prFood: Ontology principles for provenance and risk in the food domain

Belfrit Victor Batlajery, Mark Weal, Adriane Chapman
Electronics and Computer Science
University of Southampton
Southampton, United Kingdom
Email: {b.v.batlajery, m.weal, adriane.chapman}@soton.ac.uk

Luc Moreau
Informatics
King's College London
London, United Kingdom
Email: luc.moreau@kcl.ac.uk

Abstract—An ontology, a formal representation of domain knowledge, is an integral building block of the semantic web and is important in building an application within interdisciplinary domains. In the food domain, regulations require Food Business Operators (FBOs) to comply with traceability and track-ability measures when handling food in order to assess risk. In this paper, we identify several requirements to model food, food history, and risk of contamination in order to support the safety regulations of food. Using those requirements, we also identify and apply several design principles for building an ontology called prFood that encompasses interdisciplinary domains, food and risk. In order to apply safe handling of food in the food supply chain, we integrate and incorporate the pre-existing food ontologies with prFood by implementing a mapping procedure. Finally, we validate our approach by answering several use cases derived from EU and/or UK Food Regulations.

I. Introduction

The food supply chain is the systematic processing of food, consisting of all stages (processing, packaging, storage, distribution and retail, transport, handling, food preparation, and consumption) from the on-farm production to the consumption in homes, restaurants, and/or institutions (e.g. school, university, etc.) [1]. In principle, any party involved in the food supply chain must properly handle food to make sure that food is safe for consumption. Food safety itself can be defined as the efforts throughout the entire food supply chain to minimize the risk of contamination and to protect consumer's health [1] [2]. Food regulations are created to reduce the chance of food from being unsafe for human consumption. There are two reasons for having food regulations. First, to encourage all parties in the food supply chain to operate safely and reduce the risk of contamination [3]. For instance, Regulation (EC) 2073/2005 on microbiological criteria for foodstuffs requires a Food Business Operator[1] (FBO) to control the temperature when handling food. A second reason for creating the regulations is to enable traceability and track-ability of food across the food supply chain [4] [5]. As an example, Regulation 1224/2009 mandates the FBOs to keep their transactional records, which can help them or the food enforcement authority to trace where food comes from and goes to. This is important so

that food enforcement authorities can see an overview of the food supply chain and monitor the risk of contamination. Thus, a mechanism that allows food regulatory authorities to understand food has been processed, and assess the risk associated with that processing can help keeping food safe.

With a view to minimize the risk of food contamination and to support traceability and track-ability of food, 1995 Food Safety (General Food Hygiene) Regulation requires all FBOs to implement an HACCP-like system [6]. HACCP (Hazard Analysis and Critical Control Points) is intended to eliminate hazards (e.g. Salmonella, E. Coli, etc.), which may be fatal for human consumption. With HACCP-like systems, all FBOs involved in the food supply chain can review their own food processes and analyse them to anticipate an undesirable outcome. An HACCP-like system within an FBO, requires the details of food (its specification and how it should be handled) and the flow of food to be described [7]. The main objective of monitoring safety is to detect in a timely fashion, if there is an increase risk of food contamination. However, as food is traded globally, it is a challenge to have an overview of the risk over the whole food supply chain that involves many FBOs [8] [9].

This challenge motivates us to provide the food enforcement authority with a mechanism to understand food processes and have an overview over the risk across the food supply chain. In order to understand food processes and analyse the risk around it, we need to model food and its processes across the food supply chain that should allow one to perform analysis to identify potential contamination problems. Our modelling approach is through the use of semantic web technologies, which uses an ontology to model and capture knowledge about food and the risk of bacterial contamination. Therefore, we introduce our ontology, prFood, which models and captures food, including its specification, distribution, and transformation. prFood also supports risk assessment by capturing information that allows us to perform risk analysis. In order to develop prFood, we identify several design principles based on food regulations.

Overall, modelling and capturing the history of food and its associated risk provides some benefits, such as 1) Providing visibility or transparency of the interconnected food supply chain, 2) Providing crucial insight such as identifying critical

---

[1]A food business operator is a natural or legal person responsible for ensuring that the requirements of food law are met within the food business under their control [4].

food process and important food, and 3) Assessing downstream impact or tracing faults upstream. Our contributions in this work are as followed.

- Identification of the domain requirements for food regulation, including food, food history, and the risk of contamination,
- Development of the ontology design principles based on these requirements,
- Construction of prFood on the basis of those principles and validation through several use cases,
- Mapping of prFood with other food ontologies through prFoodMapping for data exploration.

The structure of the paper is as followed. The important background concepts of provenance, legacy food ontologies, and the risk framework are discussed in the Section II. Section III contains the list of requirements. In Section IV, ontology design principles identified from Food Regulations are discussed. The application of design principles and the use cases are described in Section V and Section VI successively. The paper is concluded with Section VII.

