Towards event-based traceability in provenance-aware supply chains

Monika Solanki

University of Oxford, UK
monika.solanki@cs.ox.ac.uk

Abstract. The sharing of product information plays a central role in coordinating the actions and decisions undertaken by supply chain trading partners. The Electronic Product Code Information Service (EPCIS) is an EPCglobal standard, that aims to enable the sharing of serial-level product related information via the Electronic Product Code (EPC). Central to the EPCIS data model are events that describe specific occurrences in the supply chain. This paper presents a unified, provenance-aware framework, driven by Semantic Web standards and Linked data principles for representing and sharing EPCIS events on the Web of data. Event-based linked pedigrees are utilised as the basic artifact for exchanging knowledge in supply chains. Traceability is implemented through the automated generation and validation of linked pedigrees. We exemplify our approach using the pharmaceuticals supply chain and show how counterfeit detection is an implicit part of our traceability framework.

1 Introduction

One of the most important challenges in logistics and supply chains is information integration. Sharing of data and knowledge in a standardised manner among the various stakeholders, is crucial to enable efficient management. Further, recent advances in sensor technology has resulted in wide scale deployment of RFID enabled devices in supply chains. Timely processing of RFID data facilitates efficient analysis of product movement, shipment delays, inventory shrinkage and out-of-stock situation in end-to-end supply chain processes [1]. The scanning of RFID tags in production and storage facilities generates unprecedented volumes of events as data streams, when trading partners exchange and handle products from inception through to the end-of-life phase.

This paper presents a unified, provenance-aware framework, driven by Semantic Web standards and Linked data principles for representing and sharing of supply chain data, specifically for the purposes of tracing and tracking. Traceability data in supply chains is generated when barcode and RFID readers record traces of products tagged with an EPC (Electronic Product Code), monitoring their movement across the supply chain as specific occurrences of "events". In the proposed framework, description of events is facilitated using EPCIS[1](Electronic

[1] http://www.gs1.org/gsmp/kc/epcglobal/epcis

Product Code Information Services), a standardised event oriented specifications prescribed by GS1[2] for enabling traceability [4] in supply chains. We exploit two information models: The *EPCIS Event Model* (EEM)[3] based on the EPCIS specification, that enables the sharing and semantic interpretation of event data and *CBVVocab*[4] a companion ontology to EEM for annotating the business context associated with events.

The abstraction we use for encoding and sharing traceability data is a "linked pedigree". Event-based linked pedigrees are utilised as the basic artifact for exchanging knowledge in supply chains. We propose, OntoPedigree[5] a content ontology design pattern for generating the linked pedigrees. We represent supply chain events as streams of RDF encoded linked data, while complex event patterns are declaratively specified through extended SPARQL queries. Traceability is implemented through the automated generation and validation of linked pedigrees. We exemplify our approach using the pharmaceuticals supply chain. Counterfeiting has increasingly become one of the major problems prevalent in these chains. The WHO estimates that between five and eight percent of the worldwide trade in pharmaceuticals is counterfeit [9]. Counterfeit detection is an implicit part of our traceability framework.

The paper is structured as follows: Section 2 presents our motivating scenario from the pharmaceuticals supply chain. Section 3 highlights the contextual background for the proposed framework. Section 4 provides a brief overview of various elements of the traceability framework with references to more detailed literature. Section 5 presents conclusions.

2 Motivating scenario

We outline the scenario of a pharmaceutical supply chain, where trading partners exchange product track and trace data using linked pedigrees. Figure 1 illustrates the flow of data for four of the key partners in the chain.

The *Manufacturer* commissions[6], i.e., assigns an EPC (Electronic Product Code) to the items, cases and pallets. The items are packed in cases, cases are loaded onto pallets and pallets are shipped. At the Warehouse for the *Wholesaler*, the pallets are received and the cases are unloaded. The cases are then shipped to the various *Distribution centers*. From the Distribution centers the cases are sent to retail *Dispenser* outlets, where they are received and unpacked. Finally, the items are stacked on shelves for dispensing, thereby reaching their end-of-life in the product lifecycle.

