2010

Operational techniques for implementing traceability in bulk product supply chains

Maitri Thakur
Iowa State University

Recommended Citation

Thakur, Maitri, "Operational techniques for implementing traceability in bulk product supply chains" (2010). *Graduate Theses and Dissertations*. 11527.
http://lib.dr.iastate.edu/etd/11527

Operational techniques for implementing traceability in bulk product supply chains

by

Maitri Thakur

A dissertation submitted to the graduate faculty

in partial fulfillment for the degree of

DOCTOR OF PHILOSOPHY

Co-Majors: Agricultural Engineering; Industrial Engineering

Program of Study Committee:

Charles R. Hurburgh Jr., Co-Major Professor
Sigurdur Olafsson, Co-Major Professor
Thomas J. Brumm
Bobby J. Martens
Carl J. Bern
Lizhi Wang

Iowa State University

Ames, Iowa

2010

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my major professor Dr. Charles R. Hurburgh for his guidance and encouragement throughout the course of my study. I am thankful to my co-major professor Dr. Sigurdur Olafsson and Dr. Carl J. Bern, Dr. Thomas J. Brumm, Dr. Lizhi Wang and Dr. Bobby J. Martens for their support, advice and service on the program of study committee.

I would also like to thank all my colleagues at the Grain Quality Laboratory for their cooperation and friendship during my graduate studies at Iowa State University.

Special thanks goes to Eskil Forås, Jostein Storøy, Carl Fredrik Sørensen and Finn Olav Bjørnson at SINTEF Fisheries and Aquaculture for their kind hospitality, their fresh look at my work and valuable comments and ideas. A special note of thanks to Lillian Tronsaune for making my stay in Norway so comfortable.

Finally, I would like to thank my parents Ramesh and Shashi Thakur, and my brother, Rajeev for their love, support and encouragement.

ABSTRACT

Implementation of traceability techniques in bulk food product supply chains is a complex task. A systems approach was used to develop a framework for implementation of traceability in bulk grain supply chain in the United States. A relational database model was developed to facilitate internal traceability at a grain elevator, which is one of the first nodes in a food supply chain. This data management technique could mitigate the bulk grain handling problems by recording all grain lot transformations/activities, including movement, aggregation, segregation, and destruction as well as supplier and customer information. The system can be queried to retrieve information related to incoming, internal and outgoing lots and to retrieve information that connects the individual incoming grain lots to an outgoing shipment. Next, a mathematical multi-objective mixed integer programming (MIP) model was proposed with two objective functions; to calculate the minimum levels of lot aggregation and minimum total cost of blending grain in order to meet the customer contract specifications. Constraints on the system include contract specifications, availability of grain at the shipping elevator location as well as other locations and the blending requirements. The solutions include the quantities of grain from different storage bins to be used for blending for a shipment while using the minimum number of storage bins and the total cost. The numerical results are presented for a corn shipment scenario to demonstrate the application of this model to bulk grain blending. Pareto optimal front is computed for the problem for simultaneous optimization of lot aggregation and cost of blending. This model provides an effective method for minimizing the traceability effort by minimizing the food safety risk caused by lot aggregation. Finally, a new methodology for modeling the traceability information using the UML statecharts following an event management approach in bulk food production is introduced. A generic model is presented and evaluated based on its practical application in bulk food production by providing illustrations from two supply chains; pelagic fish and grain. The statecharts are developed for frozen mackerel production and corn wet milling processes. All states and events for these processes as well as the information that needs to be captured for each transition are indentified that includes the product, process and quality information. The data capture points were identified based on the various states and events that occur during food production and are connected to product, process as well as quality information.

CHAPTER 1. GENERAL INTRODUCTION

1 Introduction

Food safety and food control continue to gain significant attention as our food supply chains and production practices become increasingly complex (Hausen et al., 2006). Food safety is in fact a very important part of public health and although several advanced surveillance and monitoring systems exist in developed countries, outbreaks of food borne diseases continue to be commonplace. Such foodborne diseases are caused by consumption of contaminated foods or beverages. There are many different types of foodborne infections as many disease-causing microbes or pathogens can contaminate foods. In addition to these, several poisonous chemicals can also cause foodborne diseases if present in food (CDC, 2005). According to the Center for Disease Control and Prevention (CDC), an outbreak of foodborne illness occurs when a group of people consume the same contaminated food and two or more of them come down with the same illness. CDC (2005) estimates that foodborne diseases cause 76 million illnesses, 325,000 hospitalizations and 5,000 deaths in the United States every year.

