Enhancing Visibility in EPCIS Governing Agri-Food Supply Chains via Linked Pedigrees

Monika Solanki, Aston Business School, Aston University, Birmingham, UK

Christopher Brewster, Aston Business School, Aston University, Birmingham, UK

ABSTRACT

Data integration for the purposes of tracking, tracing and transparency are important challenges in the agri-food supply chain. The Electronic Product Code Information Services (EPCIS) is an event-oriented GS1 standard that aims to enable tracking and tracing of products through the sharing of event-based datasets that encapsulate the Electronic Product Code (EPC). In this paper, the authors propose a framework that utilises events and EPCs in the generation of "linked pedigrees" - linked datasets that enable the sharing of traceability information about products as they move along the supply chain. The authors exploit two ontology based information models, EEM and CBVVocab within a distributed and decentralised framework that consumes real time EPCIS events as linked data to generate the linked pedigrees. The authors exemplify the usage of linked pedigrees within the fresh fruit and vegetables supply chain in the agri-food sector.

Keywords: *Agri-Food, Electronic Product Code Information Services (EPCIS), Linked Data, Ontologies, Semantic Web, Traceability, Visibility*

INTRODUCTION

One of the most important challenges in logistics and supply chains is information integration. Sharing of data and knowledge in a standardised manner along the supply chain is crucial not only to enable visibility, i.e., tracking and tracing of artifacts, but also to enable the more effective management of the supply chain. Business relations are increasingly globalised and loosely coupled thus making both standards and technologies for information exchange essential. This is particularly required in the food and agriculture sector due to the great complexity of the supply chains, and the importance of tracking and tracing for food safety and regulatory requirements.

Barcodes and more recently RFID tags have provided initial solutions to this challenge. GS1[1], a global organisation which manages barcodes, has provided an ever wider range of standards to facilitate end-to-end traceability and information sharing along the supply chain. The Electronic Product Code Information Services

DOI: 10.4018/IJSWIS.2014070102

(EPCIS)[2] and the Core Business Vocabulary (CBV)[3], collectively provide specifications for the representation of product traceability information (Främling, Parmar, Hinkka, Tätilä, & Rodgers, 2013).

EPCIS provides the technical specification to track product lots and generate messages for signalling the geographical progress and status of an item or set of items at each step of the supply chain. This is achieved by monitoring data generated within the business context of scanning a barcode or RFID tag and encapsulating it within the abstraction of an "event".

EPCIS, however, as currently designed, does not provide the framework for sharing data across a multiplicity of supply chain actors. For information to be usefully shared, it must be interlinked and made available at all stages of the supply chain (with due regard for data ownership). A limitation of the current EPCIS specification is that though it does propose a mechanism to exchange and share data by providing an XML schema against which event data can be recorded, Web services implementing the specification are tightly coupled. Achieving interoperability between Web services becomes a challenging task and incorporating any future changes in the specification results in huge maintenance overheads.

Three critical issues that act as a hindrance to enabling information exchange in existing supply chain processes are the following:

- In supply chains, particularly in the agri-food sector, the flow of data is restricted by a very conservative "need-to-know" attitude such that essentially information flows only "one up, one down".
- Although a very large number of items are scanned in food supply chains, and each actor records these events in their systems, there is no linkage of these data items across actors. This is due both to the cultural barriers to information exchange in the sector and the current set of technological solutions.
- Finally, the EPCIS XML schemas define only the structure of the data to be recorded. The semantics of data and data curation processes are informally defined. Their interpretation is left up to the individual EPCIS specification implementing engines, thereby greatly increasing the possibility of interoperability issues arising between supporting applications, e.g., validation and discovery services built over the event repositories.

In this paper we propose an information integration methodology that facilitates tracking, tracing and information sharing in the supply chain by exploiting Semantic Web standards and linked data technologies. Our approach draws requirements from real-time supply chain business processes in the agri-food sector that are involved in the tracking and tracing of perishable goods. We exploit two information models: The EPCIS Event Model (EEM)[4] based on the EPCIS specification, that enables the sharing and semantic interpretation of event data and CBVVocab[5] its companion ontology based on CBV, for modelling the business context associated with events. We incorporate the models in a distributed and decentralised framework, which facilitates tracking and tracing of goods, potentially in real time, as well as the sharing of information about individual products. We propose the concept of "Linked pedigrees" - linked datasets curated by consuming EPCIS event data, product data and location data, which enable the capture of a variety of tracking and tracing information about products as they move among the various trading partners. In addition, we present "OntoPedigree", an ontology design pattern for the data modelling of pedigrees, which can be specialised and extended to define domain specific or indeed product specific pedigree ontologies.

Linked pedigrees help to overcome a significant limitation of current schemes available for the sharing of information in the supply chain - that of information being available only from partners one up or one down in the supply chain. Dereferencing URIs makes it possible to sequentially traverse the chain of pedigrees exchanged between partners and retrieve complete traceability information, with allowance for adequate access control mechanisms. Linked pedigrees, generated using our approach can be used to derive implicit knowledge about the supply chain, such as shipment delays, inventory shrinkage and out-of-stock situations.

The paper is structured as follows: We begin by motivating the problem and identifying a set of requirements for the fresh fruits and vegetables supply chain. This is also the example scenario, which we use throughout the paper. We then discuss background and related work followed by a brief outline of the EPCIS specification from GS1. EEM and CBVVocab are next presented with their example usage. We then introduce LinkedEPCIS, a Java library and a reference implementation for EEM and CBVVocab. Following this, we outline the concept of Linked Pedigrees and present the content ontology design pattern "OntoPedigree". We demonstrate the generation of linked pedigrees using the LinkedEPCIS Java library. Examples of generation of linked pedigrees for the fresh fruit and vegetables supply chain are next illustrated. We then outline an architecture and a communication protocol for the exchange of linked pedigrees. We evaluate our approach by formalising three representative queries from our requirements analysis and finally we present our conclusions.

THE AGRI-FOOD SECTOR: FRESH FRUIT AND VEGETABLES SUPPLY CHAIN

In this section, we describe the use case, and thus provide a context for the proposed methodology. The requirements of the use case are analysed, and we show the relevance of Semantic Web standards and Linked Data technologies to the agri-food sector. Finally we outline the scenario from the perishable goods supply chain that is used throughout the paper.

Motivation

The agri-food sector is one of the most important economic sectors in Europe and worldwide remains the largest employer, the largest producer of goods, and one of the most important contributors to our environmental footprint as a species. Agricultural commodities are bought and sold via complex networks involving a very large number of heterogeneous actors. The agri-food supply chain involves the movement of agricultural products from the greenhouse, farm or ranch through aggregators and food processors, and then on to retailers and final consumers.

A multitude of stakeholders are participants in the business transactions realised through these networks. One of the key consequences of this number and heterogeneity of actors is the very poor information flow, which exists along the supply chain. Thus for example the farmer might communicate with the wholesaler or food processor but not directly with the retailer. The retailer communicates with the consumer and wholesaler but (typically) few other actors. This is of course even more accentuated in more complex supply chains where food is processed or packaged for longer-term storage. This lack of information flow has been "solved" so far by a combination of government or EC level regulation (food standards, health and safety) and third party certification (organic food certification bodies, GlobalGAP, etc.).

The lack of information has been recognised as a critical issue for a long time in the agri-food sector expressed partly in the need for greater transparency, but also in the importance given to tracking and tracing of foods in the context of health and safety, and in order to both prevent and respond to food emergencies (mad cows disease, and most recently E. Coli). Another major factor is the growing desire on the part of food consumers to know more about their food, a desire for greater food awareness.

The core objective of the methodology and framework presented in this paper is to enhance collaboration and interoperability between the actors in the agri-food, transport and logistics domains with the overall goal of enabling the sharing of data for regulatory purposes, for consumer information, as well as the tracking and tracing of agricultural products.

Requirements Analysis

The design of our framework is guided by certain requirements which have arisen out of previous work in the SmartAgriFood project[6] and reflect those aspects which we believe are essential to address the issues underlying the curation, sharing and reusability of knowledge in the agri-food supply chain. Below we outline some key requirements:

- The domain involves stakeholders with varied levels of capabilities with regards to their ICT infrastructures. The framework should provide data curation techniques capable of aggregating knowledge from a variety of sources such as legacy file-based systems, existing relational databases and sensors. Components that integrate the disparate formats into a uniform representation and that facilitate querying and reasoning over the aggregated knowledge bases are an integral part of such an infrastructure.
- The knowledge aggregated and augmented at strategic nodes within the supply chain graph is made available in a format that enables seamless interoperability and data sharing thus facilitating the optimisation of business processes, and increasing transparency across the supply chain.
- Due to the large number of processes involved, the high volume of information generated and the highly dynamic nature of information flow through the system, the data representation mechanism must be amenable to frequent updates without compromising the performance of supporting applications.
- Domain knowledge should be encapsulated as declarative and interlinked profiles of agricultural products and logistical artifacts such as pallets, containers and shipments, curated in accordance with existing regulatory standards.
- An infrastructure that provides a variety of views of the curated knowledge that is both timely and flexible, depending on the eventual intended use of the data needs to be implemented.

For the purposes of the SmartAgriFood project, we focused on three sub-areas of the agri-food supply chain, to identify use cases and requirements:

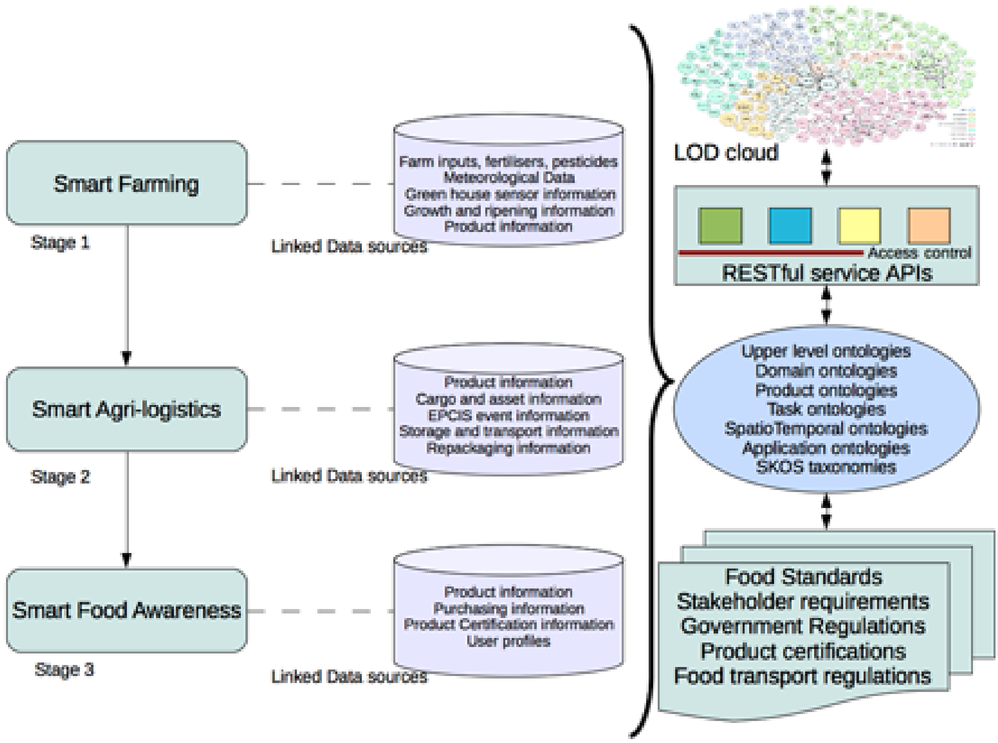
- Smart farming, focusing on precision farming, sensors and traceability;
- Smart agri-logistics, focusing on real-time virtualisation, connectivity and logistics intelligence;
- Smart food awareness, focusing on transparency of data and knowledge representation.