II. Background

In order to understand about the food processes from production to consumption, we introduce the notion of provenance. With regards of food, food provenance can be defined as a record that describes a food product and its ingredients, the processes involved in food transformation, and the organizations who are responsible in those processes from the source to consumption. Provenance plays an important role as a record of what happened to food along its lifetime, ultimately modelling the history of food.

prFood is designed to capture the history of food, and the necessary information regarding food processes and risk of bacterial contamination in order to compute risk assessments. As an example, Figure 1 shows the flow of food that travels through three consecutive processes, namely Transporting, Preparing, and Cooking. Each process is associated with a risk model, which affects the load of bacteria in food, and a set of food regulations, which regulates how food should be handled properly. In addition, each process also has risk factors[2] that affect the contamination risk of the food product. These need to be modelled as parameters in the risk model in order to calculate the risk of bacterial contamination.
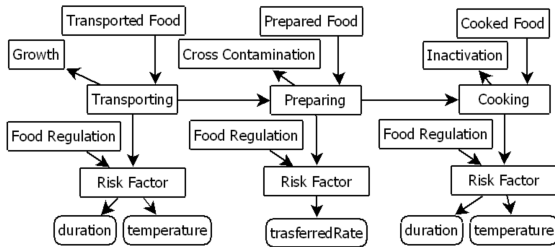

Fig. 1: An example of a food chain with prFood.

[2]A risk factor is an any aspect that contributes to the risk of contamination, such as improper storage of food, substantial time and temperature abuse of food, etc. [10].

A. Provenance

Provenance has been recognized in many domains as a way to describe how or where things are derived from [11]. W3C (World Wide Web Consortium) defines provenance as a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing in the world [12]. It contains the description of data and the processes involved during the data lifetime, such as how something is derived, who is responsible for certain actions, etc. Provenance can be used to describe the lifetime of an artefact. It can be a crucial information to determine whether the information should be trusted or not, and can provide a deep insight about data [13] [14] [15].

In order to model provenance, the W3C defines a standard language for exchanging provenance called PROV[3]. PROV is intended to facilitate the machine-processable data model for provenance [16]. PROV consists of several building blocks to describe provenance, which are PROV-DM (the conceptual data model), PROV-CONSTRAINTS (to validate PROV documents), PROV-O, PROV-XML, PROV-N (the serialization formats for PROV documents), and PROV-AQ (the mechanism to access and query PROV document). As a conceptual data model (standard interchange format), PROV-DM[4] allows provenance information to be interchanged between systems. It has three core concepts, namely *prov:Entity*, *prov:Activity*, and *prov:Agent*, and their relation is depicted in Figure 2. Example definitions of those concepts are shown in Table I.
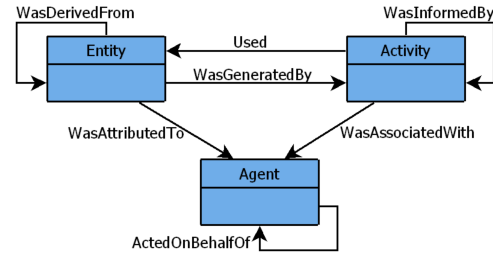

Fig. 2: The Core Concepts of PROV.

Concept	Definition	Example
Entity	A physical, digital, conceptual, or other kind of thing with some fixed aspects; entities may be real or imaginary.	Chicken, knife, egg, etc.
Activity	Something that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using, or generating entities.	Cutting, storing, mixing, etc.
Agent	Something that bears some form of responsibility for an activity taking place, for the existence of an entity, or for another agent's activity.	FBO, School, etc.

TABLE I: Definition of the entity, activity, and agent.

Although PROV is suitable for modelling the retrospective history of an entity [17], it is domain agnostic. Consequently,

[3]https://www.w3.org/TR/prov-overview/. Accessed: 22 September 2017
[4]https://www.w3.org/TR/prov-dm/. Accessed: 22 September 2017

many ontologies extend PROV-O (PROV Ontology) for their specific purposes. In their work, Ali and Moreau proposed a provenance-aware policy language (cProvl) to capture traceability data, and express policies in cloud-based service model [11]. To achieve their aims, an ontology cProv was created using PROV notation that allows traceability in cloud-based service at the service/platform level. Another work by Garijo also extends the use of PROV-O for documenting workflow plans in terms of steps and variables, known as P-PLAN [18]. Packer et al. developed a provenance ontology in social computations domain with the aim to measure a community based-service reputation [19]. This ontology is used to explain entities used in social computation. In the same domain, Markovic et al. also presents SC-PROV as a provenance vocabulary for social computation [17]. Their work expands the use of PROV-O and P-PLAN to enable descriptions of some conditions and incentives as part of the social computation plan. In this work, we extend PROV-O to enable risk calculation in the food domain.

B. Other food ontologies

Many ontologies in the food domain have been developed to model food knowledge. Pizzuti et al. summarizes the ontologies in the domain of food, then propose their ontology to trace and track a food product, known as Food Track&Trace Ontology or FTTO [20]. The FTTO ontology aims to integrate and connect the main features of the food traceability domain with the Global Track&Trace System. To achieve their goal, FTTO includes the representative food concepts in a supply chain, such as service products, processes and actors involved, in a single hierarchy. However, the notion of food contamination is missing in FTTO. Thus, although one can trace the contaminated food, it is difficult to explain what might the cause of contamination.