The food industry has undergone considerable change during the past century. New farming practices as well as new handling and processing techniques have been developed to meet the increasing consumer demand for reliable and consistently safe supply of various food products. Furthermore, consumers are giving emphasis to safety, high quality and sustainability of food products. Consumer experiences with food safety and health issues combined with an increasing demand for high quality food and feed products have resulted in an increasing interest in developing systems to improve information flow and thereby food traceability. Furthermore, consumers are giving emphasis to safety, high quality and sustainability of food products. Development of integrated systems for information exchange in the food supply chains has gained considerable importance in the past few years. Various food safety and traceability laws exist in several countries.

In the United States, after the September 11 events, the US Public Health Security and Bioterrorism Preparedness and Response Act of 2002 (the Bioterrorism Act) was passed. The Bioterrorism Act requires that all companies involved in the food and feed industry to self-register with the Food and Drug Administration and maintain records and information for food traceability purposes (US Food and Drug Administration, 2002). In Canada, federal, provincial, and territorial Ministries of Agriculture agreed on a landmark agreement, entitled the Agriculture Policy Agreement (APF) in 2003. APF has five objectives including food safety and food quality.

Can-Trace was launched in July 2003 which is a collaborative and open initiative committed to the development of traceability standards for all food products sold in Canada. The mission of Can-Trace is to define and develop minimum requirements for national whole-chain tracking and tracing standards based on the GS1 system (Can-Trace, 2003). The GS1 Global Traceability Standard is a business process standard that describes the traceability process independently, in terms of key operations for any choice of enabling data management technologies (GS1 Global Traceability Standard, 2007).

The European Union's General Food Law entered into force on January 1, 2005. The law included important elements such as rules on traceability and the withdrawal of dangerous food products from the market. Under the European Union Law, "Traceability" is defined as the ability to track any food, feed, food producing animal or substance that will be used for consumption, through all the stages of production, processing and distribution (Official Journal of European Communities, 2002).

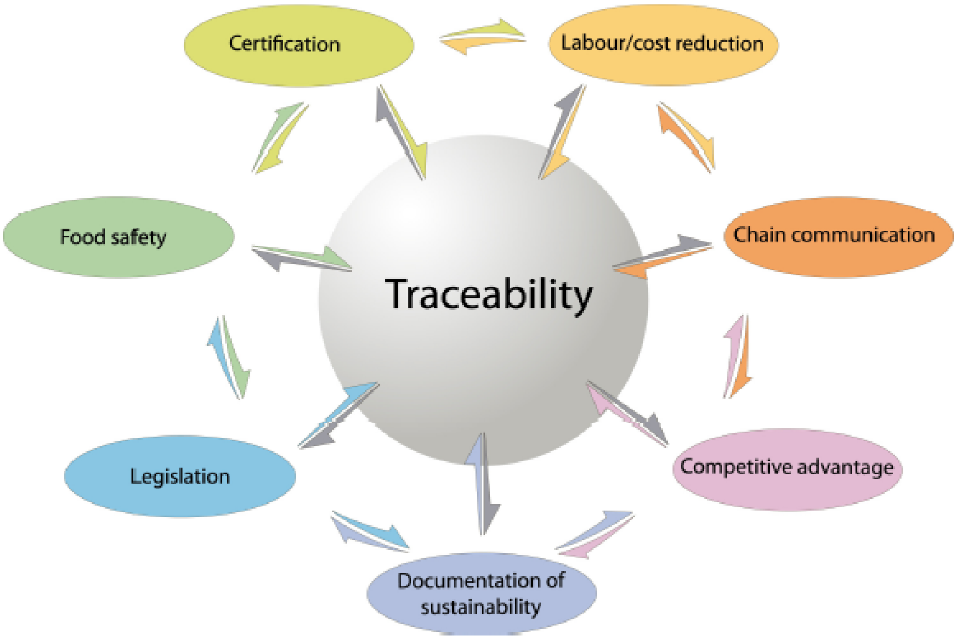


Figure 1. Motivational forces for traceability (modified from Olsen, 2009).