Based on the project's simplification of the agri-food sector, Figure 1 depicts a schematic representation of the contributions that Semantic Web and linked data could bring to each of the segments of the supply chain. As illustrated, at each stage of the chain, datasets can be described using well established, as well as domain specific, ontologies. Besides integrating stage specific and intra-stage datasets, contextually relevant datasets from the LOD[7] cloud can also be linked in, where appropriate.

Stage 1 encompasses the smart farming of agricultural products. It incorporates datasets about location, including geographical information about the farming sites, datasets about species and seeds, datasets about pesticides and fertilisers, datasets about input providers, and datasets about regulatory codes. The linked data would be described using domain specific and spatial ontologies.

Stage 2 is concerned with smart agri-logistics. Datasets and vocabularies related to the agricultural products, transport and logistics artifacts and business transactions between stakeholders would be the key semantic outputs of this stage. Standards such as EPCIS (to which we will return), domain specific ontologies, task and application ontologies for services and product ontologies such as GoodRelations[8] will play a crucial role in modelling knowledge.

Figure 1. Application of SW & LD to the agri-food supply chain



Stage 3 involves smart food awareness i.e. providing more information to end consumers. Datasets and vocabularies related to consumer preferences, carbon footprints of agricultural commodities, food certifications, food regulations and applications governing them would be crucial in describing the information models for this stage.

Upper level ontologies such as SUMO[9] and BFO[10] could be employed as metadata specification for all the three stages. Regulatory policies or domain specific rules can be specified either as part of the vocabulary description using standards such as OWL 2 RL[11] or as rules using RuleML[12], SWRL[13] or RIF[14].

Linked data applications built to serve the stakeholders rely on an unambiguous and contextual representation of the domain knowledge. This is crucial in order to provide reliable and unambiguous answers to data driven questions raised by stakeholders. Below we enumerate a few examples of informal queries that could be potentially raised against the knowledge in these three stages:

- Which products sold in the last 10 days at retailer ABC are vegetables and who are the suppliers and eventual producers?
- For meat product XYZ, identify all ingredients for batch numbers 123 to 679 and specify the source of the ingredients.
- What production parameters were implemented by farmers in the North of Spain for growing tomatoes that were distributed throughout Europe between the months of April and July, and which were rated between 6.5 - 8 by the Spanish and French food consumer organisations?
- List all suppliers of tomatoes within 50km of Barcelona.

- List the locations this tomato was "visible" at between date1 and date2.
- For all products sold between dates/times t1 and t2, and which are potential E. Coli carriers, provide a list of the producers and locations traversed by each product item.

Scenario Description

The following scenario is based on a combination of what actually happens today (Alex Kaloxylos et al., 2011), and what is in the process of occurring with the growth of precision farming, Farm Management Information Systems (Alexandros Kaloxylos et al., 2012), and the wider user of sensors in agricultural production (Ruiz-Garcia, Lunadei, Barreiro, & Robla, 2009; Taylor, K. et al., 2013).

The lifecycle of fresh fruit and vegetables until they reach the end consumer is a complex process because of the number of actors involved and the diverse set of heterogeneous data that is (or that can be potentially) produced. For example, inside a farm, a perishable product such as a tomato could generate large volumes of data (e.g., environmental conditions, date of plantation and harvest, fertilizers, energy used, pesticides used etc). We outline an agri-food supply chain scenario for fresh fruit and vegetables, in particular, tomatoes. The tomato supply chain involves thousands of farmers, hundreds of traders and few retail groups, with information infrastructures in place to record data about the agricultural goods, shipments, assets and cargo. Specifically for fresh fruit and vegetables, it is of utmost importance that up-to-date information about the shelf life of assets as well as the quality conditions under which the cargo is shipped and stored is transparently made available to key stakeholders, potentially in real time.

Figure 2 shows a generalised food chain scenario with a reduced level of complexity. This scenario covers 90% of the general supply scenarios for fresh food products[15].

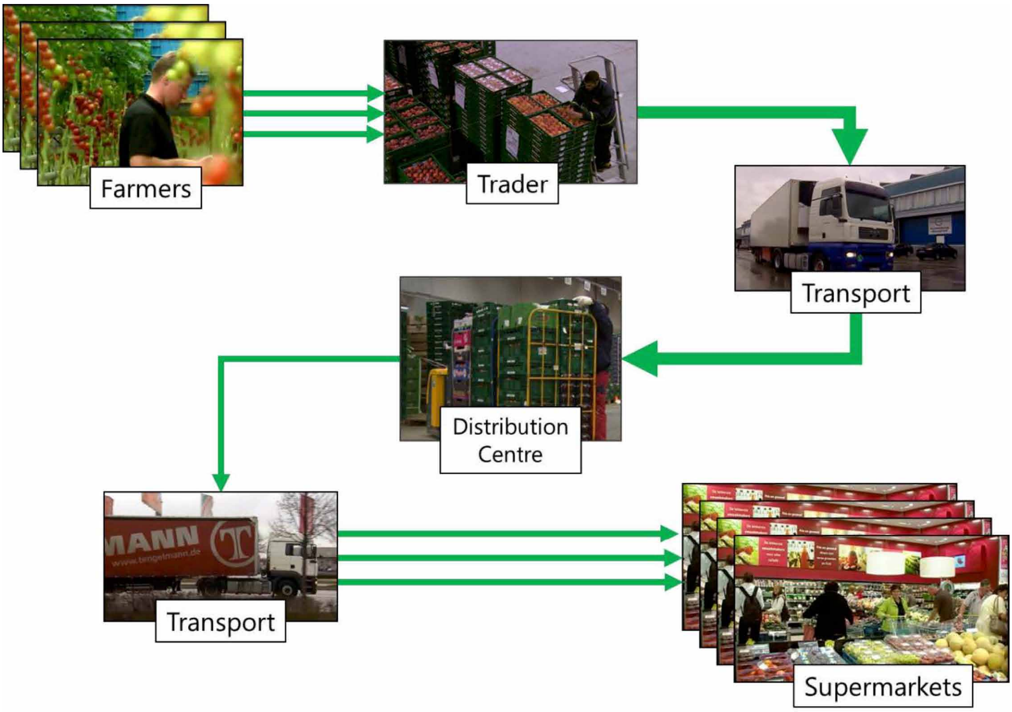
Single trays are packed at agricultural production and are aggregated to larger product batches (pallets) at trader or bundling companies. One pallet can include different trays with the same product of a specific cultivar (e.g. tomatoes of the type "Cherry tomatoes") from different suppliers. At the distribution centers these pallets are disaggregated and re-aggregated into dollies that contain single crates with different products for a specific retail outlet as illustrated in Figure 3.

In the retail outlets these single crates are placed in the fruit and vegetable section. Ever greater use is being made of returnable packaging consisting of rigid or foldable trays (such as those provided by Europool System) where plastic crates/trays are filled on farm and are used throughout the supply chain system.

In order to exemplify the application of the EEM and CBVVocab information models for the generation of linked pedigrees as proposed further in the paper, we consider the following (hypothetical) tomato supply chain trading partners:

- *Franz farmer* and *Bob farmer* produce agricultural commodities (fruits, vegetables, flowers). In particular, they specialise in growing tomatoes. The packaging of tomatoes is done in punnets (or trays), each of which are tagged with a RFID label. Shipment of tomatoes to downstream partners is done in crates, each of which is tagged with a RFID label.
- *Joe trader* is a tomato trader and sources tomatoes from national and international agricultural productions. Joe trader bundles tomatoes procured from multiple farmers to larger product batches before dispatching them to distribution centers.
- *Freshfoods Inc.* sources tomatoes from multiple traders and splits up large product batches to smaller batches for distribution to retail supermarkets.
- *Orchards* is a supermarket that receives fresh produce from distribution centers such as Freshfoods Inc.
- Transport service providers execute transport processes and transport products from different actors to their destination in the chain.

Figure 2. Generalised agri-food chain scenario for tomatoes



The use of returnable packaging items in the supply chain reduces some effort i.e. tomatoes do not need to be repacked, but the number of actors and possible aggregations/splits remains just as complex. One of the main advantages of returnable packaging is that the addition of RFID to such trays and pallets increases the tracking and tracing potential within the supply chain (Reiche, R. and Lehmann, R. J. and Fritz, M. and Schiefer, G., 2011).

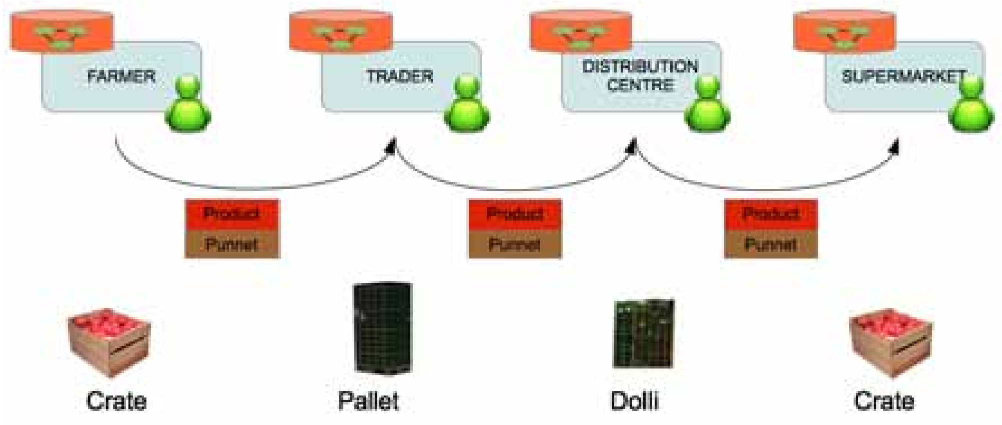
The supply chain scenario and associated activities outlined above give rise to various EPCIS event types, which are recorded via RFID/barcode readers installed at various strategic points in the business location. Event-based linked pedigrees as proposed in this paper provide the information abstraction required to record and exchange tracking and tracing information between the various partners in the sup- ply chain.

BACKGROUND AND RELATED WORK

In this section we present a review of a) related work in the domain of EPCIS events, b) the notion of pedigrees as used in the pharmaceuticals sector and c) a brief account of supply chain information integration, tracking and tracing.

In (Ilic, Alexander and Andersen, Thomas and Michahelles, Florian, 2009), the authors present a supply chain visualisation tool for the analysis of EPCIS event data. Several kinds of scenarios are discussed such as ensuring pair-wise shipping and receiving confirmation, maintaining dwell-time consistency so that the time difference between two events for objects such as perishable goods is under a certain threshold. A map-based interface is provided for the identification of locations with specific supply chain issues determined by a rule-based analysis.