EPCIS events are internally recorded for various business steps at each of the trading partner's premises and used for the generation of linked pedigrees. When the pallets with the cases are shipped from the manufacturer's premises to

[2] `http://www.gs1.org/`
[3] `http://purl.org/eem#`
[4] `http://purl.org/cbv#`
[5] `http://purl.org/pedigree#`
[6] associates the serial number with the physical product

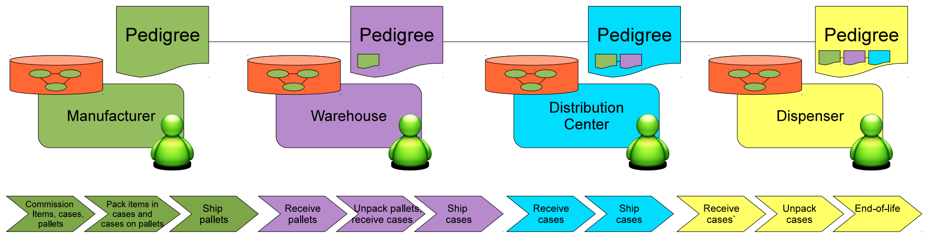

Fig. 1. Trading partners in a pharmaceutical supply chain and the flow of information

the warehouse, pedigrees encapsulating the set of EPCIS events encoding traceability data are published at an IRI based on a predefined IRI scheme. At the warehouse, when the shipment is received, internal EPCIS events corresponding to the receipt of the shipment are recorded. The IRI of the pedigree sent by the manufacturer is dereferenced to retrieve the pedigree. IRIs of the events corresponding to the transaction (shipping) and consignment (goods) information encapsulated in the pedigree are also dereferenced to retrieve the event specific information for the corresponding business steps. When the warehouse ships the cases to the distribution center, it incorporates the IRI of the manufacturer's pedigree in its own pedigree definition. As the product moves, pedigrees are generated with receiving pedigrees being dereferenced and incorporated, till the product reaches its end-of-life stage. Note that pedigrees sent by a distributor may include references to the pedigrees sent by more than one warehouse.

3 Preliminaries

3.1 EPCIS

An Electronic Product Code (EPC)[7] is a universal identifier that gives a unique, serialised identity to a physical object. EPCIS is a ratified EPCglobal[8] standard that provides a set of specifications for the syntactic capture and informal semantic interpretation of EPC based product information. As the EPC tagged object moves through the supply chain, RFID readers record and transmit the tagged data as "events". Given the scenario in Section 2, we are concerned with three types of EPCIS[9] events: *ObjectEvent* represents an event that occurred as a result of some action on one or more entities denoted by EPCs, i.e., commissioning of an object *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 cases are aggregated to a

[7] `http://www.gs1.org/gsmp/kc/epcglobal/tds/tds_1_6-RatifiedStd-20110922.pdf`
[8] `http://www.gs1.org/epcglobal`
[9] Please refer the specification for details.

pallet) *TransactionEvent* represents an event in which one or more entities denoted by EPCs become associated or disassociated with one or more identified business transactions, i.e., the shipping of a pallet of goods in accordance to the fulfillment of an order.

3.2 The EEM ontology

EEM is an OWL 2 DL ontology for modelling EPCIS events. EEM conceptualises various primitives of an EPCIS event that need to be asserted for the purposes of traceability in supply chains. A companion standard to EPCIS is the Core Business Vocabulary(CBV) standard. The CBV standard supplements the EPCIS framework by defining vocabularies and identifiers that may populate the EPCIS data model. *CBVVocab* is an OWL ontology that defines entities corresponding to the identifiers in CBV. Development of both the ontologies was informed by a thorough review of the EPCIS and the CBV specifications and extensive discussions with trading partners implementing the specification. The modelling decisions [7] behind the conceptual entities in EEM highlight the EPCIS abstractions included in the ontology. It is worth noting that in previous work [12] we have already defined a mapping between EEM and PROV-O[10], the vocabulary for representing provenance of Web resources. This implies that when a constraint violation is detected, the events in the history can be interrogated using PROV-O for recovering provenance information associated with the events.

