UK. However, it is recognised that food safety regulations differ across countries, and future iterations of FATO will consider regulation variability among different countries. Strategies to address this include the mapping and alignment of FATO with other major regulatory frameworks (e.g., FDA) to ensure semantic interoperability. Additionally, our approach as illustrated in Figure 1 is transferable to other related domains such as environmental traceability (Jia et al. 2023). The authors are currently exploring the development of an environmental traceability ontology to help food companies report their environmental performance rigorously, transparently and efficiently, as part of the SEEBEYOND project (Jia and Evans 2022), and has been included in the Food Data Transparency roadmap in the UK (Department for Environment and Rural Affairs 2024).

... However, it is recognised that food safety regulations differ across countries, and future iterations of FATO will consider regulation variability among different countries. Crucially, FATO's structural reliance on global standards such as GS1/EPCIS and FoodOn ensures that its core framework supports international trade, allowing for the seamless integration of diverse regulatory requirements in future extensions.

Future development of FATO will focus on integrating provenance information using PROV-O (Lebo et al. 2013) to determine the food business employees responsible for each declaration. Moreover, a digital platform would be needed in the future to provide a user interface for food businesses to either verify their standardised food data schema against the ontology or provide data to populate the ontology. The digital platform can also be packaged as a service-based application (Baryannis and Plexousakis 2013, 2014; Baryannis et al. 2017) to facilitate integration in existing platforms of food businesses. Moreover, the data collected from the platform will be used for developing a human-in-the-loop AI approach (Wu et al. 2022) to gain insights and improve both the ontology and the standardised FPIF iteratively.

Acknowledgements

Lili Jia is funded by the Economic and Social Research Council (ESRC) Policy Fellowship during 1 February 2022 and 31 July 2023. Grant reference is ES/W008440/1.

References

Alobaid A, Garijo D, Poveda-Villalón M, Santana-Perez I, Fernández-Izquierdo A and Corcho O (2019) Automating ontology engineering support activities with ontoology. *Journal of Web Semantics* 57: 100472. DOI:https://doi.org/10.1016/j.websem.2018.09.003.

Baryannis G, Kritikos K and Plexousakis D (2017) A specification-based QoS-aware design framework for service-based applications. *Service Oriented Computing and Applications* 11(3): 301–314. DOI:10.1007/s11761-017-0210-4.

Baryannis G and Plexousakis D (2013) WSSL: A Fluent Calculus-Based Language for Web Service Specifications. In: Salinesi C, Norrie MC and Pastor Ó (eds.) *25th International Conference on Advanced Information Systems Engineering (CAiSE 2013)*, *Lecture Notes in Computer Science*, volume 7908. Springer Berlin Heidelberg. ISBN 978-3-642-38708-1, pp. 256–271. DOI:10.1007/978-3-642-38709-8_17.

Baryannis G and Plexousakis D (2014) Fluent Calculus-based Semantic Web Service Composition and Verification using WSSL. In: Lomuscio A et al. (eds.) *9th International Workshop on Semantic Web Enabled Software Engineering (SWESE2013), co-located with ICSOC 2013*, *Lecture Notes in Computer Science*, volume 8377. Springer International Publishing Switzerland, pp. 256–270. DOI:10.1007/978-3-319-06859-6_23.

Begen FM, Barnett J, Barber M, Payne R, Gowland MH and Lucas JS (2018) Parents' and caregivers' experiences and behaviours when eating out with children with a food hypersensitivity. *BMC public health* 18(1): 1–10.

Blomqvist E, Hammar K and Presutti V (2016) Engineering Ontologies with Patterns - The eXtreme Design Methodology. In: Hitzler P, Gangemi A, Janowicz K, Krisnadhi A and Presutti V (eds.) *Ontology Engineering with Ontology Design Patterns*, *Studies on the Semantic Web*, volume 25. IOS Press. ISBN 978-1-61499-676-7, pp. 23–50.

Conrado AB, Patel N and Turner PJ (2021) Global patterns in anaphylaxis due to specific foods: a systematic review. *Journal of Allergy and Clinical Immunology* 148(6): 1515–1525.

Department for Environment and Rural Affairs (2024) FDTP: towards consistent, accurate and accessible environmental impact quantification for the agri-food industry. URL https://www.gov.uk/government/groups/food-data-transparency-partnership.

Dooley DM, Griffiths EJ, Gosal GS, Buttigieg PL, Hoehndorf R, Lange MC, Schriml LM, Brinkman FSL and Hsiao WWL (2018) FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *npj Science of Food* 2(1): 23–. DOI:10.1038/s41538-018-0032-6.

EuroFIR AISBL (2025) FoodExplorer: The EuroFIR Food Data Interface. URL https://www.eurofir.org/FoodEXplorer/.

European Food Safety Authority (2015) The food classification and description system foodex 2 (revision 2). *EFSA Supporting Publications* 12(5): 804E. DOI:https://doi.org/10.2903/sp.efsa.2015.EN-804.