The other ontology is developed by Markovic et al. [21]. Their ontology aims to monitor food safety by documenting constraints that may be associated with an HACCP plan[5]. Data about time and temperature is captured within their ontology to monitor how meat is treated in the kitchen as a means to provide an alert system if regulation is abused. In addition, their work also extended other existing ontologies with food safety ontology to model HACCP based food preparation. However, the constraints are limited to the cooking process in the kitchen alone. Consequently, other food processes, such as transporting or storing are not modelled, making it difficult to describe what happened with food before the cooking process.

In a similar vein, Xiangwei and Lin developed an ontology to model HACCP knowledge in a food cold chain [22]. Although their ontology models the information in multiple processes, it is limited to the processes that maintain the low-temperature (cold chain). Thus, processes with a high-temperature (cooking) or cross-contamination between food products are not modelled in this ontology. In addition, as cold chain is often performed by a food enterprise, this ontology does not focus on the consumer at home, where most food contamination is likely to occur [23] [24] [25]. Overall, most of the legacy food ontologies lack the knowledge of quantifying harmful bacteria as a way to assess the risk of contamination, which is a key difference from prFood.

In practice, most FBOs have their own data model that was built to capture the information about food they handle. For example, a distribution company may have a set of data, which captures the distribution of food to its customers. This set of data may have an ontology too. Therefore, mapping prFood with pre-existing ontologies is needed to integrate our data with other data.

C. MPRM as a risk framework

As prFood is designed to convey the notion of risk, we adopt the risk framework proposed by Nauta called Modular Process Risk Model (MPRM) [26]. MPRM is a process-driven framework to estimate the risk of food contamination based on how food is handled. It was developed as a tool to statistically model the transmission of micro-organisms from one food process (e.g. cooking, retailing, transporting, etc.) to another process in the food supply chain.

With this framework, one can predict an expected number of micro-organisms and start to evaluate the risk of contamination due to microbial exposure across the food supply chain. As a result, identifying paths leading to the higher number of micro-organisms becomes feasible, and preventive actions can be performed in the process where high level of micro-organisms is introduced [27] [28].

In MPRM, there are 6 basic processes that can affect the microbial number after the food process. They are Growth, Inactivation, Partitioning, Mixing, Removal, and Cross Contamination. Growth and inactivation are 2 basic microbial processes that strongly depend on the characteristic of micro-organisms investigated and surrounding environmental conditions. Meanwhile, partitioning, mixing, removal, and cross-contamination are 4 handling processes that are assigned depending on how food is handled.

III. REQUIREMENTS FOR PRFOOD

In the process of modelling food, we use a dataset about the food supply chain in Hampshire County that managing kitchens for approximately 600 schools. Those kitchens prepare and provide menus to pupils at schools. There are also several distributors, who deliver the food ingredients to be processed into a meal. Those distributors also get supplied by the food suppliers (inside or outside the United Kingdom), who produce the food products. In this context, we identify several FBOs (e.g. suppliers, distributors, kitchens, etc.) and requirements from the EU or UK food regulations that they should comply.

A. Requirement A: Traceability in the food supply chain

In order to comply with European legislation regarding traceability (Regulation (EU) 178/2002, the principle of food and feed traceability), all FBOs need to keep the transactional

---

[5]A HACCP plan is a written HACCP based procedure.

documents safely. These should include the details of the sold food products, the suppliers, and the customers. Keeping those documents safe makes it easier for an FBO or an food enforcement authority to reconstruct past events and satisfiy the concept of *one-up/one-down*[6]. In fact, food provenance potentially allows us to do more than *one-up/one-down* because it can describe the entire history of food from its source. This requirement can be satisfied by looking at transactional documents (e.g. invoice, order, etc.) that identify the seller and buyer of food. It is exemplified in Use Case 1.

B. Requirement B: Nutritional information about food shall be provided to consumers

To comply the Regulation (EU) 1169/2011 about food information to consumers, an FBO must provide nutritional information to its costumers. Nutritional information on food products helps consumers choose which food is more nutritious, leading to have a healthy life. For the consumers with a special diet, nutritional information can help them with food that suitable with their health condition. To comply with this requirement, our approach should be able to check the compositional ingredients of nutritional information in a food product. This is demonstrated in Use Case 2.

C. Requirement C: An FBO shall minimize the risk of contamination

Regulation (EC) 2073/2005 on microbiological criteria for foodstuffs requires all FBOs to control harmful bacteria levels. This regulation applies to all FBOs involved in the processing, manufacturing, handling, and distribution of food, including retailers and caterers. The growth of bacteria is highly affected by temperature. Therefore, to comply with this regulation, it is important to control the temperature when handling food. As a guideline, Regulation (EC) 852/2004, Regulation (EC) 853/2004, and The Food Hygiene Regulations 2006 provide the requirement of temperatre control that should be met. These regulations are intended to help identify acceptance handling and temperature boundary for both the FBO and food enforcement authorities. Therefore, by checking the temperature used in the food processes (Use Case 4), helps an FBO to comply with the regulation.