The ISO 22005 Food Traceability Standard states that each company know who their immediate supplier is and to whom the product is being sent, based on the principle of one up and one down. It states that food safety is the joint responsibility of all the actors involved (International Organization for Standardization, 2007). Thus, all the actors involved in the food supply chain are required to store necessary information related to the food product that links inputs with outputs, so that when requested, the information can be provided to the food inspection authorities on a timely basis. Regulations such as those in place in the EU are not the only driving forces for traceability there a many other driving forces such as its implications for food safety and are shown in Figure 1. In order to achieve a fully traceable supply chain, it is important to develop systems for chain traceability as well as internal traceability. This includes linking, to the best extent possible, units of output with specific units of input. Each supply chain actor should have a record keeping system that would enable them to trace back their ingredients and track forward the products so as to determine the cause of the problem or to efficiently recall the associated (or contaminated) food products.

2 Problem Statement

Despite the published literature on food traceability, there is a lack of research in development of bulk product traceability systems. These limitations range from addressing bulk product traceability challenges as different from other food products that are not handled and processed in bulk as well as a lack of data management systems as techniques for ensuring operational efficiency of bulk product management including handling and processing to ensure a holistic approach to development of traceability systems. It is essential to address the traceability of bulk products from a standpoint of data management strategies, costs and operational techniques that can be implemented by the industry. It is based on these needs that a series of related studies were carried out in this research.

3 Objectives

The objectives of this research were to develop operational techniques for implementing traceability systems in bulk product supply chains. These objectives were achieved by a series of research studies described in the next section.

4 Dissertation Organization

This dissertation consists of four articles and a general literature review in the field of food traceability systems. The research studies address the following objectives:

(1) Review current understanding of traceability systems implementation in the food industry (Chapter 2).
(2) Develop a framework for implementing traceability in bulk grain supply chain in the US using a systems approach (Chapter 3).
(3) Develop a database model to facilitate internal traceability at a grain elevator (Chapter 4).
(4) Develop a multi-objective optimization technique for balancing cost and traceability in bulk grain handling (Chapter 5).
(5) Develop an event management approach for modeling traceability information in bulk product supply chains (including grain and pelagic fish) using UML statecharts (Chapter 6).

In addition, two related articles are included in the appendix of this document. The first article presents a data mining technique for recognizing patterns in foodborne disease outbreaks and the second article presents modeling of traceability information in a soybean value chain. Although, not a part of the main document, these articles are related to the field of food safety and traceability and have been published in the Journal of Food Engineering.

5 Practical Implications

The deliverables from this dissertation provide operational strategies for implementing traceability systems in the bulk product supply chains, grain industry in particular. The database model developed in this research can be implemented by any grain elevator to facilitate internal traceability. The model can be easily modified for other food products and can be easily implemented along with existing logistics and inventory management techniques in food production and processing industry. Additional cost of traceability systems has been a topic of debate in the food industry. The optimization model developed in this research provides an effective way of balancing cost and traceability at a grain elevator. Again, this model can be used for other bulk products. Finally, modeling of traceability states and events in food production provides an effective technique for identification of critical traceability points where information needs to be stored. This model also provides a method for integrating product, process and quality information in one system. The output from this model can be used by systems such as EPCIS (Electronic Product Code Information Systems) for capturing data throughout food supply chains.

This dissertation contributes to the existing knowledge in the field of food traceability and specifically focuses in implementation of internal traceability systems in bulk product supply chains.

References

Can-Trace, 2003. Agriculture and Agri-Food Canada. <http://www.can-trace.org>

CDC, 2005. Foodborne Illness, Centers for Disease Control and Prevention. <http://www.cdc.gov/ncidod/dbmd/diseaseinfo/foodborneinfections_g.htm>.

GS1 Global Traceability Standard, 2007. Business Process and System Requirements for Full Chain Traceability. <http://www.gs1.org/traceability/gts>

Hausen, T., Fritz, M., Schiefer, G., 2006. Potential of electronic trading in complex supply chains: An experimental study. International Journal of Production Economics, 104, 580-597.

International Organization for Standardization, 2007. New ISO Standard to Facilitate Traceability in Food Supply Chains. ISO 22005:2007.