Figure 3. Flow of products in the tomato Supply Chain



In (Nguyen, Tuyen and Lee, Young-Koo and Jeong, Byeong-Soo and Lee, Sungyoung, 2007) a data model and algorithm for managing and querying event data has been proposed. A virtual object layer is introduced on top of relational databases. The data model is illustrated as an extended entity relationship diagram and is close in spirit to EEM as proposed in this paper. A critical limitation of this model is that it is overlayed on top of relational databases and is not available in a form that can be shared and reused between organisations as linked data. Further, EPCIS vocabularies are stored in a large number of tables thereby rendering the relational layer bulky and cumbersome. In contrast, EEM provides a lightweight ontological model for describing the semantic relationships between EPCIS entities.

In (Turchi, S. and Ciofi, L. and Paganelli, F. and Pirri, F. and Giuli, D., 2012) the authors propose to use the InterDataNet (IDN)[16] framework for applying a RESTful architecture and linked data principles to facilitate the sharing of EPCIS data. An IDN document is viewed as a graph with nodes that reference data from various sources. Each node is referenced using an HTTP URI. EPC-data originating from EPCIS events is represented as IDN documents. RESTful services referred to as IDN-Compliant applications provide the capture and query interface implementations. The proposed approach suffers from several critical limitations. No reusable and shared data model is provided that can serve as the metadata layer for representing and reasoning over EPC-data. The encapsulation of information as IDN documents imposes an additional layer, which may significantly affect performance of querying applications.

The Fosstrack[17] EPCIS modules provide a set of implementations for the EPCIS 1.0.1 specification. The implementations include an EPCglobal-certified EPCIS Repository, Query and Capture clients and a Web adapter for REST Web services. A limitation of the existing Fosstrack Java capture library is that any attributes added as an extension element to the existing schema cannot be directly captured and queried for through the interfaces provided in the API. The LinkedEPCIS library proposed in this paper, below, overcomes this limitation by virtue of the underlying linked data curation methodology that facilitates the addition of new triples to the dataset.

Conventionally, pedigrees can be paper based or documents exchanged electronically (e-pedigree). The use of pedigrees for tracking and tracing commodities is most widely prevalent in the pharmaceutical industry[18]. Pedigree or electronic pedigree (e-pedigree) is an audit trail that records the path and ownership of a drug as it moves through the supply chain, in which each stakeholder involved in the manufacture or distribution of the drug adds to the pedigree. The Pedigree standard[19], ratified by EPCglobal, provides an XML schema for the description of the life history of a product across arbitrary supply chains. Recently the concept of an "Event-based Pedigree"[20] has been proposed that utilises EPCglobal's EPCIS specification for capturing events in the supply chain and generating pedigrees based on a relevant subset of the captured events. The generation of linked pedigrees as presented in this paper, builds on the event-based pedigree approach.

The importance of tracking and tracing in the food supply chain has been identified by many authors (Regattieri, Gamberi, & Manzini, 2007; Martini, Mietzsch, Giannerini, Papaekonomou, & Kunisch, 2010; Främling et al., 2013). Closely related to this has been recognition of the importance of transparency across the supply chain, i.e. the uninterrupted flow of information along the supply chain so as to enable proper functioning without unintended crises (Tim Verwaart et. al., 2013). A slightly different perspective on this is provided by the term "visibility" but which nonetheless refers to the availability of information about a product in the supply chain (Joris Hulstijn and Sietse Overbeek and Huib Aldewereld and Rob Christiaanse, 2012). The use of Semantic Web technologies for capturing and managing data across the supply chain was first proposed in (Brewster, Glaser, & Haughton, 2005) although the focus was on the environmental impact of food in the organic food supply chain. In (He, Tan, Lee, & Li, 2009) the authors present a solution that utilises both RFID and GPS for tracking and tracing of international shipments. In (Ruiz-Garcia, Steinberger, & Rothmund, 2010), the authors present a prototype implementation for tracking and tracing batches of agricultural products based on the use of a number of web standards (OGC for geospatial data), but the architecture is centralised and there is no reference to GS1 standards. The CASSANDRA project[21] proposes the concept of a virtual data pipeline that connects entities and gathers and distributes data according to predefined conditions. In (Hofman, 2011), the authors present a contextual architecture for making supply chain data available to applications designed for customs authorities.

THE EPCIS SPECIFICATION: A BRIEF OVERVIEW

The EPCIS specification is an EPCglobal standard that supports a detailed representation of the location and state of a product as it moves between organisational boundaries and provides for sharing this information between entities or partners, in a technology-supplier independent way. The standard is data carrier neutral and can be used to exchange data found from RFID tags, barcodes and other data carriers.

An Electronic Product Code (EPC)[22] is a universal identifier that gives a unique, serialised identity to a specific physical object. In most instances, EPCs are encoded on barcodes or RFID tags which can be used to track all kinds of objects including: trade items, fixed assets, documents, or reusable transport items. EPCIS identifiers for events, products and locations are currently represented using URNs. Various formats for URNs that can be used for identifying the EPCs have been prescribed in the GS1 EPC Tag Data Standard[23]. A typical identifier for classes of products is GTIN (Global Trade Item Number)[24]. An individual serialized item on the other hand is identified using SGTIN (Serialized Global Trade Item Number), e.g., two cases of the same product will have the same GTIN, but individual products in the cases will have different SGTINs. For example, a SGTIN is represented as shown below:

urn:epc:id:sgtin:0614141.112345.400

Similar identifiers[25] are prescribed for RFID read points and business locations using SGLN, (Serialized Global Location Number).

The EPCIS specification defines two kinds of data: event data and master data. Event data arises in the course of carrying out business processes, it grows over time and is captured through the EPCIS capture interface and made available for query through the EPCIS query interfaces. Some examples of event data that can be represented using the EPCIS specification are illustrated below:

- *At time T, the association of the following case tags to the following pallet tag was created at palletizer #3, to fulfill order #1234.*
- *Between the time the case crossed the first beam and the second beam at location L, the following tag was read.*
- *At Time T, Object X was observed at Location L.*

Master data is additional data that provides the necessary context for interpreting the event data. An example of master data is: *Location L refers to the distribution center located at 123 Elm Street, Anytown, US.*

The EPCIS standard defines an abstract generic event and several specialised physical event types, arising from supply chain activity across a wide variety of industries.

- *EPCISEvent* represents the generic EPCIS event.
- *ObjectEvent* represents an event that occurred as a result of some action on one or more entities denoted by EPCs.
- *AggregationEvent* represents an event that happened to one or more EPC-denoted entities that are physically aggregated (constrained to be in the same place at the same time, as when crates are aggregated to a pallet).
- *QuantityEvent*[26] represents an event concerned with a specific number of objects all having the same GTIN, but where the individual instances are not identified. For example a quantity event could report that an event happened to 200 boxes of widgets, without identifying specifically which boxes were involved.
- *TransactionEvent* represents an event in which one or more entities denoted by EPCs become associated or disassociated with one or more identified business transactions.
- *TransformationEvent* represents an event in which one or more physical or digital objects are fully or partially consumed as inputs and one or more objects are produced as outputs.

Each EPCIS event, recorded and registered against RFID tagged or barcoded artifacts has four information dimensions. It encapsulates the "what", "when", "where" and "why" of these artifacts at the barcode or RFID scan point.

THE EEM AND CBVVOCAB ONTOLOGIES

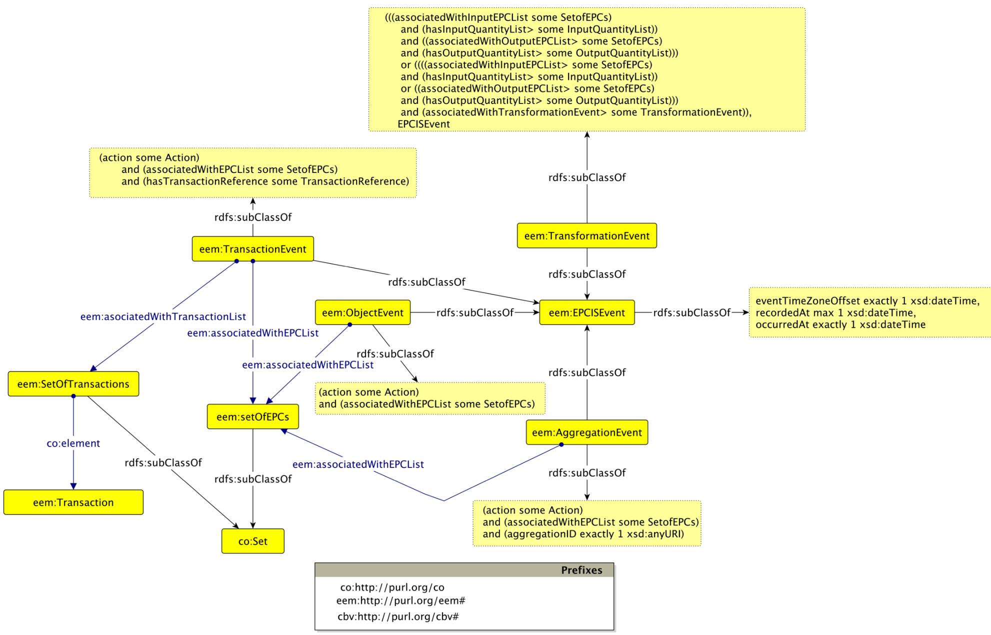
In (Monika Solanki and Christopher Brewster, 2013b) we presented EEM, an OWL DL ontology that defines EPCIS events, their attributes and relationships. It enables the sharing and semantic interpretation of EPCIS events on the Web of data. A detailed explanation on the rationale and modelling decisions governing EEM can be found in (Monika Solanki and Christopher Brewster, 2013b). In this section, for the purpose of completeness, we present a brief overview of the most important classes and properties in EEM and its companion ontology CBVVocab.

EEM defines a generic event class and specialised event classes corresponding to the various event types. EPCISEvent is the parent class of all events. ObjectEvent, AggregationEvent, TransformationEvent and TransactionEvent are specialised classes of EPCISEvent. Figure 4 illustrates the event classes in EEM.

The class EPC provides a placeholder for EPCs represented using various URI schemes. The list of EPCs is represented by SetOfEPCs, specialising from Set[27].

The class Action denotes the activity undertaken on objects tagged with electronic product codes and represented by SetOfEPCs. The set of actions[28] associated with an event are asserted with the individuals ADD, OBSERVE and DELETE.

Figure 4. EPCIS event classes as represented in EEM



An aggregation event is associated with a unique aggregation URI, while a transformation event is associated with specific quantities of inputs and outputs.

An event takes places within a certain business context that may or may not involve transactions. The class Transaction encapsulates references to transactions and their types. The collection class SetOfTransactions represents the set of transactions associated with an event.

The business information associated with EPCIS events includes identifiers for the EPC readers, business locations, and location of the EPC readers that recorded the events, business steps or processes that generate the events and the state or disposition of the tagged objects.