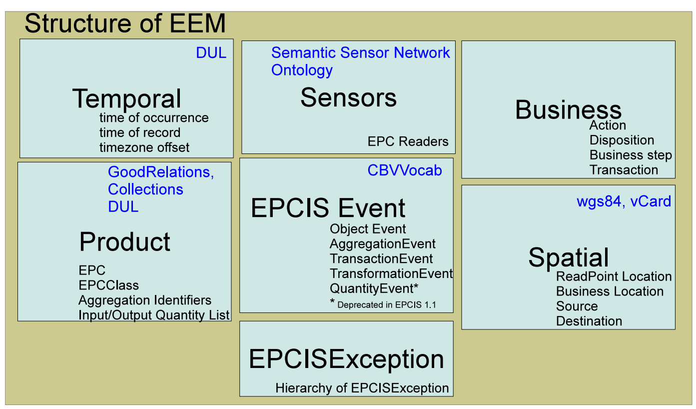

Fig. 2. Structure of EEM and its alignment with external ontologies (noted in blue coloured text)

The EEM ontology structure and its alignment with various external ontologies is illustrated in Figure 2. The ontology is composed of modules that define

[10] `http://www.w3.org/ns/prov-o`

various perspectives on EPCIS. The *Temporal* module captures timing properties associated with an EPCIS event. It is aligned with temporal properties in DOLCE+DnS Ultralite (DUL)[11]. Entities defining the EPC, aggregation of EPCs and quantity lists for transformation events are part of the *Product* module. The GoodRelations[12] ontology is exploited here for capturing concepts such as an `Individual` Product or a lot (collection) of items, `SomeItems` of a single type. Information about the business context associated with an EPCIS event is encoded using the entities and relationships defined in the *Business* module. RFID readers and sensors are defined in the *Sensor* module. The definitions here are aligned with the SSN[13] ontology. The *EPCISException* module incorporates the hierarchy of the most commonly observed exceptions [13] occurring in EPCIS governing supply chains.

For further details on EEM and its applications in real world scenarios, the interested reader is referred to [6, 7, 11, 12].

3.3 Linked pedigrees

A Pedigree is an (electronic) audit trail that records the chain of custody and ownership of a drug as it moves through the supply chain. Each stakeholder involved in the manufacture or distribution of the drug adds visibility based data about the product at their end, to the pedigree. Recently the concept of "Event-based Pedigree"[14] has been proposed that utilises the EPCIS specification for capturing events in the supply chain and generating pedigrees based on a relevant subset of the captured events. In previous work [6] we introduced the concept of linked pedigrees in the form of a content ontology design pattern, "OntoPedigree". We proposed a decentralised architecture and presented a communication protocol for the exchange of linked pedigrees among supply chain partners. In [11], we extended OntoPedigree to include provenance metadata as illustrated in Figure 3 and proposed an algorithm for the automated generation of linked pedigrees. For the purpose of completeness, we briefly recall the axiomatisation of a linked pedigree in Figure 4.

The definition highlights the mandatory and optional restrictions on the relationships and attributes for every pedigree that is exchanged between stakeholders. Based on these, we define the requirements on the constraints to be validated for the pedigrees.

4 The traceability framework

In this section we provide a brief overview of the traceability framework. Interested readers are referred to specific past work that cover the various aspects in further detail.

---

[11] `http://ontologydesignpatterns.org/ont/dul/DUL.owl`
[12] `http://purl.org/goodrelations/v1`
[13] `http://purl.oclc.org/NET/ssnx/ssn`
[14] `http://www.gs1.org/docs/healthcare/Healthcare_Traceability_Pedigree_Background.pdf`