D. Requirement D: An FBO shall monitor their food processes to review and understand risk around food

Controlling and monitoring food ensures the HACCP plan still produce food that is safe for consumption. The monitoring can be done by conducting a food sampling. Therefore, it is crucial to keep the record of sampling reports to be able to justify the risk of contamination. This is known as discontinuous monitoring. When the result of the sampling is not accepted, it may indicate that HACCP should be revisited in order to meet the minimum standard of safety food.

Another procedure of monitoring is continuous monitoring, by which the monitoring is performed in a timely basis.

Continuous monitoring procedure is preferable as it helps assure produced products have met the acceptance criteria. An example of this type of monitoring is temperature monitoring, which control or maintain the temperature level at the acceptance level. As the same as Requirement C, monitoring the temperature in the food process can help to reduce the chance of food contamination is shown in Use Case 4.

E. Requirement E: HACCP principles must be applied in food processes

According to 1995 Food Safety (General Food Hygiene), it is required for an FBO to have an HACCP-like system when handling food and the guideline for handling or storage of food is stated in Regulation (EU) 931/2011. Overall, in any stage along the food supply chain, HACCP principles must be applied to identify and minimize the hazard. This includes identifying potential harmful bacteria, establish control/critical limit, establish monitoring/verification procedures, and establish corrective action for each of the food processes in food chain [7]. Having the complete details about HACCP is beneficial to investigate the food supply chain if sampling food results in a bad outcome. Checking the instruction or the HACCP description in food product is a way to comply this requirement (Use Case 3).

IV. Ontology design principles

On the basis of the food requirements in the previous section, we identify several principles in developing an ontology that models food, food history, and allows computation of risk. The objective is to model the food ecosystem and capture all necessary information regarding provenance and the risk factors associate with bacterial contamination. These risk factors are known as the parameters, which allow one to perform risk calculation and provide an explanation about the history of food. As a result, PROV-O becomes the core ontology of prFood to support the provenance of food.

As pre-existing food ontologies are independently developed from PROV-O and prFood, we also provide a mapping procedure between prFood and those legacy ontologies. The mapping process is done by using an ontology, prFoodMapping, which maps the related classes of the legacy food ontologies as the subclasses of the classes defined in prFood.

A. Provenance is a backbone of the ontology

The core requirement of traceability in the food supply chain means that provenance serves as a backbone in designing prFood. In order to convey the notion of provenance in prFood, PROV-O is extended to the food domain by defining all classes in prFood as subclasses of PROV-O. In prFood, we define the food concept (e.g. food, bacteria, sampling report, etc.) as the subclasses of *prov:Entity* and the processes related with food (e.g. analysing, cooking, etc.) as the subclasses of *prov:Activity*. Moreover, all the parties/actors in the food supply chain (e.g. laboratory, retailer, inspector, etc.) are defined as subclasses of *prov:Agent*. With provenance as an ontology backbone, we address Requirement A and Requirement E.

---

[6]A *one-up/one-down* concept is a concept in tracing and tracking food to explain the connection between inputs and outputs of food an FBO treats.

B. Collaborative and automatic risk assessment

To be able to perform automatic risk assessment, we define some properties that capture the parameters for the purpose of risk calculation. These parameters (a.k.a risk factors) contribute to the growth of harmful bacteria when handling food. In addition, mapping between prFood and the legacy food ontologies allows collaborative risk assessment. The datasets captured with the legacy food ontologies from different systems can be mapped with prFood and exploited when we perform risk calculation. This principle addresses Requirement C, Requirement D, and Requirement E.

C. Ontology share-ability/Expansible through meta-modelling

Based on the principle of share-ability, we provide a design that allows other ontologies to be integrated with prFood by mapping procedure. As a result, although the legacy food ontologies are developed independently from prFood, it does not limit us to query and return information that supports risk calculation. With this principle, we address Requirement B.

D. Independence from the ontology datasets

In principle, prFood and the ontology datasets are independent each other. Some food systems may have developed their own food ontologies to help them capture and describe data about food; yet, their captured information can be used for the purpose of risk assessment with mapping procedure. With this principle, we address Requirement B and Requirement D.

V. APPLICATION DESIGN PRINCIPLES

From the application point of view, prFood is developed on the basis of PROV to convey the notion of provenance in the food domain. The modelling of prFood is done by extending PROV-O to match a real food ecosystem, capturing necessary information about food, its history, and information that supports risk assessment, in particular risk of contamination. As an ontology, prFood allows sharing vocabulary between heterogeneous systems to trace and track provenance information of food. Because prFood captures the food provenance in standardized format (PROV-O, OWL, RDF), it brings some benefits, such as monitoring the risk factors can be performed automatically and food information can be exchanged with the common understanding among systems [21].

However, many food ontologies already exist, and our mechanism for risk calculation is more powerful if we can easily extend it into specialized food sub-domains. Therefore, we provide a mapping mechanism, by which we map the classes and properties in prFood to equivalent classes and properties in each legacy food ontology. This process allows SPARQL queries to exploit the legacy food ontologies via prFoodMapping.