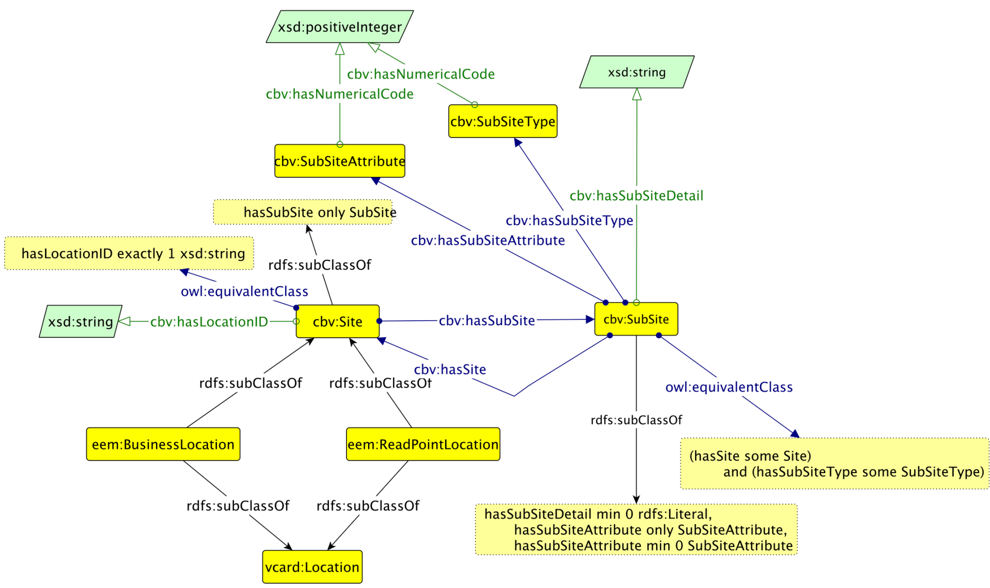
The BusinessLocation and ReadPointLocation classes capture physical location details and specialise from the Location class defined in the vcard[29] vocabulary. The EPCReader class represents readers with physical and logical identifiers.

A companion standard to the EPCIS standard is the CBV standard. The EPCIS data model includes business step, disposition and business transaction type fields that may be included in any EPCIS event, but does not define the specific identifiers that can populate those fields. CBV defines specific values that may be used to populate each of those fields. EEM defines classes BusinessStep, BusinessTransactionType and Disposition corresponding to the fields. Individual assertions capture the field values. In order to keep the EPCIS elements distinct from the CBV elements, we provide these definitions in a separate ontology, CBVVocab.

Representing Physical Locations

As highlighted earlier, the EPCIS specification defines two kinds of data: *event data* and *master data*. The description of physical locations is a part of the master data. The CBV standard defines master data attributes for capturing information associated with a physical location

Figure 5. Ontological representation of Location Master Data



that may be used to describe locations for an EPCIS event.

CBV identifies two levels of granularity for locations - *Site* and *Sub-site*, both of which may be used to describe a *ReadPoint* or a *BusinessPoint* location. A Site denotes a physical location such as a distribution center or a retail store, whereas a Sub-site denotes a physical location such as a back room or the sales room of a retail store. CBV specifies attributes that may be used to further describe a location. Figure 5 illustrates the entities in the ontological model for describing locations.

Site and SubSite denote the location types for ReadPointLocation and BusinessPointLocation entities and are identified using the hasLocationID datatype property. Individual assertions of SubSiteType and SubSiteAttribute entities are used to associate the sub-site types and attributes with the SubSite entity. The datatype property hasNumericalCode ranges over the numerical codes assigned to sub-site types and attributes in the CBV specification.

Figure 6 illustrates an example of a ReadPointLocation in the Turtle serialisation.

Properties

EEM defines several kinds of properties for events, to capture the relationship between entities based on the four information dimensions.

Event Specific Properties

EEM defines properties relating events to their business context. While many properties are common among the four specific event types, some are specific to certain events. For example, the hasAggregationID property is defined only for the AggregationEvent. While hasTransactionReference is required to be asserted for a TransactionEvent, it is optional for the other event types.

Besides the implicit relationships described in the EPCIS specification, EEM defines some additional relationships over events. We believe the proposed extensions will facilitate EPCIS event identification, querying and rule based reasoning. A crucial attribute missing from the EPCIS specification is the eventID for an event. This is especially required because of

Figure 6. Representing a ReadPointLocation

```
@prefix mde:<http://company1.com/data/epcis/location/sgln/>
@prefix add:<http://company1.com/data/epcis/location/address/>
@prefix vcard:<http://www.w3.org/2006/vcard/ns#>
@prefix wgs84:<http://www.w3.org/2003/01/geo/wgs84_pos#>
mde:ReadPointLoc1 a eem:ReadPointLocation;
                  a vcard:Location;
                  eem:LocationType mde:0614141.00300.0.
                  wsg84:latitude "66.66"^^xsd:decimal;
                  wsg84:longitude "-533"^^xsd:decimal;
                  vcard:adr  add:MyAddress1.

add:MyAddress1 a vcard:Address;
               vcard:country-name "Germany"^^xsd:string;
               vcard:street-address "abc-street"^^xsd:string;
               vcard:postal-code "GER-123"^^xsd:string.

mde:0614141.00300.0 rdf:type cbv:Site;
                    cbv:hasLocationID "0614141003006"^^xsd:string;
                    cbv:hasSubSite mde:ReadPointSubSite1.

mde:0614141.00300.1 rdf:type  cbv:SubSite;
                     cbv:hasSubSiteType  cbv:PackagingArea;
                     cbv:hasSubSiteAttribute cbv:Groceries;
                     cbv:hasSubSiteAttribute cbv:Frozen;
                     cbv:hasSubSiteDetail "shelf 456 in the
                       backroom of store 123"^^xsd:string
```

the large volume of event data that is gen- erated and needs to be interlinked and queried. A systematic identification system assigns every event a unique eventID. This can then be used to construct URIs for events in order to publish event data as linked data and link event data with master data. EEM further extends EPCIS semantics with properties used to define rules over event actions.

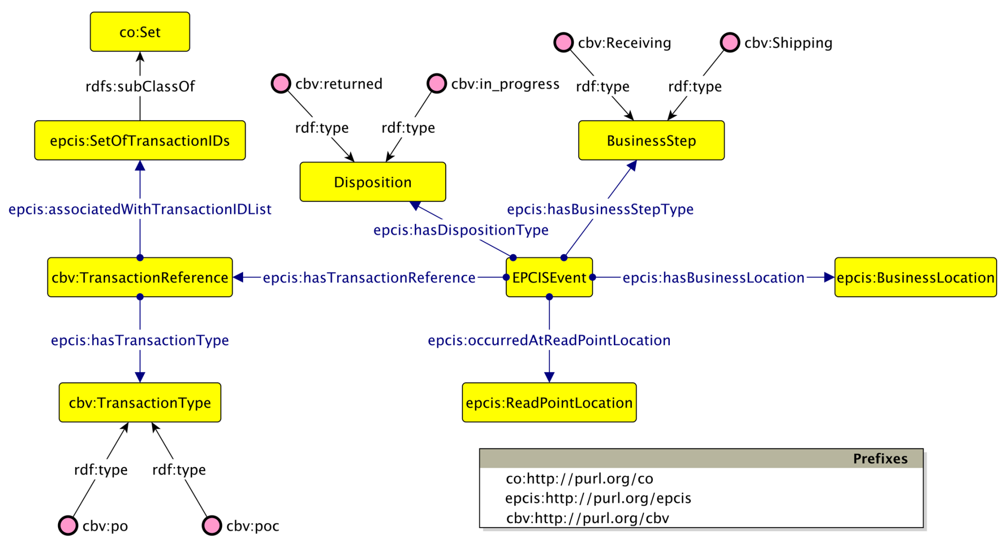
Temporal Properties

An EPCIS event is associated with three types of timing properties: eventOccurredAtTime signifies the date and time at which the EPCIS capturing applications asserts the event occurred, eventRecordedAtTime captures the date and time at which this event was recorded by an EPCIS Repository (optional). Additional business context is provided through the property eventTimeZoneOffset, the time zone offset in effect at the time and place the event occurred.

Location Properties

The hasBusinessLocation and hasReadPointLocation object properties connect the business and read point locations respectively to an event. A business location or a read point itself is identified using the hasLocationID data type property with the property range being xsd:anyURI.

Figure 7. Business context entities, relationships and representative individuals



Business Context Properties

The BusinessStep and Disposition entities relate to an event through the hasBuisnessStepID and hasDispositionType property respectively. Individual assertions for these entities are provided in the CBVVocab ontology and are used to populate the range values for the properties. Every Transaction entity is related to a TransactionType entity through the hasTransactionType relationship. Values for transaction types are provided by the CBV standard and asserted in the CBV ontology. Figure 7 provides an illustration of the entities, relationships and some representative individuals for the entities.

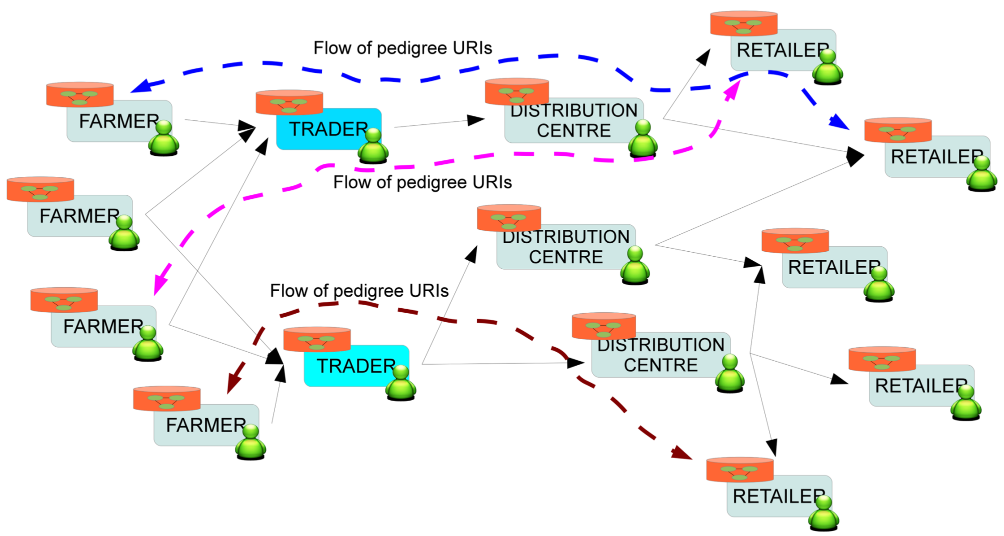
LINKED PEDIGREES

As explained earlier in the paper, a pedigree is a record that traces the ownership and transactions of a product as it moves among various trading partners. Conventionally, pedigrees can be paper based or documents exchanged electronically (e-pedigree)[30]. Analogous to the notion of pedigree, in (Monika Solanki and Christopher Brewster, 2013a) we briefly proposed the concept of "linked pedigrees".

A linked pedigree is a dataset, identified using an HTTP URI, described and accessed using linked data principles and represented in RDF. Linked pedigrees encapsulate the knowledge required to trace and track products in the supply chain on a Web scale as well as capture a variety of other types of relevant data. They aim to bridge the interoperability gap prevalent in existing supply chain models. They provide a domain independent data model for the sharing of knowledge among Semantic Web enabled systems deployed for the tracking, tracing and data capture concerning commodities as they physically flow downstream in the supply chain. The pedigree also makes it possible for data to flow upstream, something, which is currently severely limited in the agri-food sector.