Official Journal of the European Communities, 2002. Regulation (EC) No 178/2008 of the European Parliament and the Council of 28 January 2002.

Olsen, P., 2009. Food Traceability Process Mapping. Standard methods for analyzing material flow, information flow and information loss in food supply chains. In Harmonizing methods for food traceability process mapping and cost/benefit calculations related to implementation of electronic traceability systems, Nofima report 15/2009. Tromsø, Nofima – Norwegian Institute of Food Fisheries and Aquaculture.

US Food and Drug Administration, 2002. The Bioterrorism Act of 2002.

CHAPTER 2. LITERATURE REVIEW

1 Importance of Traceability

Traceability is a preventive, necessary, supplement of food safety systems, which increases the efficiency of a food company, when used correctly. In practice traceability means collection, documentation, maintenance and application of information related to all processes in the supply chain, which guarantees for the consumers the information on origin and life history of a product (Opara and Mazaud, 2001）. USDA Economic Research Service states that besides ensuring a safe food supply use of a traceability system results in lower cost of distribution systems, reduced recall expenses, and expanded sales of products with attributes that are difficult to discern and in every case, the benefits of traceability translate into larger net revenues for the firm (Golan et al., 2004). Traceability is required for controlling crisis situations by enabling effective recalls, delivering precise information to consumers and regulatory authorities and for safety of consumers (EVIRA, 2007). A well thought-out traceability system is fundamental for achieving optimal benefits from quality control, production control and to fulfill consumer demands (Moe, 1998).

Some early research focuses on the importance of traceability for firms. Fisk and Chandran (1975) first gave several reasons why traceability should be considered a source of competitive advantage for firms. Traceability can open opportunities for firms to improve their product quality (Florence and Queree, 1993). Traceability used in an active way indicates the use of tracking information to optimize and control processes that must be seen as a tool for managing quality information through the entire supply chain (Jansen-Vullers et al., 2003).

Besides food producers and processors, consumers mostly gain hidden benefits from traceability that include effective achievement of food safety and an increased effectiveness of recall in case of emergencies (FSA, 2002). Food safety is the most important motivation for traceability. Food manufacturers develop and adopt internal traceability systems and traceability chains mainly to improve food safety, since traceability can be seen as a subsystem and its presence is essential to the management of food quality (Peri, 2002). Traceability is an essential tool for ensuring both production and product quality (Becker, 2000; Wall, 1994).

Moe (1998) showed that a good traceability system can provide several competitive advantages that include improvement in process control, better use of raw materials by linking the end product and raw material data, avoiding the mixing of high-quality and low-quality raw materials and easier quality auditing process.

2 Supply Chain Traceability

The ISO 22005 Food Traceability Standard requires that each company know who their immediate supplier is and to whom the product is being sent, on the principle of one-up and one-down. It states that food safety is the joint responsibility of all the actors involved (International Organization for Standardization, 2007). Thus, all the actors involved in the food supply chain are required to store necessary information related to the food product that link inputs with outputs, so that when demanded, the information can be provided to the food inspection authorities on a timely basis. For effective supply chain operations, the activities of all partners in the supply chain must be synchronized. This synchronization can be achieved only by implementation of a system that facilitates information sharing on various activities that add value long the supply chain and the coordination between internal and external partners within the chain (Williamson et al., 2004; Gunasekaran and Ngai, 2004). The general Food Law (Official Journal of European Communities, 2002) requires chain traceability systems. The guidance on the implementation of EC Food Law Regulation Article 18 (Guide 178/2002) declares that “it is in the logic of Article 18 that a certain level of internal traceability would be put in place by food business operators”.

2.1 Internal traceability

Previous research has emphasized the importance of internal traceability systems (Moe, 1998). Internal traceability is related to the ability to trace product information internally within a company, and has typically the following characteristics (Martínez-Sala et al., 2009): (1) It is within one company and at one geographical location. (2) It gets a lot of information from the production management systems. (3) There are few privacy issues. Many companies have good routines and software systems for keeping track of internal traceability. This kind of software is often linked with dedicated production management software and general Enterprise Resource Planning (ERP) systems.

The analysis of existing traceability systems shows that only a few links in a supply chain are using software for internal traceability and the diversity of these systems makes the integration difficult (Bechini et al., 2005). Typical production processes within a food company are made up of different transformations of raw materials into a finished product ready for shipment. For food traceability purposes, it is important to record which input factors have been used to produce which output products (Senneset et al, 2007).