All classes described in prFood are subclasses of either an Entity (*prov:Entity*), an Activity (*prov:Activity*), or an Agent (*prov:Agent*). Meanwhile, the object properties (the properties between objects) in prFood are described in verbal form of past tense to express the past evidence associated with food. We also define data properties (the link properties between

object and literal) to capture the data about food or risk factors. The snippet of the classes, object properties, and data properties in prFood is shown in Figure 3. We also adapt several properties from PROV-O, such as *prov:wasDerivedFrom*, *prov:wasGeneratedBy*, and *prov:used* to explicitly capture how a certain thing came from. For instance, the transformation of food can be modelled in triple as followed.

prFood:CookedFood   prov:wasDerivedFrom   prFood:StoredFood

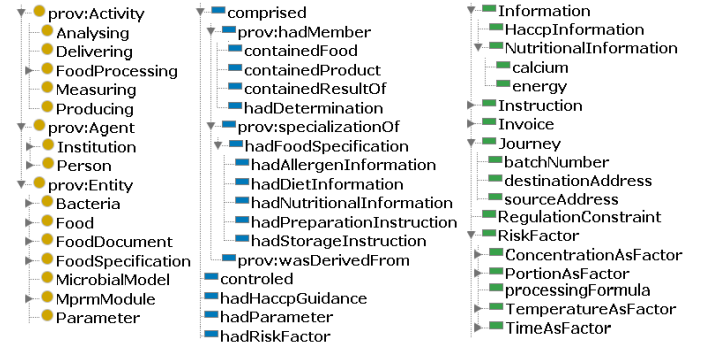This makes provenance an important aspect in prFood and Requirement A is addressed.

Fig. 3: Classes (left), Object Properties (middle), and Data Properties (right) in prFood.

All classes of food processes defined in prFood (*prFood:FoodProcessing*) will have some data properties to detail how the process is performed with a view for investigation if contamination outbreak occurs. For example, *prFood:HaccpInformation* is a data property that explicitly describes the HACCP plan in food processes. In addition, class *prFood:FoodProcessing* is linked to class *prFood:Parameter*, which explicitly capture all parameters, such as the duration and temperature. Those parameters are the important factors to automatically assess the risk of contamination, and known as the risk factors. The class *prFood:Parameter* will be put against class *prFood:FoodRegulation* that defines the acceptance level of the risk factors. In regard to the automatically risk assessment, class *prFood:MicrobialModel* captures the mathematical formulation to estimate the number of bacteria in food after each food process is performed. The calculation formula depends on the type of bacteria (*prFood:Bacteria*), the MPRM basic module (*prFood:MprmModule*), and the parameters/risk factors (*prFood:Parameter*). The ability to combine tracing and assessing risk addresses Requirement C, Requirement D, and Requirement E.

To be able to explain the specification of food, we provide class *prFood:FoodSpecification* that is a superclass of *prFood:AllergenInformation*, *prFood:NutritionalInformation*, etc. Each of the subclasses is attributed with the data properties that capture the related information about the details of food (e.g. *prFood:energy* is a data property in class *prFood:NutritionalInformation* that captures the amount of

energy in food). This design addresses Requirement B.

VI. USE CASES

In this section, we present several use cases to reflect the requirements presented in section III. The use cases are also based on the real requirements from EU and UK food regulations. Each use case is validated by SPARQL queries to answer the requirements, validating our design of prFood.

Use Case 1: Can we establish the chain of sellers and buyers for a food product every where?

The final aim of this use case is to check whether an FBO keeps their transactional documents in complying with Requirement A. In order to achieve that, we identify the transactional documents (e.g. invoice, order, etc.) as they contain information about the distributors (sellers) and customers (buyers) of the food product. Each FBO involved in the food supply chain plays its own role to process and handle food. Therefore, by exploring transactional documents, we are able to build a chain of sellers and buyers of food product.

To extend the result, we also identify all of the transactional documents from other food legacy ontologies. These documents are the equivalent documents with transactional document in prFood; thus, are the subclasses of transactional document from prFood in prFoodMapping. Listing 1 and Listing 2 show the SPARQL queries to answer our first use case.

```
PREFIX prFood:<https://provenance.ecs.soton.ac.uk/food#>
SELECT DISTINCT *
WHERE {
  ?invoice rdf:type prFood:Invoice ;
           prFood:wasReleasedBy ?distributor ;
           prFood:wasTriggeredBy ?orderdoc .
  OPTIONAL{ ?orderdoc rdf:type prFood:Order ;
                       prFood:wasReleasedBy ?customer .}
}
```
Listing 1: SPARQL query to return the invoices details from prFood RDF.

```
PREFIX prFood:<https://provenance.ecs.soton.ac.uk/food#>
PREFIX prFoodMapping:<https://provenance.ecs.soton.ac.uk/FoodMapping#>
SELECT DISTINCT *
WHERE {
  ?subInvoice rdfs:subClassOf prFood:Invoice .
  ?subWasReleasedBy rdfs:subPropertyOf prFood:wasReleasedBy .
  OPTIONAL{?otherInvoice rdf:type ?subInvoice ;
                         ?subWasReleasedBy ?distributorORcostumer .}
}
```
Listing 2: SPARQL query to return the invoices a legacy food ontology via mapping.