When applied to the domain of agri-food, a linked pedigree defines a dataset containing certified information about each distribution of a unit, i.e., individual goods, packaged cartons, crates, pallet loads, consignments and complete shipments of agri-food. Considering the supply chain as a graph, stakeholders at every node and edge of the graph define a linked pedigree. Figure 8 illustrates the flow of pedigrees, upstream and downstream of the supply chain.

Figure 8. Flow of pedigrees among stakeholders participating in three different supply chains



Pedigree definitions include URIs for product, transaction and consignment information curated in the stakeholder's knowledge base. Additionally, apart from the pedigree initiated and created by the first stakeholder in the supply chain, e.g., the farmer, all other linked pedigrees include URIs to the pedigree datasets for the stakeholders in the immediate upstream or downstream of the supply chain.

While the notion of pedigrees is a general one, they are usually tailored to address the needs of specific domains or industries. Rather than define a generic ontology for pedigrees, in this paper we propose a content ontology design pattern "OntoPedigree" that can be incorporated while building domain specific pedigree ontologies. Content ontology (CP) design patterns (Gangemi, 2005; Presutti, Valentina and Gangemi, Aldo, 2008) are reusable ontological artifacts that aim to provide solutions to recurrent, domain specific modelling problems (Staab & Studer, 2009). A repository of content ontology design patterns[31] can be found at the Ontology Design Patterns Portal. In the following sections, we present the pattern intent, supporting competency questions and the conceptual entities defining OntoPedigree.

Intent

OntoPedigree[32] provides a minimalistic abstraction which defines conceptual entities for the modelling of knowledge required to enhance data integration and visibility in a supply chain. The pattern can be specialised to define domain specific pedigrees.

Competency Questions

The pattern aims to provide answers to the following competency questions:

- Who is the creator of the pedigree?
- What is the supply chain creation status of a given pedigree?
- Which are the business transactions recorded against a particular consignment?
- What are the events associated with pedigrees created between dates X and Y?
- Which products have been shipped together?
- Which other pedigrees are included in the received pedigree?

Conceptual Entities

Some of the key concepts and relationships encapsulating the data model defined by the pattern are:

- Pedigree: A concept defining the notion of a pedigree as an entity capturing certified information for a stakeholder in the supply chain.
- PedigreeCreationStatus: An entity defining the status of creation of a pedigree. Valid instances include InitialPedigree, IntermediatePedigree or FinalPedigree.
- Creator: A Dublin Core[33] annotation entity linking the pedigree to its creator.
- Date: A Dublin Core annotation entity linking the pedigree to the date of its creation .
- hasSerialNumber: An entity linking the pedigree to its uniquely assigned serial number.
- hasTransactionInfo: An entity defining the relationship between the pedigree and the conceptual entity capturing transaction information related to the sale, movement, return or transfer of goods.
- hasProductInfo: An entity defining the relationship between the pedigree and the conceptual entity capturing product information.
- hasConsignmentInfo: An entity defining the relationship between the pedigree and the conceptual entity capturing consignment information.
- hasReceivedPedigree: An entity defining the relationship between the pedigree and any receiving pedigrees from stakeholders in the upstream or downstream of one or more participating supply chains. Examples of such receiving pedigrees include,
    - A pedigree defined by a repacking trader, who combines goods from several consignments received from upstream stakeholders.
    - A pedigree defined by a downstream retailer who returns goods to an upstream trader.

Figure 9 illustrates the graphical representation of OntoPedigree. It depicts the entities defined for the pattern and their relationships with entities from PROV-O, EEM and GoodRelations.

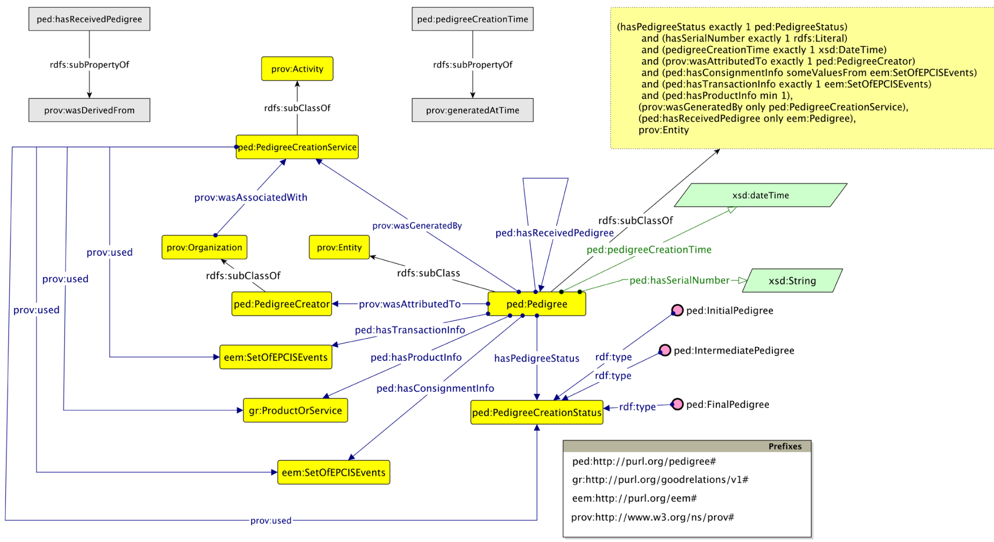
The axiomatisation of the concept Pedigree is illustrated below:

```
Class: ped:Pedigree
SubClassOf:
(hasPedigreeStatus exactly 1
ped:PedigreeStatus)
  and (hasSerialNumber exactly
1 rdfs:Literal)
  and (pedigreeCreationTime
exactly 1 xsd:DateTime)
  and (prov:wasAttributedTo
exactly 1 ped:PedigreeCreator)
  and (ped:hasConsignmentInfo
some eem:SetOfEPCISEvents)
  and (ped:hasTransactionInfo
exactly 1
eem:SetOfEPCISEvents)
  and (ped:hasProductInfo min
1),
 (prov:wasGeneratedBy only
ped:PedigreeCreationService),
 (ped:hasReceivedPedigree only
eem:Pedigree),
  prov:Entity
```

EXCHANGING LINKED PEDIGREES

We propose an open, scalable and decentralised architecture for enabling real time data visibility in supply chains. We assume that information management systems with dedicated Web service interfaces are in place for the capture and visualisation of the relevant data. Every actor in the supply chain manages its own datastore, i.e., there is no central repository for holding integrated datasets. We also assume that the services are equipped with interfaces through which semantically enriched queries, e.g., SPARQL queries can be executed and results can be obtained in one or more of the several

Figure 9. Graphical Representation of OntoPedigree

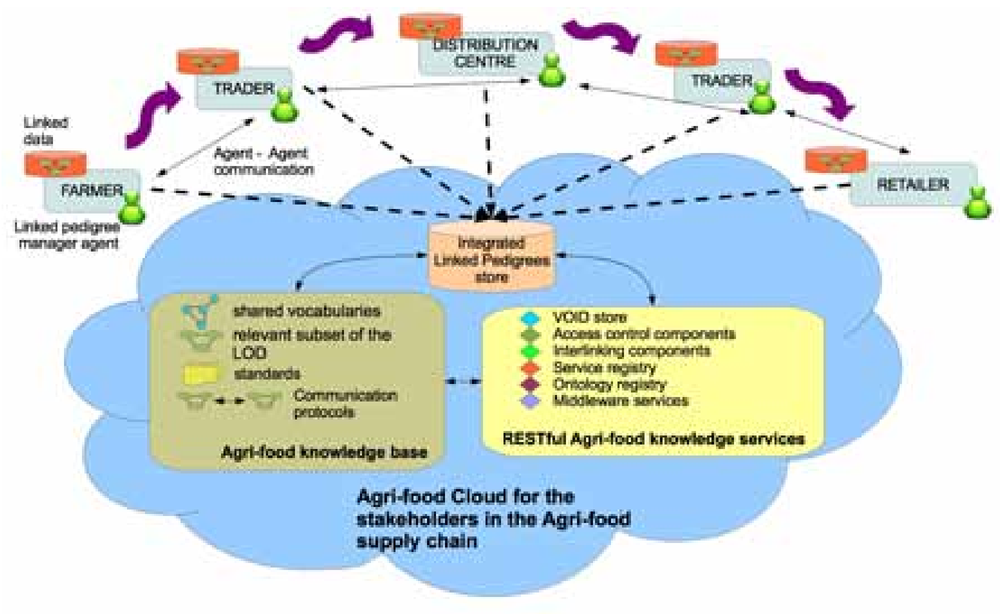


standard formats such as RDF/XML, Turtle, JSON or JSON-LD.

Linked pedigrees can be obtained via a push model, i.e., an upstream actor sends the URI of the pedigree to a downstream actor or a pull model, i.e., a downstream actor requests pedigree information from an upstream actor. In this paper we assume a pull model for retrieving pedigrees. The high level architecture as exemplified for the agri-food supply chain is illustrated in Figure 10. Shared data models, vocabularies, Web based and mobile application components are provided as cloud-based services. Below we provide an account of key components comprising the architecture:

- **Linked Pedigree Manager agent**: The pedigree manager agent is the central component that interfaces between the EPCIS event store and external systems. Some of the responsibilities of the agent include:
  - Generating pedigrees using the Linked Pedigree Generator component.
  - Querying linked pedigrees using REST Web service interfaces from upstream/downstream stakeholders and locally corroborating the electronic information recorded on received physical goods against the query results. Besides the pedigree being checked against the goods received or sent by supply chain partners, inspection/checking of pedigrees may be routinely undertaken by third parties. The manager agent is responsible for mediating between pedigree checking requests and event data held in the event store.
  - On receiving a request to provide a pedigree against a consignment of physical goods, generating pedigrees on the fly from the knowledge curated in the event data stores of the stakeholder. This requires assigning the pedigree a URI and including outgoing links to external datasets. Pedigrees may also be curated by the agent before goods are shipped.
- **Knowledge Services**: The management, update, query and access of the knowledge repository is facilitated via a set of Web services with RESTful interfaces.

