A General Framework for Food Traceability

A. Bechini, M.G.C.A. Cimino, B. Lazzerini, F. Marcelloni, A. Tomasi
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni
University of Pisa
Via Diotisalvi 2, 56122 Pisa
Fax: +39 050 2217600; Tel: +39 050 2217599
{a.bechini, m.cimino, b.lazzerini, f.marcelloni, a.tomasi}@iet.unipi.it

Abstract

In this paper, we propose a generic data model for food traceability. We discuss the characteristics of our model with respect both to tracing and tracking requirements, and to quality control. Since traceability is based on reliable and faithful exchange of documents among the units of the supply chain, we also consider how electronic business-to-business standards, such as ebXML, can help support data homogeneity as well as system interoperability. Finally, we highlight how the implementation effort can be accordingly reduced.

1. Introduction

Article 3 of the EU General Food Regulation, which will be applied from 1 January 2005, defines traceability as the "ability to trace and follow a food, feed, food-producing animal or substance intended to be, or expected to be incorporated into a food or feed, through all stages of production, processing and distribution" [1]. A preliminary study on the traceability in the food chain drawn up by the Food Standard Agency, an independent food safety watchdog set up by an Act of Great Britain Parliament in 2000 to protect the public's health and consumer interests in relation to food, identifies three basic characteristics for traceability systems [2]: i) univocal identification of units/batches (denoted as lots in the following) of all ingredients and products, ii) information on when and where they are moved and transformed, and iii) a means to link these data. In practice, traceability systems are record keeping procedures that show the path of a particular product or ingredient from supplier(s) into the business, through all the intermediate steps which process and combine ingredients into new products and through the supply chain to consumers. Both products and processes may form key components in a traceability system with information stored in relation to each.

Though several efforts have been devoted to the development of effective traceability systems in the last years, a preliminary analysis of existing systems reveals some open problems [3][4]. For instance, typically only a few units of a supply chain are provided with a traceability system and often these systems are proprietary, thus making integration quite difficult. Legal requirements are often not or insufficiently represented. Further, a large amount of traceability systems are not adequately equipped for timely and accurate tracing of products through the chain. Finally, these systems are more aimed at guaranteeing the system certification than preventing incidents and contributing to the real improvement of food safety.

In this paper, we discuss a generic data model for food tracing and tracking. To overcome the limitations of current traceability systems, we have developed our model considering legal requirements, data communication and quality issues as specifications. To support data homogeneity, scalability and interoperability, we have used the electronic business eXtensible Markup Language (ebXML) [5], which is becoming a sort of electronic business-to-business standard.

2. Traceability issues and related data model

In one of the most advanced theoretical treatment of traceability, Kim *et al.* describe the core of a traceability system as the ability to trace both lots and activities [6]. Thus, each data model coping with traceability has to define lot and activity as basic entities. Further, the model has to allow lot tracking and tracing. The term *"tracking"* is used to specify the ability to follow the downstream path of a product along the supply chain (possibly according to given criteria). This is a crucial feature for an efficient recall of products, e.g., for possible safety defects. The term *"tracing"* refers to the ability to determine the origin and characteristics of a particular product. This can be obtained by moving upstream the supply chain. Tracing is especially useful to detect the cause of quality problems. Figure 1 shows the scenario of a contamination event that has been detected at the end of a very simple supply chain (only four segments are considered). In the figure, a circle denotes a *traceability lot* (*lot*, for short), that is a unit of the food product processed or packaged in a similar condition, or a mass of products that share such characteristics as type, category, size, package and place of origin. A rectangle represents an *activity*, such as production, preparation, distribution and sale, which may have *N* lots as input and may deliver *M* lots as output. An edge indicates the relation between a lot and an activity. In practice, edges allow describing the route of lots. The supply chain unit, where the activity is performed, is responsible for the activity itself and for the corresponding outcoming lots. In the following, the unit is denoted as *responsible actor*. Since we consider that each lot is generated by an activity, we can conclude that each lot has a responsible actor. As regards traceability purposes, this actor is also responsible for the reliability of the traceability data related to the lot.