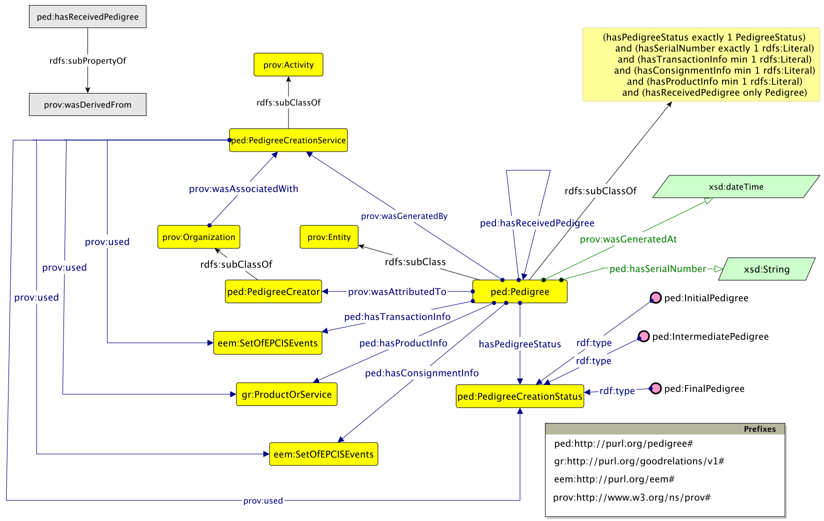

Fig. 3. Graphical Representation of Provenance based OntoPedigree

4.1 Automated generation of linked pedigrees

We have proposed a pedigree generation algorithm based on complex processing of continuous and real time streams of RFID data in supply chains. Our streams comprise of events annotated using RDF/OWL vocabularies. Event streams are generated, as products tagged with RFID identifiers are scanned and handled by diverse trading partners, in various phases of an end-to-end supply chain. Annotating streams using standardised vocabularies ensures interoperability between supply chain systems and expands the scope to exploit ontology based reasoning over continuously evolving knowledge.

In the proposed approach, we represent supply chain events as streams of RDF encoded linked data, while complex event patterns are declaratively specified through extended SPARQL queries. In contrast to existing approaches [3,5,8] where an element in a stream is a triple, our streams comprise of events where each event is represented as a named graph [2]. A linked pedigree is considered as a composition of named graphs, represented as an RDF dataset[15].

A detailed explanation of the methodology can be found in [11].

4.2 Validation of traceability artefacts

Supply chain data is inherently very sensitive to adhoc integration with third party datasets. For a specific stakeholder, effectiveness of the business workflows and decision support systems utilised within its supply chain operations, that ultimately govern the timely fulfillment of its contractual obligations, is directly

[15] http://www.w3.org/TR/rdf11-datasets/

```
Prefix ped: <http://purl.org/pedigree#>
Prefix prov: <http://www.w3.org/ns/prov-o>

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 someValuesFrom 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
```

Fig. 4. Manchester syntax serialisation of OntoPedigree

dependent on the quality and authenticity of the data received from other partners. Before traceability datasets received from external sources and partners can be incorporated and integrated with the supply chain datasets generated internally within an organisation, to be further shared downstream, they need to be validated against information recorded for the physical goods received as well as against bespoke rules, defined to ensure the quality, uniformity, consistency and completeness of datasets exchanged within the supply chain.

We have proposed a methodology for validating the traceability data sent from one stakeholder to another in the supply chain. Our approach is motivated by four main requirements: (1) The validation should be supply chain domain agnostic, i.e, the constraints must be reusable independently of the goods being tracked. (2) The representation and sharing of traceability data must conform to standards most commonly deployed in supply chains (3) The architecture must be scalable to handle large volumes of streaming traceability data and (4) The constraints must be formalised using widely used Semantic Web standards that are fit-for-purpose. While constraints can be represented using expressive formalisms such as temporal logics, adopting a unified mechanism for representing domain knowledge and constraints eliminates impedance mismatch between the representations, avoids the need for an intermediate mapping language, makes the addition of new constraints easier and simplifies implementation requirements.

In the proposed approach, we show how linked pedigrees received from external partners can be validated against constraints defined using SPARQL queries and SPIN[16] rules. To the best of our knowledge, validating constraints on (real

[16] http://www.w3.org/Submission/spin-overview/

time) supply chain knowledge has so far not been explored both within the Semantic Web and supply chain communities.