Figure 10. A high level architecture for the exchange of linked pedigrees in agri-food supply chain



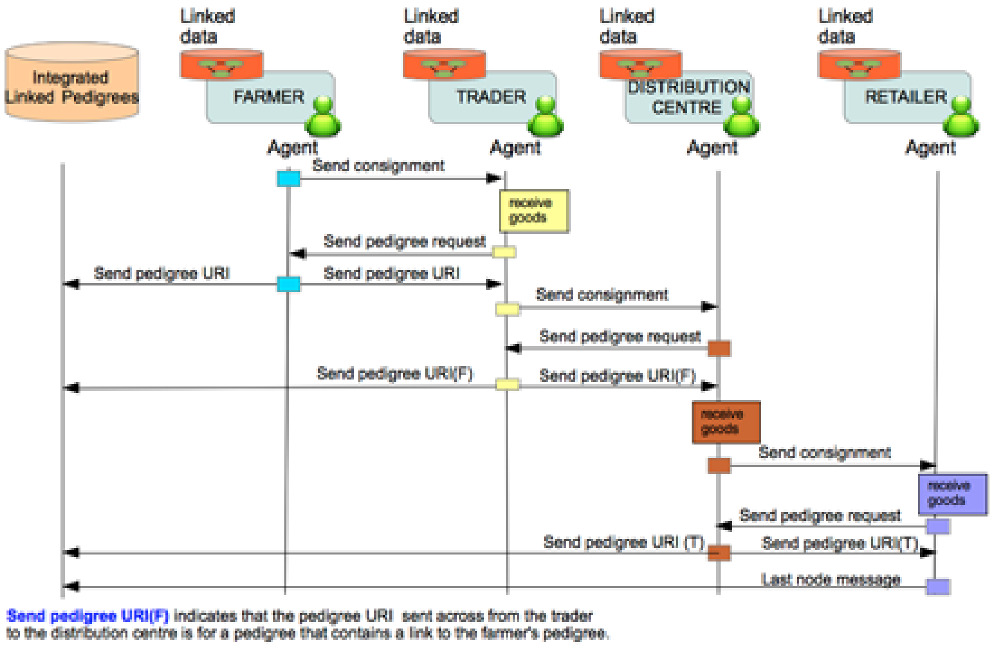
Be- sides these functionalities, the service suite provides standalone components or "apps" that can be integrated within the IT infrastructure for individual stakeholders. Examples include components implementing access control, dataset interlinking middleware services and dataset metadata (VOID[36]) stores.

- **Integrated Linked Pedigree Store**: The integrated linked pedigree store provides an overarching, governing service, thereby giving an end-to-end context to the pedigree transactions taking place within individual supply chains. It can be observed in Figure 10 that the architecture is decentralised, i.e., there is no central datastore. The pedigrees in their definition include the URIs of the pedigree received from the upstream or downstream stakeholders. Linked pedigrees can be sequentially traversed, to eventually construct an ordered chain of pedigrees in the supply chain, by dereferencing the URI corresponding to the hasReceivedPedigree relationship for every actor. However access control restrictions mean that it may not be possible for stakeholders themselves to obtain complete information related to products and consignments from every other stakeholder. The integrated linked pedigree store mitigates this limitation, should the need arise, by acting as the governing authority and providing a service that can facilitate the end-to-end dereferencing of linked pedigrees in the supply chain.

Additionally, the store can also provide validation services for establishing the conformance of the information recorded on received physical goods against the results returned by querying the received pedigree URI, should a stakeholder agent not be available or equipped to perform the validation. Figure 11 illustrates an example of the protocol as applied to the agri-food supply chain.

- Each pedigree is assigned a unique URI following domain specific URI design principles (Public Sector Information Domain

Figure 11. An example of a linked pedigree communication protocol among actors in the agri-food supply chain



of the CTO Council's cross Government Enterprise Architecture., 2009).

- The EPCIS implementation at the trader's end records the receipt of the goods as an event. The trader's agent requests pedigree information about the consignment from the farmer.
- The farmer agent creates the pedigree dataset with the assigned URI and adds relevant pedigree information to it. The information is represented as a combination of literal values and URIs representing local and global entities. Pedigrees can be generated offline and curated in the triple store. Alternatively they can be generated on the fly when a request is received.
- The URI for the pedigree is then sent to the trader as well as to the linked pedigree store. It is assumed that digital certification and authentication procedures for messages exchanged between stakeholders as well as the linked pedigree store are in place.
- The messages containing pedigree URIs are electronically authenticated by the trader agent and the URIs are dereferenced.
- The trader dispatches the goods to the next stakeholder in the chain, i.e., the distribution center.
- On receiving a pedigree request from the distribution center, the trader includes in its pedigree, the URI to the pedigree it has received from the farmer.
- The process of creating the pedigree and adding the URI of the receiving pedigree is undertaken at every link/node in the supply chain where new information is generated.
- At every link/node, the pedigree URI is also sent to the linked pedigree store, which is responsible for keeping an account of all the pedigrees, exchanged between stakeholders participating in a specific supply chain transaction.
- At the last point in the supply chain, i.e., the retailer, the final pedigree informa-

tion, along with an end-of-supply-chain message is created and forwarded to the linked pedigree store.
- On receiving the final pedigree message, the linked pedigree store consolidates and contextualises the pedigrees received for that specific instance of the supply chain. It generates a linked data instance that aggregates all the pedigrees involved in the supply chain and seals the aggregated linked pedigree dataset for future references.

A Note on Privacy, Access Control and Non-Repudiation

The decentralised nature of the architecture and the message oriented communication protocol make security and privacy important considerations. Due to the inherent nature of tracking and tracing data being commercially sensitive, it is assumed message exchange will be appropriately secured via digital signatures and deferencing of pedigree URIs and event data URIs corresponding to various elements of a pedigree will be controlled via access controlled restrictions. We do not address these issues further in the paper.

THE LINKEDEPCIS LIBRARY

In this section we present an API framework for conveniently incorporating the EEM model in EPCIS capture and query applications. LinkedEPCIS[34] is a Java library for capturing EPCIS events as linked data. The type hierarchy in LinkedEPCIS is based on the entities defined in the EEM data model. Every event curated in our event data triple store, using the library, is systematically assigned an HTTP URI. LinkedEPCIS has been built over the Sesame framework[35]. Sesame is an open source Java framework for storage and querying of RDF data. Development of the library has been based on the EPCIS standard specification (Version 1.1 of May, 2014). The main objectives of LinkedEPCIS are:

- Provide a linked data oriented reference implementation for the EPCIS specification.
- Provide a set of Java classes that generate linked data from EPCIS events.
- Provide a set of “templated” SPARQL queries for querying over the linked data.
- Provide examples of RESTful Web services that can be exploited by EPCIS implementing organisations to deploy their own set of services.

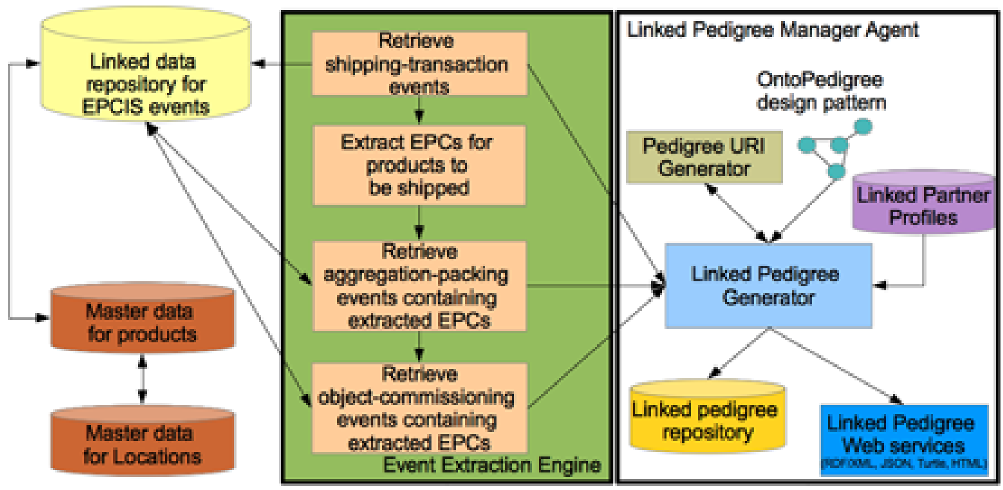
The most significant classes in the LinkedEPCIS library are EPCISEvent and EPCISCommon. EPCISEvent encapsulates the attributes and operations common to all EPCIS event types. These operations include setting various attributes such as: temporal properties, disposition, action and transaction, assigning a reader to the event, setting location data, defining extensions to events and finally persisting the event as linked data. EPCISCommon provides a set of operations for the internal generation and manipulation of the linked data model.

Central to the data model generated through the LinkedEPCIS library is the Graph interface from the Sesame API. LinkedEPCIS records data about events as triples/statements and attaches them to a Graph, which can be persisted as files or dumped in a dedicated EPCIS events triple store. Besides the attributes for events predefined in the EPCIS specification, extensions are supported by retrieving the current Graph and attaching new triples.

GENERATING LINKED PEDIGREES USING LINKEDEPCIS

A real time application of the LinkedEPCIS library in the supply chain domain is a service for generating linked pedigrees associated with shipped consignments. Supply chain trading partners can maintain and share information about product movement throughout the life of the product by exchanging linked pedigrees. The technology also makes it possible to exchange

Figure 12. LinkedPedigree application architecture



a great deal more information about products in an integrated fashion.

The LinkedPedigree Web service application provides a linked data representation of pedigrees in various serialisation formats such as RDF/XML, JSON, Turtle and NTriples. Pedigrees are assigned URIs, which can be forwarded to relevant supply chain trading partners.

Using the LinkedEPCIS library's event capture API, EPCIS events are captured and integrated with relevant master data, before curating them in a triple store. The LinkedEPCIS query library is then used for systematically querying the linked data to generate linked pedigrees. Figure 12 illustrates the architecture underlying the LinkedPedigree application.

Some of the key components of the architecture are:

- **Linked Data Repositories:** The repositories hold linked datasets for the EPCIS events captured by the trading partner for its supply chain operations, master data about the products and locations and profiles of the trading partners. In most cases the event data will be integrated with master data about the products to be shipped and locations. The event and master data repositories may be distributed. Partner profiles will be integrated during the runtime generation of linked pedigrees.
- **Event Extraction Engine:** The engine is responsible for extracting events from the linked EPCIS event repository using the LinkedEPCIS API or executing bespoke SPARQL queries at the SPARQL endpoint.
- **Linked Pedigree Generator:** The linked pedigree generator combines the extracted event data and generates a linked pedigree dataset, integrated with the profiles of the sending and receiving trading partners. It automates the generation of linked pedigrees by querying the triple store for event URIs corresponding to the commissioning, shipping, receiving and transaction events[36]. Additionally, URIs for the product master data and location master data are retrieved. The pedigree status is dynamically set, based on the supply chain node at which the pedigree is generated.

The application generates pedigrees following the protocol outlined below:

- Before consignments are shipped to supply chain partners, data for the shipping events related to individual consignments

Figure 13. Generating Linked Pedigrees using EPCIS event URIs

Pedigree Component	Linking relationship	Resource identifier
Product information	`hasProductInfo`	Product data URIs Serialised product data URIs
Consignment information	`hasConsignmentInfo`	Commissioning events - Object event/Aggregation event URIs
Transaction information	`hasTransactionInfo`	Shipping events - Transaction event URIs

are retrieved from the LinkedEPCIS triple store. It is worth noting that linked data for the shipping events contain triples asserting transaction information for the consignments. Shipping events are therefore usually registered as transaction events in the EPCIS event repository.

- The transaction events datasets are interrogated to ex- tract EPCs of the products scheduled to be shipped.
- The triple store is queried for aggregation events, which include the EPCs retrieved from the transaction events. Aggregation events usually correspond to packing of goods into consignments.
- The object events corresponding to the "commissioning" business step for the EPCs are retrieved. Commissioning of EPCs is usually undertaken during production of goods.
- The Linked Pedigree Creator aggregates the shipping-transaction, aggregation-packing and object-commissioning events into a linked pedigree dataset. Pedigree datasets are asserted following the OntoPedigree ontology design pattern.
- The Pedigree URI Generator assigns a URI to the linked pedigree dataset.
- URIs for linked pedigree datasets can now be shared among the supply chain trading partners. Linked Pedigrees can be retrieved via content negotiation in one of the several serialisation formats.