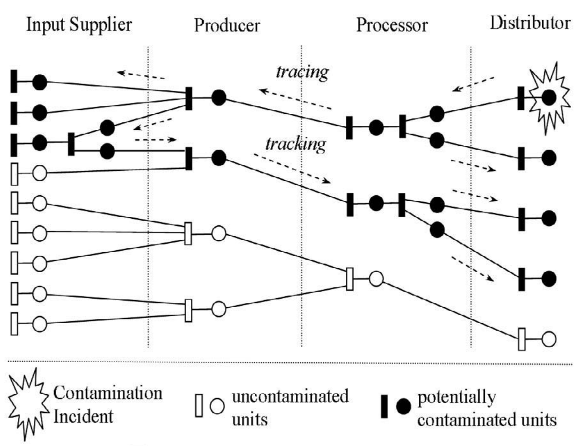

Figure 1. Typical scenario for a product recall in a supply chain.

The presence of an efficient traceability system allows constraining the product recall only to the products actually affected by contamination. Tracing and tracking capabilities are therefore crucial to confine the reaction to possible hazards and reduce the recovery cost.

The scenario shown in Figure 1 requires the adoption of an appropriate data model, which is general and suitable for any food type. Moreover, such a model has to be able to support the univocal identification of traceability lots and activities, and the recording of lots and activities, and their relations. To take the today's requirements on food quality for health care into consideration, additional data that are not strictly necessary to traceability are typically stored. For instance, for a cooking activity, oven temperature and humidity can be considered as important parameters in case of hazard.

Each lot is identified by a *global identifier*, which has to be univocal within the supply chain. To avoid a centralized administration of the identifiers, we adopt a solution that is inspired to the approach used in the EAN/UCC standard. We assume that each actor is uniquely identified in the supply chain by an *actor identifier*. We allow an actor to freely associate an identifier (*traceable entity identifier*) with each traceable entity, i.e. either an activity or a lot, the actor is responsible for. As regards lot identification, in case of an actor that gives off different products, the identifier might be composed, for instance, of the type of the product and a progressive number. The only constraint we impose is that the identifier is univocal within the amount of lots managed by the actor. Actually, this is a minimal requirement to guarantee traceability of a lot. The *global identifier* is composed of the *actor identifier* and the *traceable entity identifier*.

The model is described in Figure 2 by using a UML notation. Here, two distinct packages are shown: Traceability and Quality. The former contains the entities that allow tracing and tracking the product route. The latter groups together the components related to lot quality. The Traceable Entity is an abstract class which models the basic characteristics of the two entity types involved in traceability: lots and activities. The field TE_ID implements the traceable entity identifier. The association relation *is managed by* enforces a traceable entity to be always associated with a responsible actor. This constraint guarantees the univocal identification of the traceable entity, as described above. Furthermore, Traceable Entity has also an association with Site, which has its own univocal identifier. This relation imposes that each lot is contained into a site. Thus, at each stage of the supply chain, the traceability system is able to retrieve the information about the site where the lot has been processed or stored. Both Site and Responsible Actor are characterized by a number of attributes, which summarize all the information needed for traceability. Classes Lot and Activity are derived from Traceable Entity. The association relation *is generated from* means that each lot may be generated from one or more lots. The generation is ruled by an activity.

Figure 3 shows an example of the objects used to record an activity: a distributor purchases a red wine cask from a producer, and carries it to her/his storehouse by a truck. The input and the output lots of the activity definitely are the same cask lot. However, producer and distributor typically identify the lot in a different way. Moreover, producer and distributor are respectively responsible for the input and the output lot. Thus, from the traceability standpoint, input and output lots are different.