A detailed illustration of the proposed approach can be found in [10, 13]

5 Conclusions

Data visibility in supply chains has received considerable attention in recent years. In the healthcare sector, visibility of datasets that encapsulate track and trace information is especially important in addressing the problems of drug counterfeiting. In this paper we have shown how Semantic Web standards, ontologies and linked data can be utilised to represent and process real time streams of supply chain knowledge, thereby significantly contributing to the vision. Provenance, which is a critical aspect of supply chain knowledge is an integral part of our framework. We have shown how we exploit this knowledge for the validation of constraints that are defined to ensure the quality, uniformity, consistency and completeness of datasets exchanged between supply chain partners. We have performed an exhaustive evaluation of the framework which have been reported in various works outlined in the paper. Our results provide very useful insights in improving the overall efficiency of the supply chain. It is worth noting that while we have chosen the healthcare sector as a case study, our approach is domain independent and can be widely applied to most scenarios of traceability.

References

1. F. M. Alexander Ilic, Thomas Andersen. EPCIS-based Supply Chain Visualization Tool. Auto-ID Labs White Paper WP-BIZAPP-045, 2009.
2. J. J. Carroll, C. Bizer, P. Hayes, and P. Stickler. Named graphs, provenance and trust. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05. ACM, 2005.
3. Davide Francesco et al. C-SPARQL: a Continuous Query Language for RDF Data Streams. *Int. J. Semantic Computing*, 2010.
4. K. Främling, S. Parmar, V. Hinkka, J. Tätilä, and D. Rodgers. Assessment of EPCIS Standard for Interoperable Tracking in the Supply Chain. In T. Borangiu, A. Thomas, and D. Trentesaux, editors, *Service Orientation in Holonic and Multi Agent Manufacturing and Robotics*, volume 472 of *Studies in Computational Intelligence*, pages 119–134. Springer Berlin Heidelberg, 2013.
5. Le-Phuoc, Danh et al. A native and adaptive approach for unified processing of linked streams and linked data. In *Proceedings of the 10th International Conference on The Semantic Web*, ISWC'11. Springer-Verlag, 2011.
6. Monika Solanki and Christopher Brewster. Consuming Linked data in Supply Chains: Enabling data visibility via Linked Pedigrees. In *Fourth International Workshop on Consuming Linked Data (COLD2013) at ISWC*, volume Vol-1034. CEUR-WS.org proceedings, 2013.
7. Monika Solanki and Christopher Brewster. 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, 2013.

8. Rinne, Mikko et al. Processing Heterogeneous RDF Events with Standing SPARQL Update Rules. In *OTM Conferences (2)*, Lecture Notes in Computer Science. Springer, 2012.
9. E. W. Schuster and R. Koh. Track and Trace in the Pharmaceutical Supply Chain. Auto-ID Labs, Massachusetts Institute of Technology Cambridge, MA.
10. M. Solanki and C. Brewster. A Knowledge Driven Approach towards the Validation of Externally Acquired Traceability Datasets in Supply Chain Business Processes. In *Proceedings of the 19th International Conference on Knowledge Engineering and Knowledge Management(EKAW)*, volume LNAI 8876, pages 503–518. Springer International Publishing Switzerland, 2014.
11. M. Solanki and C. Brewster. EPCIS event based traceability in pharmaceutical supply chains via automated generation of linked pedigrees. In Peter Mika et al., editor, *Proceedings of the 13th International Semantic Web Conference (ISWC)*. Springer-Verlag, 2014.
12. M. Solanki and C. Brewster. Modelling and Linking transformations in EPCIS governing supply chain business processes. In Hepp, Martin; Hoffner, Yigal (Eds.), editor, *Proceedings of the 15th International Conference on Electronic Commerce and Web Technologies (EC-Web 2014)*. Springer LNBIP, 2014.
13. M. Solanki and C. Brewster. Monitoring EPCIS Exceptions in linked traceability streams across supply chain business processes. In *Proceedings of the 10th International Conference on Semantic Systems(SEMANTiCS)*. ACM-ICPS, 2014.