Figure 13 highlights how various elements of the linked pedigree are populated.

Figure 14. Subset of EPCIS events captured in the tomato supply chain

	Supply chain operation	EPCIS event type	Business Step	Disposition	Action type
	Franz farmer				
1.	Commissioning punnets for tomatoes	Object event	commissioning	active	ADD
2.	Packaging punnets in crates	Aggregation event	packing	in_progress	ADD
3.	Storing crates	Quantity event	storing	in_progress	-
4.	Loading and shipping crates	Transaction event	shipping	in_transit	ADD
	Joe trader				
5.	Receiving crates	Object event	receiving	in_progress	OBSERVE
6.	Stocking crates	Object event	stocking	in_progress	OBSERVE
7.	Loading and shipping crates	Transaction event	shipping	in_transit	ADD
	Fresh food Inc.				
5.	Receiving crates	Object event	receiving	in_progress	OBSERVE
6.	Stocking crates	Object event	stocking	in_progress	OBSERVE
7.	Aggregating, loading and shipping crates	Transaction event	shipping	in_transit	ADD
	Orchards supermarket				
8.	Receiving crates	Object event	receiving	in_progress	OBSERVE
9.	Store crates	Object event	storing	in_progress	OBSERVE
10.	Arranging punnets on shelves	Aggregation event	stocking	sellable_accessible	DELETE
11.	Selling punnets	Object event	retail_selling	retail_sold	OBSERVE

EVALUATION

Our evaluation of the proposed approach focuses on the following criteria:

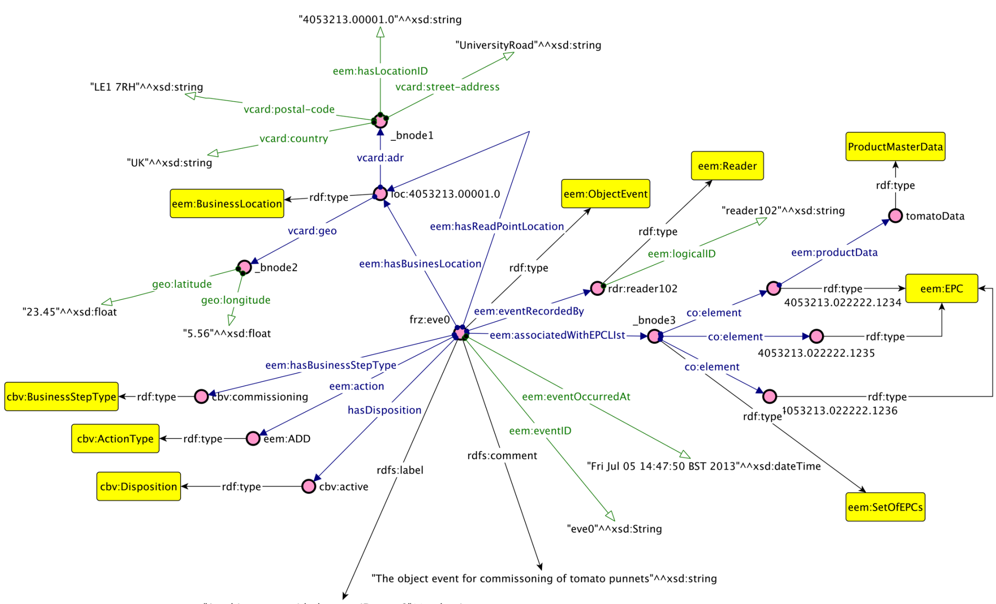
- The knowledge exchanged between stakeholders in the agri-food supply chain via linked pedigrees must facilitate the answering of competency questions highlighted earlier.
- The benefits provided by the proposed approach must address the shortcomings of existing systems for traceability within agri-food.

EPCIS Events for the Tomato Supply Chain

In order to address the first criteria, we acquired a large representative sample of EPCIS event datasets, which we augmented with a simulated set of events for the various stakeholders in the tomato supply chain. It is worth noting that for evaluating supply chain business processes, it is common practice to conduct experiment evaluations on simulated EPCIS datasets seeded with actual event data (Schapranow, Matthieu-P. et al., 2011; Müller, Jürgen and Tröger, Ralph and Zeier, Alexander and Alt, Rainer, 2009). The practice is both reliable and representative of real world scenarios. The datasets were integrated with background knowledge about tomato varieties, environmental conditions, stakeholder profiles and spatial data for locations. Integrated event datasets were stored in OWLIM 5.4 as distinct named graphs for each of the stakeholders. Figure 14 presents a subset of the type of EPCIS events captured in the tomato supply chain.

Figure 15 illustrates an Object event captured at the EPCIS implementation deployed at Franz farmer's packaging utility and expressed using EEM. The Object event relates to the commissioning of punnets for tomatoes.

Figure 15. An EEM object event representation from the tomato supply chain



GENERATION OF LINKED PEDIGREES

We generated linked pedigrees using the LinkedEPCIS library and the integrated EPCIS event datasets. A snippet of the pedigree sent from the trader to the retailer includes information received from the farmers, upstream in the supply chain as illustrated below.[37].

The significant advantage of exchanging traceability information using linked pedigrees over conventional mechanisms is that the pedigree received by FreshFood Inc. from Joe trader includes URIs to pedigree datasets provided by Franz farmer and Bob farmer, even though they are not FreshFood Inc's one-up or one-down partners. Consuming EPCIS event data curated as linked data to generate and exchange linked pedigrees as outlined above can help in a number of ways, above all in tracking and tracing scenarios, but also in providing a conduit for more detailed data and information about products, in providing information flow upstream to producers, and increasing the efficiency of the overall supply chain.

Querying Pedigrees for Traceability

In order to answer the various traceability queries outlined earlier, we formulated SPARQL queries and executed them over the integrated datasets. The evaluations were made on Mac OSX 10.9.2, 1.7GHz Intel core i5, 4GB 1333 MHz DDR3. We selected three representative queries from our requirements analysis of the agri-food domain. They are outlined below:

Query 1: Which products sold in the last 10 days at retailer ABC are vegetables and who are the suppliers and eventual producers?

The above query exploits the SPARQL 1.1 feature of property paths to retrieve all the received pedigrees that are included in the final pedigree. Received pedigrees include informa-

Pedigree sent from a farmer to the trader

```
###http://fispace.aston.ac.uk/franzfarmer/pedigrees/FranzToma-
toFarmerPedigree123
fsc:FranzTomatoFarmerPedigree123 rdf:type ped:Pedigree;
ped:hasSerialNumber "tomPed123"^^xsd:String;
ped:hasStatus ped:InitialPedigree;
ped:hasConsignmentInfo fci:FranzFarmerObjectEvent10,
fci:FranzFarmerAggregationEvent6;
ped:hasTransactionInfo fti:FranzFarmerShippingEvent12;
ped:hasProductInfo ftp:FranzTomatoesMay2013Data.
```

Pedigree sent by the trader to the retailer

```
###http://fispace.aston.ac.uk/joetrader/pedigrees/JoeTomatoTra-
derPedigree456
jsc:JoeTomatoTraderPedigree456 rdf:type ped:Pedigree;
ped:hasSerialNumber "joeTradePed456"^^xsd:String;
ped:hasStatus ped:IntermediatePedigree;
ped:hasConsignmentInfo jci:JoeTraderObjectEvent20,
jci:JoeTraderObjectEvent30;
ped:hasTransactionInfo jti:JoeTraderTransactionEvent40;
ped:hasProductInfo jpi:JoeTradesMay2013Info;
ped:hasReceivedPedigree fsc:FranzTomatoFarmerPedigree123,
bsc:BobTomatoFarmerPedigree123.
```

Query 1

```
SELECT ?products ?partners WHERE{
?pedigree a ped:Pedigree;
prov:wasAttributedTo http://www.abc.net;
ped:hasStatus ped:FinalPedigree;
ped:hasProductInfo ?products;
ped:hasReceivedPedigree+ ?rpedigree;
?rpedigree prov:wasAttributedTo ?partners.
?products prod:typeOfGood ?description;
prod:status prod:Sold;
prod:soldAt ?time.
FILTER((contains(str(?description), ''vegetable'')))
&& (now()-?time)<=10)
}
```

tion about the supply chain partner (supplier or producer) that generated the pedigree.

Query 2: List the locations this tomato went through between date1 and date2.

Query 3: For all products sold between dates/times t1 and t2, and which are potential E. Coli carriers, provide a list of the producers and locations traversed by each product item.

CONCLUSION

Data integration in supply chains, particularly for tracking, tracing and transparency, has received considerable attention in recent years. Information systems are now being designed to facilitate the process of making data available in real time to stakeholders in the supply chain, while keeping access control restrictions in place. In this paper, we have shown how Semantic Web standards, ontologies and linked data can be utilised to curate and represent real time supply chain knowledge, thereby significantly contributing to the vision.

The representation of EPCIS events on the Web of data is an important step towards achieving the objectives of sharing traceability information between trading partners and detecting inconsistencies in supply chains on a Web scale. We have incorporated EEM and CBVVocab in the generation of linked pedigrees, that provide the ontological primitives required to represent EPCIS events using Semantic Web standards. EEM is an OWL DL ontology and builds on foundational modelling decisions based on our requirements analysis of the supply chain sector.

Query 2

```
SELECT ?location WHERE{
?pedigree a ped:Pedigree;
ped:hasProductInfo ?product.
ped:hasConsignmentInfo ?epcList.
?product eem:hasSGTIN ?epc.
?epcList co:element ?epc.
?event eem:associatedWithEPCList ?epcList.
?event eem:hasBusinessLocation ?location.
?event eem:eventOccurredAt ?date.
FILTER((?date >= date1) && (?date <= date2))
}
```

Query 3

```
SELECT DISTINCT ?product ?location WHERE{
?pedigree a ped:Pedigree;
ped:hasStatus ped:FinalPedigree;
ped:hasProductInfo ?product;
?product prod:status prod:Sold;
prod:typeOfGood ?description;
prod:soldAt ?time.
eem:hasSGTIN ?epc.
?epcList co:element ?epc.
?event eem:associatedWithEPCList ?epcList.
?event eem:hasBusinessLocation ?location.
FILTER(contains(str(?description), ``E. Coli'') &&
((?time >= t1) && (?time <= t2)))
}
```

The capture and curation of EPCIS events linked datasets is realised using the LinkedEPCIS library, which can be integrated with existing RFID and EPCIS implementations. We have exemplified the use of the EEM model and LinkedEPCIS library by modelling and curating events from the agri-food supply chain. It is worth noting that both the model and the library can be generically applied to any supply chain governed by the EPCIS specification. We have introduced the concept of "linked pedigrees" within the framework of an open, scalable and decentralised architecture and proposed a design pattern "OntoPedigree" that provides a minimalistic abstraction for designing domain specific pedigree ontologies. Finally, we have presented a linked pedigrees communication protocol and an exemplifying usecase from the perishable food logistics domain. It is worth noting that our approach is domain independent and can be widely applied to most scenarios of traceability.

In this paper we have laid out a formally rigorous foundation for the use of Semantic Web and linked data standards in agri-food supply chains. As part of our future work, we hope to exploit this for reasoning and inference over the EPCIS event data.