In Listing 1, we query the instances of class *prFood:Invoice* and try to find its distributor and customer. To find the distributor of the invoice, we use *prFood:wasReleasedBy* that is a subclass of *prov:wasAttributedTo* to indicate an activity that associates with agent. To find the customer of the invoice, we query the order through *prFood:wasTriggeredBy*; subsequently, query the customer of that order by using *prFood:wasReleasedBy*.

In Listing 2, the mapping procedure is performed, in which we query all subclasses of *prFood:Invoice* in PrFoodMapping

(the mapping ontology). These subclasses are the equivalent classes of *prFood:Invoice* in the legacy food ontologies. In this case, the subclass of *prFood:Invoice* is *xyzDistributor:InvoiceXYZ*. Next, we query all subproperties of *prFood:wasReleasedBy* in prFoodMapping to identify its equivalent property in the legacy food ontology. In this case, the subproperty of *prFood:wasReleasedBy* are *xyzDistributor:hasCustomer* and *xyzDistributor:hasDistributor*. After we query this class, which is equivalent with *prFood:Invoice*, we can immediately query the distributor and the customer through *prFoodMapping:subWasReleasedBy*, because the details of distributor and the customer are described in the invoice class in the legacy food ontology.

With *prov:wasAttributedTo*, we can extend the case by identifying the party that performs a food process along the food supply chain. Each food process (e.g. *prFood:Cooking*) results in a certain type of food transformations (e.g. *prFood:CookedFood*). Thus, we can identify who performs which food process and what is the result of that process by using provenance.

Use Case 2: Does an FBO provide consumers with nutritional information about food they market?

In this use case, we query the nutritional information about the marketed food products to comply Requirement B. In most cases, the description of the food product comprises the characteristic of the product, including the nutritional information. Therefore, the nutritional information of food can be retrieved by querying the description of the food product and filtering out unnecessary information.

With the same procedure as in Use Case 1, we first query prFood to find all information about the food product, including the nutritional information. To retrieve the nutritional information from the other food legacy ontology, we query prFoodMapping to find the equivalent classes of nutritional information. Once those classes are identified, we explore them to find all the necessary information. Listing 3 and Listing 4 show the SPARQL queries for Use Case 2.

```
PREFIX prFood:<https://provenance.ecs.soton.ac.uk/food#>
SELECT DISTINCT *
WHERE {
  ?food  rdf:type  prFood:Food ;
         prFood:hadNutritionalInformation  ?nutInfo .
  OPTIONAL{ ?nutInfo  rdf:type  prFood:NutritionalInformation ;
                      ?componentOfnutInfo  ?valueOfNutInfo .}
  FILTER(?componentOfnutInfo != rdf:type)
}
```
Listing 3: SPARQL query to nutritional information from prFood.

```
PREFIX prFood:<https://provenance.ecs.soton.ac.uk/food#>
PREFIX prFoodMapping:<https://provenance.ecs.soton.ac.uk/FoodMapping#>
SELECT DISTINCT *
WHERE {
  ?subFood  rdfs:subClassOf  prFood:Food .
  OPTIONAL{ ?otherFood  rdf:type  ?subFood ;
                        ?otherNutInfo  ?otherNutInfoValue .}
  FILTER(?otherNutInfo != rdf:type && ?otherNutInfo != rdfs:comment &&
         ?otherNutInfo != rdfs:label)
}
```
Listing 4: SPARQL query to nutritional information from a food legacy ontologies via mapping

Listing 3 shows the SPARQL query to a class that represents the food product (*prFood:Food*) and has information about its nutritional information (*prFood:NutritionalInformation*). Listing 4 shows the mapping procedure to identify the subclass of *prFood:NutritionalInformation* in prFoodMapping ontology. The subclass of *prFood:NutritionalInformation* represents the equivalent class of *prFood:NutritionalInformation* in another legacy food ontologies. Therefore, once we identify the subclasses of *prFood:NutritionalInformation*, we can start exploring the nutritional information in the other legacy food ontologies.

Use Case 3: Can we see the proper actions and corrective actions in each food process as the HACCP Guidance?

This use case is in compliance with Requirement E, where we query the details of how should FBOs properly handle food and what should they do in case of unexpected event. When describe the characteristic of product, its supplier will also describe how the product should be treated. The guidance provided is, most of the time, coherent with HACCP guidance as a standard regulation to protect food from being contaminated. Thus, having the description of the product will explain how we should handle and process the product.

We begin to query our food ontology, prFood, and identify the class that explain about how to handle the food according to the HACCP procedure. This class is also mapped in our mapping ontology, prFoodMapping, as a superclass for all the classes in the legacy food ontologies that also explain about how the food product should be handled.

```
PREFIX prFood:<https://provenance.ecs.soton.ac.uk/food#>
PREFIX prFoodMapping:<https://provenance.ecs.soton.ac.uk/FoodMapping#>
SELECT DISTINCT *
WHERE {
  ?food  rdf:type  prFood:Food ;
         prFood:hadHaccpGuidance  ?haccpClass .
  ?haccpClass  prFood:HaccpInformation  ?haccpDetails .
  prFoodMapping:subFoodOf  rdfs:domain  prFood:Food .
  prFoodMapping:sameHaccpGuidanceAs  rdfs:range  ?otherHaccp .
  OPTIONAL {?subFood  rdf:type  prFood:Food ;
                      ?otherHaccp  ?otherHaccpDesc .}
}
```

Listing 5: SPARQL query to return HACCP details from prFood and other legacy food ontology.