FactMR (2022) Free from food market. URL https://www.factmr.com/report/3272/free-from-food-market/.

Fernández M, Overbeeke C, Sabou M and Motta E (2009) What makes a good ontology? a case-study in fine-grained knowledge reuse. In: Gómez-Pérez A, Yu Y and Ding Y (eds.) *The Semantic Web*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 61–75.

Fernández-Izquierdo A and García-Castro R (2019) Themis: a tool for validating ontologies through requirements. In: *International Conference on Software Engineering and Knowledge Engineering*. pp. 573–578.

GS1 (2022) Epcis standard. Technical report, GS1. URL https://ref.gs1.org/standards/epcis/.

Jia L and Evans S (2021a) Improving food allergen management in food manufacturing: An incentive-based approach. *Food Control* 129: 108246. DOI:10.1016/j.foodcont.2021.108246.

Jia L and Evans S (2021b) Prevent food allergy alerts: an incentive-based approach. Technical report, Food Standards Agency. DOI:10.46756/sci.fsa.flm647. Project ref. FS301087.

Jia L and Evans S (2022) Standardising environmental metrics in the food and drink sector: an incentive-based regulatory approach. Technical report, Environmental Agency. Project ref. ENV6005288R.

Jia L, Prabhu J and Evans S (2023) How food companies can better measure their sustainability. *Harvard Business Review* 05(13).

Karpatne A, Kannan R and Kumar V (2022) *Knowledge Guided Machine Learning: Accelerating Discovery using Scientific Knowledge and Data*. Boca Raton, FL: Chapman and Hall/CRC.

Kotis KI, Vouros GA and Spiliotopoulos D (2020) Ontology engineering methodologies for the evolution of living and reused ontologies: status, trends, findings and recommendations. *The Knowledge Engineering Review* 35: e4.

Lack G (2008) Epidemiologic risks for food allergy. *Journal of Allergy and Clinical Immunology* 121(6): 1331–1336. DOI: https://doi.org/10.1016/j.jaci.2008.04.032.

Lebo T, Sahoo S, McGuinness D, Belhajjame K, Cheney J, Corsar D, Garijo D, Soiland-Reyes S, Zednik S and Zhao J (2013) PROV-O: The PROV ontology. *W3C recommendation* 30.

Pearson S, Brewer S, Godsiff P and Maull R (2021) Food Data Trust: A framework for information sharing. Technical report, Food Standards Agency. DOI:10.5281/zenodo.4575565. Project ref. FS301083.

Pizzuti T, Mirabelli G, Sanz-Bobi MA and Goméz-Gonzaléz F (2014) Food Track & Trace ontology for helping the food traceability control. *Journal of Food Engineering* 120: 17–30. DOI:https://doi.org/10.1016/j.jfoodeng.2013.07.017.

Pomponi D et al. (2020) The WHO/IUIS Allergen Nomenclature: Providing a Common Language. *Molecular Immunology* 120: 102–111. Official description of the IUIS nomenclature system.

Poveda-Villalón M, Gómez-Pérez A and Suárez-Figueroa MC (2014) OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)* 10(2): 7–34.

Redan BW (2020) Processing aids in food and beverage manufacturing: Potential source of elemental and trace metal contaminants. *Journal of agricultural and food chemistry* 68(46): 13001–13007.

Shimizu C, Hammar K and Hitzler P (2022) Modular Ontology Modeling. *Semantic Web* Pre-press(Pre-press): 1–31. DOI: 10.3233/SW-222886.

Soon JM, Brazier AK and Wallace CA (2020) Determining common contributory factors in food safety incidents–a review of global outbreaks and recalls 2008–2018. *Trends in Food Science & Technology* 97: 76–87.

Taylor SB, Christensen G, Grinter K, Sherlock R and Warren L (2019) The Allergen Bureau VITAL Program. *Journal of AOAC INTERNATIONAL* 101(1): 77–82. DOI:10.5740/jaoacint.17-0392.

Turchet L, Bouquet P, Molinari A and Fazekas G (2022) The smart musical instruments ontology. *Journal of Web Semantics* 72: 100687. DOI:https://doi.org/10.1016/j.websem.2021.100687.

US Food and Drug Administration (2025) Title 21, Code of Federal Regulations: Food and Drugs. Electronic Code of Federal Regulations (eCFR). URL `https://www.ecfr.gov/current/title-21`.

Valarakos AG, Karkaletsis V, Alexopoulou D, Papadimitriou E, Spyropoulos CD and Vouros G (2006) Building an allergens ontology and maintaining it using machine learning techniques. *Computers in Biology and Medicine* 36(10): 1155–1184. DOI: https://doi.org/10.1016/j.compbiomed.2005.09.007.

Vrandečić D, Pinto S, Tempich C and Sure Y (2005) The DILIGENT knowledge processes. *Journal of Knowledge Management* 9(5): 85–96.

Wu X, Xiao L, Sun Y, Zhang J, Ma T and He L (2022) A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems* 135: 364–381. DOI:https://doi.org/10.1016/j.future.2022.05.014.