Distributed and decentralised systems have been the focus of research for the last few decades. They are of particular importance in the agri-food sector due to a long history of independence and distrust of central authority. There are several issues such as trust relationship between actors, access control mechanisms and performance optimisation that need to be considered. In this paper we have abstracted from those issues. Our aim was to show the relevance of consuming supply chain event data to the problem of real time tracing and tracking in supply chains.

As part of our future work, we are looking into refining the EEM model to the OWL QL or OWL RL profile in order to facilitate querying and reasoning. We have developed bespoke SWRL rules over EPC list, actions and events, in order to materialise intuitive predicates, which are currently not a part of the EPCIS specification. These will soon be implemented and integrated within the LinkedEPCIS library. While we strongly believe that linked pedigrees can make a significant difference to current approaches to data visibility, tracking and tracing in supply chains, much work still needs to be done. We are investigating the use of rule based reasoning to enable real-time checking of pedigrees and identify problems such as "dwell-time consistency" before shipments are sent out or received. We are also refining our event model to enable optimised query retrieval over large event datasets

ACKNOWLEDGMENT

The research described in this paper has been partially supported by the EU FP7 FI PPP projects, SmartAgriFood (http://smartagrifood.eu) and FISpace (http://www.fispace.eu/)

REFERENCES

Brewster, C., Glaser, H., & Haughton, B. (2005). Aktive food: Semantic web based knowledge conduits for the organic food industry. In Proceedings of the ISWC Workshop Semantic Web Case Studies and Best Practice for eBusiness (SWCASE 05), 4th International Semantic Web Conference, 7 November. Galway, Ireland.

Främling, K., Parmar, S., Hinkka, V., Tätilä, J., & Rodgers, D. (2013). Assessment of EPCIS Standard for Interoperable Tracking in the Supply Chain. In T. Borangiu, A. Thomas, & D. Trentesaux (Eds.), *Service orientation in holonic and multi agent manufacturing and robotics* (Vol. 472, pp. 119–134). Springer Berlin Heidelberg. doi:10.1007/978-3-642-35852-4_8

Gangemi, A. (2005). Ontology design patterns for semantic web content. In *The Semantic Web - ISWC 2005*. Springer. doi:10.1007/11574620_21

He, W., Tan, E., Lee, E., & Li, T. (2009). A solution for integrated track and trace in supply chain based on RFID and GPS. In IEEE Conference on Emerging Technologies Factory Automa- tion, 2009. ETFA 2009. doi:10.1109/ETFA.2009.5347146

Hofman, W. (2011). Supply Chain Visibility with Linked Open Data for Supply Chain Risk Analysis. In Proceedings of the 1st Workshop on IT Innovations Enabling Seamless and Secure Supply Chains (WITNESS2011) (Vol. 769). CEUR Workshop proceedings.

Ilic, A., Andersen, T., & Michahelles, F. (2009, January). Increasing Supply-Chain Visibility with Rule- Based RFID Data Analysis. *IEEE Internet Computing*, *13*(1), 31–38. doi:10.1109/MIC.2009.10

Joris Hulstijn and Sietse Overbeek and Huib Aldewereld and Rob Christiaanse. (2012). Integrity of Supply Chain Visibility: Linking Information to the Physical World. In CAiSE Workshops (p. 351-365).

Kaloxylos, A., et al. (2011). First report on smart farming architectural requirements and subsystem (Tech. Rep.). SmartA- griFood Project.

Kaloxylos, A., Eigenmann, R., Teye, F., Politopoulou, Z., Wolfert, S., Shrank, C., & Kormentzas, G. et al. (2012). Farm management systems and the Future Internet era. *Computers and Electronics in Agriculture*, *89*(0), 130–144. doi:10.1016/j.compag.2012.09.002

Martini, D., Mietzsch, E., Giannerini, G., Papaekonomou, V., & Kunisch, M. (2010, October). D 2.1: European tracking and tracing backbone solution requirements analysis (Deliverable No. D2.1). Transparent Food Project.

Monika Solanki and Christopher Brewster. (2013a). Consuming Linked data in Supply Chains: Enabling data visibility via Linked Pedigrees. In Fourth International Workshop on Consuming Linked Data (COLD2013) at ISWC (Vol. Vol-1034). CEUR-WS.org proceedings.

Monika Solanki and Christopher Brewster. (2013b). Representing Supply Chain Events on the Web of Data. In Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE) at ISWC. CEUR-WS.org proceedings.

Müller, J., Tröger, R., Zeier, A., & Alt, R. (2009). Gain in Transparency Versus Investment in the EPC Network-analysis and Results of a Discrete Event Simula- tion Based on a Case Study in the Fashion Industry. In *Proceedings of the 2009 International Conference on Service-oriented Computing*. Berlin, Heidelberg: Springer-Verlag.

Nguyen, T., Lee, Y.-K., Jeong, B.-S., & Lee, S. (2007). Event Query Processing in EPC Infor- mation Services. In Proceedings of the 2007 Third International IEEE Conference on Signal-Image Technologies and Internet- Based System (pp. 159–166). Washington, DC, USA: IEEE Computer Society.

Presutti, V., & Gangemi, A. (2008). Content Ontology Design Patterns as Practical Building Blocks for Web Ontologies. In *Proceedings of the 27th International Conference on Conceptual Modeling*. Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-87877-3_11

Public Sector Information Domain of the CTO Council's cross Government Enterprise Architecture. (2009). Designing URI Sets for the UK Public Sector (Tech. Rep.). Chief Technology Officer Council.

Regattieri, A., Gamberi, M., & Manzini, R. (2007). Trace- ability of food products: General framework and experi- mental evidence. *Journal of Food Engineering*, *81*(2), 347–356. doi:10.1016/j.jfoodeng.2006.10.032

Reiche, R., Lehmann, R. J., Fritz, M., & Schiefer, G. (2011). Business process support using RFID-based information services in the distribution of fresh fruits and vegetables. In GIL-Jahrestagung in Oppenheim 2011. Gesellschaft fÃijr Informatik in der Land- Forst- und ErnÃd'hrungswirtschaft e.V.

Ruiz-Garcia, L., Lunadei, L., Barreiro, P., & Robla, I. (2009). A review of wireless sensor technologies and applications in agri- culture and food industry: State of the art and current trends. *Sensors (Basel, Switzerland)*, *9*(6), 4728–4750. doi:10.3390/s90604728 PMID:22408551

Ruiz-Garcia, L., Steinberger, G., & Rothmund, M. (2010). A model and prototype implementation for tracking and tracing agricul- tural batch products along the food chain. *Food Control*, *21*(2), 112–121. doi:10.1016/j.foodcont.2008.12.003

Schapranow, Matthieu-P. et al. (2011). A Formal Model for Enabling RFID in Pharmaceutical Supply Chains. In Hawaii inter- national conference on system sciences. IEEE Computer Soci- ety.

Staab, S., & Studer, R. (Eds.). (2009). *Ontology design patterns* (2nd ed.). Springer.

Taylor, K., Griffith, C., Lefort, L., Gaire, R., Compton, M., Wark, T., & Trotter, M. et al. (2013, November). Farming the web of things. *IEEE Intelligent Systems*, *28*(6), 12–19. doi:10.1109/MIS.2013.102

Turchi, S., Ciofi, L., Paganelli, F., Pirri, F., & Giuli, D. (2012). Designing EPCIS through Linked Data and REST principles. In *20th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 2012 (p. 1-6).

Tim Verwaart et. al. (2013). SmartAgriFood Delieverable D100.4 Final Strategic Overview (Tech. Rep.). SmartAgriFood Project.

ENDNOTES

1 http://www.gs1.org/
2 http://www.gs1.org/gsmp/kc/epcglobal/epcis
3 http://www.gs1.org/gsmp/kc/epcglobal/cbv
4 http://purl.org/eem#
5 http://purl.org/cbv#
6 http://www.smartagrifood.eu
7 http://lod-cloud.net/
8 http://purl.org/goodrelations/v1
9 http://www.ontologyportal.org/
10 http://www.ifomis.org/bfo
11 http://www.w3.org/TR/owl-profiles/OWL_2_RL
12 http://ruleml.org/
13 http://www.w3.org/Submission/SWRL/
14 http://www.w3.org/2005/rules/
15 Cf. for example http://www.europoolsystem.com/158/The-system
16 http://www.interdatanet.org/wiki/index.php/Main_Page
17 https://code.google.com/p/fosstrak/
18 http://www.axway.com/products-solutions/b2b/life-sciences-solutions/epedigree
19 http://www.gs1.org/gsmp/kc/epcglobal/pedigree
20 http://www.gs1.org/docs/healthcare/Healthcare\ _Traceability_Pedigree_Background.pdf
21 http://www.cassandra-project.eu/
22 http://www.gs1.org/docs/epcglobal/an_overview_of_EPC.pdf
23 http://www.gs1.org/gsmp/kc/epcglobal/tds/tds_1_6-RatifiedStd-20110922.pdf
24 http://www.gs1.org/barcodes/technical/id-keys/ gtin
25 We do not include details of these identifiers. The interested reader is referred to the Tag Data standard
26 Deprecated in EPCIS 1.1
27 http://purl.org/co/. We specialise from a Set rather than a List as it does not contain duplicates.
28 The interested reader is referred to the EPCIS standard for details.
29 http://www.w3.org/2006/vcard/ns\#
30 Such a paper based pedigree is the "grain passport" still in use in the UK arable farming sector, cf. http://publications.hgca.com/publications/documents/cropresearch/Combinable_Crops_Passport_2011.pdf.
31 http://ontologydesignpatterns.org/wiki/Submissions:ContentOPs
32 http://purl.org/FIspace/pedigree#
33 http://purl.org/dc/elements/1.1
34 https://github.com/nimonika/LinkedEPCIS
35 http://rdf4j.org/
36 The interested reader is referred to the EPCIS standard for further details.
37 We exclude prefixes to save space.

Monika Solanki is a Senior Research Fellow in the Operations and Information Management academic group at Aston Business School. She has a Ph.D. in Computer Science from De Montfort University, UK. Her current research focuses on developing and implementing stream processing algorithms and models for tracking and tracing in supply chains, using event based pedigrees over GS1's EPCIS standards, Semantic Web standards and Linked Data technologies. Dr Solanki has extensive experience in designing and developing ontological domain models, curating linked datasets, implementing supporting Semantic Web applications and Web services for a number of interdisciplinary projects spanning the boundaries of Computer Science, Bioenergy, Archaeology and Plant Biology. She has published several papers in international conferences, workshops and journals, proposing methodologies, formal models and concrete implementation infrastructures for building Semantic Web and linked data applications. Dr Solanki is also a visiting associate lecturer at University of Leicester, UK where she is the originator and convener of an MSc module on Semantic Web and Linked data. Besides Semantic Web, Linked data and knowledge representation, her other research interests include temporal logics, formal verification (model checking) and runtime monitoring of Web service behaviours.

Christopher Brewster is a Senior Lecturer in Information Technology in the Operations and Information Management Group, Aston Business School, Aston University, Birmingham. His main current research focus is linked data and open data in the agrifood supply chain, with a particular interest in how data integration in the supply chain can mitigate against food crises, lack of sustainability and food waste. He has a PhD in Computer Science from the University of Sheffield, UK.