In Listing 5, we query *prFood:HaccpGuidance* with its equivalent classes in the legacy food ontologies. The equivalent classes of *prFood:HaccpGuidance* in legacy food ontologies will be the subclasses of *prFood:HaccpGuidance* in the prMappingFood to imply that the subclasses of *prFood:HaccpGuidance* hold the same notion as class *prFood:HaccpGuidance*.

Use Case 4: Can an FBO provide a food enforcement authority with the temperature details?

This use case is intended to address Requirement C and Requirement D. When handling food, the temperature needs to be maintained at the level where harmful bacteria cannot grow. In this use case, we present how to query the risk factors (temperature and duration) of handling food in retailer.

Although in this use case we only query the risk factors in the retailing process, it should work for the other food processes too (e.g. cooking, transporting, etc.).

We begin by identifying the class that represents the process in the food supply chain from prFood (e.g. *prFood:Retailing*). Subsequently, we explore the parameters of the process through the class property. Finally, we explore the risk factors associate with that process through temperature and time properties. We also query the subclasses of the process in prMappingFood, in order to identify the equivalent retailing class in the other legacy food ontologies.

```
PREFIX prFood:<https://provenance.ecs.soton.ac.uk/food#>
SELECT DISTINCT *
WHERE {
  ?retailing  rdf:type  prFood:Retailing ;
              prFood:hadParameter  ?retailingParameter .
  ?retailingParameter  rdf:type  prFood:Parameter ;
              prFood:duration  ?durationValue ;
              prFood:deviceTemperature  ?deciveTempValue .
}
```

Listing 6: SPARQL query to return the risk factors from prFood.

In Listing 6, we query a class with type *prFood:Retailing*. This class represents the retailing process in a retailer. Subsequently, we query the parameter (*prFood:Parameter*) used in retailing process through the property *prFood:hadParameter*. After we receive the class *prFood:Parameter*, we explore the duration and temperature used in retailing process.

We also query the subclasses of *prFood:Retailing* in order to identify the equivalent retailing class in the other legacy food ontologies. After identifying the equivalent retailing class in other food ontologies, we explore the class that is also equivalent to *prFood:Parameter* in order to extract the duration and temperature of that class.

VII. CONCLUSION

In this paper, we present an ontology with an ability to capture a cohesive view (*stage-by-stage* of the food processes) in the food supply chain, as well as capturing the provenance of food and the risk factors for the purpose of risk calculation. prFood is an ontology that models the food ecosystem, including its production, transformation, distribution, and potential risk of contamination. Its purpose is different from the existing legacy food ontologies as it models the knowledge of food and facilitates the quantification of risk.

From a provenance point of view, prFood shows that provenance can be contextualized in terms of food and risk. As prFood takes into account the notion of provenance, it facilitates us to characterize how food moves and is transformed across the food supply chain; ultimately, increasing the transparency. As a result, prFood makes the history of food visible for checking/auditing, verification, and further investigation. Additionally, modelling food in provenance standardized format is easier to convert it into graph representation for further analytical methods. However, this is beyond the scope in this paper.

prFood provides the ability to regulatory authorities to overview risk in the interconnected food supply chain. It

supports the identification of critical food process, and the capability to assess downstream impact or tracing faults upstream automatically. However, the capability of automatically perform risk assessment is again beyond the scope of this paper, and we address this as our future work. With the notion of provenance and risk assessment, prFood is able to not only draw the history of food, but also predict the risk of contamination in process food has been gone through.

Through the development of prFood, identified the food requirements for food safety and provided some use cases to reflect those requirements. These use cases are the examples cases any food system need to satisfy. Based on food requirements, we also identified several design principles and used them to build prFood that model the food supply chain. We demonstrated our model with sample data to satisfy our use cases and meet the requirements derived from food safety legislation.

Finally, our approach needs to be tested in the real situation with a real food supply chain and real data captured as food travel from one process to another processes. Thus, a small experiment with multiple FBOs will be our future work. Furthermore, we can apply our method to analyze risk and present the effectiveness of our methodology.