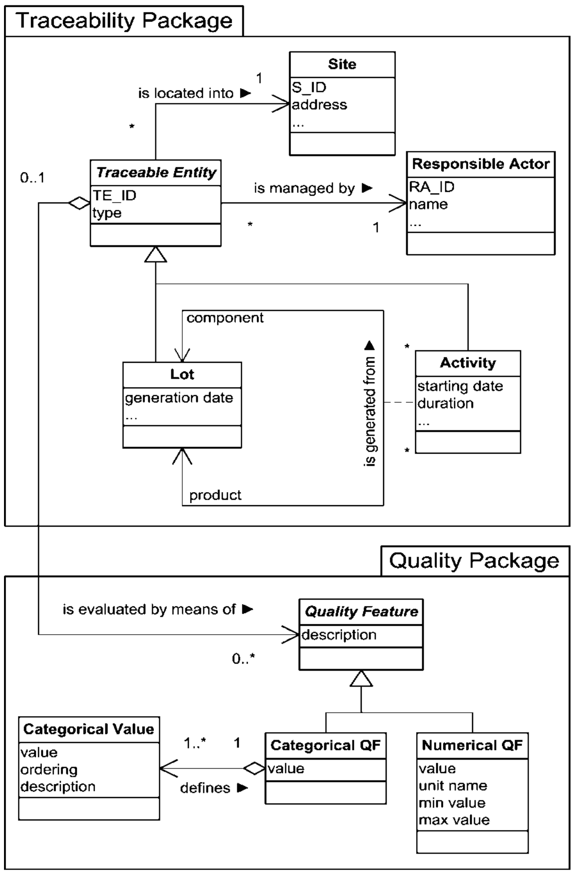

Figure 2. UML class diagram of the traceability data model.

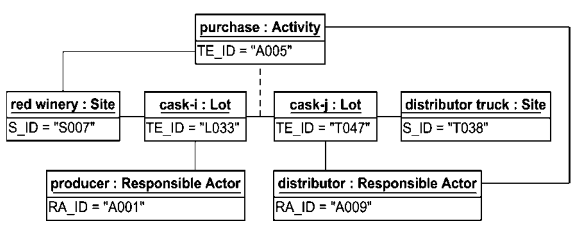

Figure 3. Objects involved in recording the actual execution of a simple activity.

The purchase activity can be described by the two sequence diagrams shown in Figure 4. The first diagram refers to a distributed architecture without a central database (data trustee). Here, the unit responsible for an activity is also responsible for recording the relation between input and output lots. The producer communicates the global identifier of the input lot to the distributor, which provides to associate it with the global identifier of the output lot. This association allows tracing and tracking the lot. Typically, the global identifier is attached as bar code or RFID tag to the lot. Thus, part of the communication between supply chain units generally consists of reading the identifier by appropriate readers.

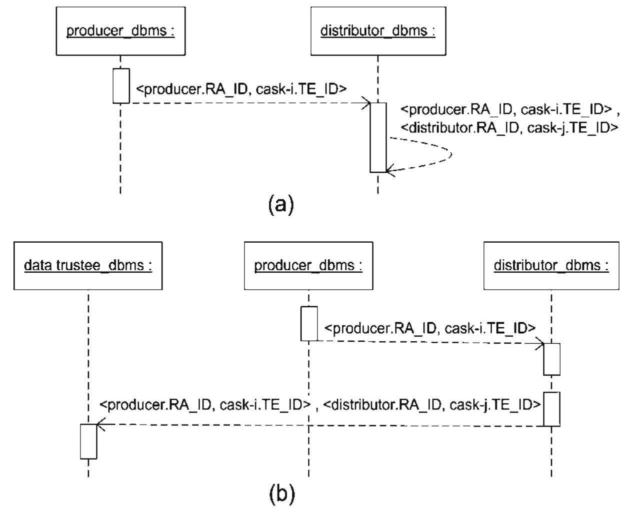

Figure 4. Sequence diagram of a purchase activity; (a) distributed architecture (b) centralized architecture.

The second diagram assumes the existence of a central data trustee, which is responsible for the traceability data. This architecture requires that each supply chain unit responsible for an activity provides the data trustee with all the information related to the activity. In particular, this information must allow data trustee at least to associate the input lot(s) with the output lot(s).

In both architectures, when tracing and tracking are required, supply chain units have to communicate with each other and possibly with the data trustee in order to retrieve the history of a lot. Traceability can succeed only if data are exchanged in a secure and reliable way.

Quality is defined by the ISO 9000 standard as the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs. To take quality issues into account, we introduced the Quality Package shown in Figure 2. This package contains the abstract class Quality Feature, which is characterized by a description of the feature itself and a collection of methods to set and retrieve feature values. Values can be either categorical or numerical. Categorical QF and Numerical QF concrete classes implement features that can assume, respectively, categorical and numerical values. Categorical QF contains a set of Categorical Value objects, which define the possible values. A Categorical Value is characterized by the value, a description, and an ordering number. This last item can be used whenever ordered categorical values are needed.

Numerical QF is qualified by the value, the unit name (for instance, Kg for "weight" quality factor), and the minimum and maximum values. This class organization allows dealing uniformly with different quality features.

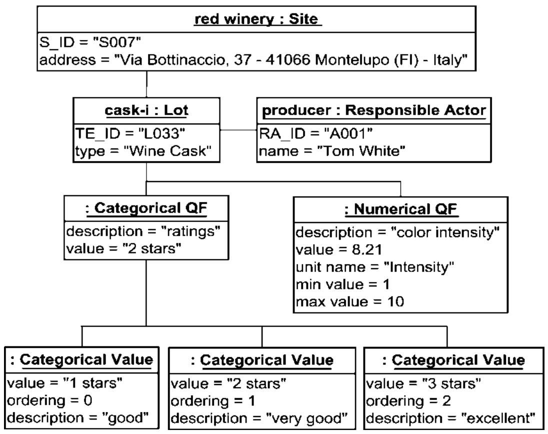

Figure 5. Example of objects related to quality features.

Figure 5 shows an example of object diagram that describes the quality features "color intensity" and "ratings" associated with lot *wine cask*. Color intensity can assume numerical values in the interval [1,10]. Ratings take the wine excellence into account. Here, excellence is evaluated using three values: one star, two stars, and three stars, which correspond, respectively, to good, very good and excellent.

3. Traceability and e-business standards

In practically exploiting the proposed data model, we have to keep in mind that each responsible actor actually belongs to a certain enterprise. The flow of product lots along the supply chain is associated with information exchanges among responsible actors and possibly third-party organizations. Because of its significance for the concerned enterprises, the traceability (and quality) information must always be transmitted in a secure and reliable way. Such a kind of communication, especially when different enterprises are involved, can be regarded as an e-business transaction. For this reason, the adoption of e-business protocols becomes both appropriate (because of the very nature of transactions) and convenient (as the interaction takes advantage of an established technological support). ebXML [5] is a modular suite of specifications that enables enterprises to conduct business over the Internet more easily and efficiently. The ebXML specification provides organizations with a common, extensible, and automated method of exchanging business messages, conducting trading relationships, communicating data using common terms, and defining and registering business processes — such as ordering, shipping, and invoicing. There are several reasons that may lead to the employment of ebXML (or similar standards/specifications) in a traceability system as well: the most important one is that traceability information exchanges are actually B2B transactions, thus naturally supported by ebXML. Moreover, ebXML directly allows overcoming the heterogeneity of the participating information systems, providing the basic means for interoperability. In case a B2B infrastructure is already present, the traceability system can be built more easily upon it, taking advantage of the existing support and of the standard data treatment.

Finally, ebXML is a promising standard also for e-government distributed applications [5]: in this perspective, monitoring activities from official organizations would be made easier by a common communication infrastructure among the actors taking part to the traceability system. We have developed and implemented a prototype of a traceability system. Some ebXML documents used in this prototype can be viewed at http://www.ing.unipi.it/~o1553499/trace.htm.

4. Conclusions

Based on a possible scenario of a product recall in a supply chain, we have proposed a generic data model for food traceability. We have described the basic classes of the model and shown some samples of instantiation. We have developed a prototype of a traceability system and have verified that the model is able not only to trace and track a product, but also to provide a useful support to quality control. We are currently applying the prototype to a real vegetable supply chain in Tuscany.

References

[1] Regulation (EC) n. 178/2002 of the European Parliament and of the Council of 28 January 2002, ch. I, art. 3, par.15.

[2] Food Standards Agency, "Traceability in the Food Chain – A preliminary study," March 2002, http://www.foodstandards.gov.uk/news/newsarchive/traceability.

[3] M. de Castro Neto, M.B. Lima Rodrigues, P. Aguiar Pinto, I. Berger, "Traceability on the web – a prototype for the Portuguese beef sector", EFITA 2003 Conference, 5-9 July 2003, Debrecen, Hungary, pp. 607-611.

[4] C.A. van Dorp, "Tracking and Tracing Business Cases: Incidents, Accidents and Opportunities", EFITA 2003, 5-9 July 2003, Debrecen, Hungary, pp. 601-606.

[5] ebXML official website – http://www.ebxml.org.

[6] H.M. Kim, M.S. Fox, M. Gruninger, "Ontology of Quality for Enterprise Modelling" In Proceedings of WET-ICE, Los Alamitos, CA, USA, pp. 105-116, IEEE 1995.