ACKNOWLEDGEMENT

REFERENCES

[1] World Health Organization and Agriculture Organization of the United Nations, "Exposure assessment of microbiological hazards in food," ATM Forum Contribution 94-0735R1, 2008. [Online]. Available: {http://www.who.int/foodsafety/publications/micro/MRA7.pdf}

[2] O. I. Aruoma, "The impact of food regulation on the food supply chain," *Toxicology*, vol. 221, no. 1, pp. 119–127, April 2006, nutraceuticals and Functional Foods Regulations in the United States and Around The World. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0300483X0600045X

[3] Food Standards Agency, "Food Law Code of Practice (England)-April 2015," Food Standards Agency, Report, April 2015. [Online]. Available: https://www.food.gov.uk/sites/default/files/Food%20Law%20Code%20of%20Practice%20-%202015_1.pdf

[4] European Parliament and the Council, "Regulation (EC) No 178/2002 of The European Parliament and of The Council," European Parliament and the Council, Report, January 2002. [Online]. Available: http://www.food.gov.uk/sites/default/files/multimedia/pdfs/1782002ecregulation.pdf

[5] A. Regattieri, M. Gamberi, and R. Manzini, "Traceability of food products:general framework and experimental evidence," *Journal of Food Engineering*, vol. 81, pp. 347–356, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0260877406006893

[6] J. E. Ehiri and G. P. Morris, "Food safety control strategies: A critical review of traditional approaches," *International Journal of Environmental Health Research*, vol. 4, no. 4, pp. 254–263, 1994. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/09603129409356824

[7] M. D. Pierson, *HACCP: principles and applications*. Springer Science & Busines, 1992.

[8] M. M. Aung and Y. S. Chang, "Traceability in a food supply chain: Safety and quality perspectives," *Food Control*, vol. 39, no. Supplement C, pp. 172 – 184, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0956713513005811

[9] G. Senneset, E. Forås, and K. M. Fremme, "Challenges regarding implementation of electronic chain traceability," *British Food Journal*, vol. 109, no. 10, pp. 805–818, 2007.

[10] W. H. Organization, *Risk assessments of Salmonella in eggs and broiler chickens*. Food & Agriculture Organization, 2002, vol. 2.

[11] L. Moreau and M. Ali, "A provenance-based policy control framework for cloud services," May 2014. [Online]. Available: https://eprints.soton.ac.uk/364997/

[12] L. Moreau, P. Groth, J. Cheney, T. Lebo, and S. Miles, "The rationale of prov," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 35, no. Part 4, pp. 235 – 257, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570826815000177

[13] M. D. Allen, A. Chapman, L. Seligman, and B. Blaustein, "Provenance for collaboration: Detecting suspicious behaviors and assessing trust in information," in *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2011 7th International Conference on*. IEEE, 2011, pp. 342–351.

[14] J. Golbeck and J. Hendler, "A semantic web approach to the provenance challenge," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 5, pp. 431–439, 2008.

[15] A. Toniolo, F. Cerutti, N. Oren, T. J. Norman, and K. Sycara, "Making informed decisions with provenance and argumentation schemes," in *Proceedings of the Eleventh International Workshop on Argumentation in Multi-Agent Systems*, 2014.

[16] P. Groth and L. Moreau, "Prov-overview. an overview of the prov family of documents," 2013.

[17] M. Markovic, P. Edwards, and D. Corsar, "Sc-prov: A provenance vocabulary for social computation," in *International Provenance and Annotation Workshop*. Springer, 2014, pp. 285–287.

[18] D. Garijo and Y. Gil, "Augmenting prov with plans in p-plan: scientific processes as linked data." CEUR Workshop Proceedings, 2012.

[19] H. S. Packer, L. Drăgan, and L. Moreau, "An auditable reputation service for collective adaptive systems," in *Social Collective Intelligence*. Springer, 2014, pp. 159–184.

[20] T. Pizzuti, G. Mirabelli, M. A. Sanz-Bobi, and F. Goméz-Gonzaléz, "Food track & trace ontology for helping the food traceability control," *Journal of Food Engineering*, vol. 120, pp. 17–30, 2014.

[21] M. Markovic, P. Edwards, M. Kollingbaum, and A. Rowe, "Modelling provenance of sensor data for food safety compliance checking," in *International Provenance and Annotation Workshop*. Springer, 2016, pp. 134–145.

[22] M. Xiangwei and Z. Lin, "An ontology development for haccp knowledge description and sharing in food cold chain," 2015.

[23] C. Byrd-Bredbenner, J. Berning, J. Martin-Biggers, and V. Quick, "Food safety in home kitchens: a synthesis of the literature," *International journal of environmental research and public health*, vol. 10, no. 9, pp. 4060–4085, 2013.

[24] C. Griffith, D. Worsfold, and R. Mitchell, "Food preparation, risk communication and the consumer," *Food control*, vol. 9, no. 4, pp. 225–232, 1998.

[25] L. J. Kagan, A. E. Aiello, and E. Larson, "The role of the home environment in the transmission of infectious diseases," *Journal of community health*, vol. 27, no. 4, pp. 247–267, 2002.

[26] M. J. Nauta, "A modular process risk model structure for quantitative microbiological risk assessment and its application in an exposure assessment of bacillus cereus in a repfed," *RIVM Rapport 149106007*, 2001.

[27] J. Dawber, B. Horn, and J. Brown, "Performing a quantitative microbial risk analysis using second-order monte carlo simulation," 2009.

[28] A. S. R. Duarte, "The interpretation of quantitative microbial data: meeting the demands of quantitative microbiological risk assessment," Ph.D. dissertation, National Food Institute, Technical University of Denmark, 2013.