2.2 Chain traceability

Chain traceability refers to the exchange of product information between different actors in a food value chain. Figure 2 shows the principles of internal traceability and chain traceability. Traceability systems can be set up to increase transparency in the supply chains (Meuwissen et al., 2003). McKean (2001) stated that the information must be transferred throughout the chain and properly identified to the appropriate food products. The research also stated that continued development of electronic data storage and management makes extended traceability activities possible and increasingly cost effective. One of the basic prerequisites of both internal and chain traceability is the unique identification of raw materials, semi finished products and finished products (Senneset et al., 2007). As the basis for chain traceability, the identities of traceable units must be recorded at reception and shipping as shown in Figure 2.

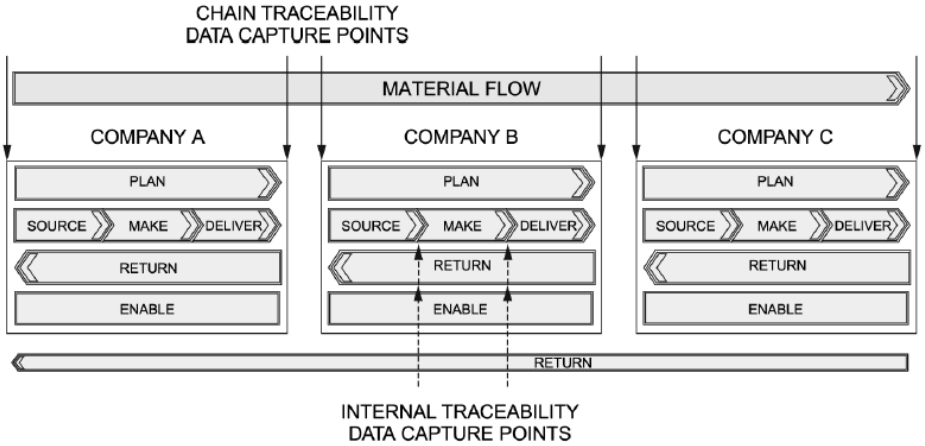


Figure 2. Location of traceability data points (Senneset et al., 2007).

3 Concept of a Traceable Unit

The concept of a traceable unit (TU) was first introduced by Kim et al. (1999) where a TRU was defined as a batch of any resource. Under the TRACE project, a TU can be defined as any item upon which there is a need to retrieve predefined information and that may be priced, or ordered, or invoiced at any point in a supply chain. In practice, it refers to the smallest unit that is exchanged between two parties in the supply chain (TraceFood Wiki, 2009). Each traceable unit must be uniquely identified. In order to capture and retrieve traceability information when required, this information must be associated with a uniquely identified TU (Thakur and Donnelly, 2010).

4 TraceFood Framework

The TraceFood Framework developed under the European Commission sponsored TRACE project provides a toolbox with principles and guidelines for how to implement electronic chain traceability. The framework consists of the following components (TraceFood Wiki, 2009):

(a) Principle of unique identifications
(b) Documentation for joining and splitting (transformations) of units
(c) Generic language for electronic exchange of information
(d) Sector-specific language for electronic information exchange
(e) Generic guidelines for implementation of traceability
(f) Sector-specific guidelines for implementation of traceability

Based on this framework, the implementation of chain traceability requires industry analysis to understand the material flow, information flow and information handling practices. Using this method, based on the industry analysis, recommendations can be provided for new sector-specific data terminology and what information needs to be recorded by each link and communicated to other links in the chain. To enable effective, electronic information exchange, work needs to be carried out on a sector-specific level. Analysis of what product information the particular food sector already records should be carried out and a method and format for identifying this product information should be developed in a standard form (Donnelly, 2009). The need for such systems has already been identified throughout the food industry, but particularly in areas where the authenticity of a product is in question. The viability of such non-proprietary standards were shown in the TraceFish project (CEN 14659, 2003; CEN 14660, 2003; Denton, 2003) where both sector-specific standards (for captured fish and farmed fish) and generic standards (for electronic coding and request-response scheme) were developed. The TraceFish work established sector-specific data models that not only contain information about data elements (including the relationship between them) relevant for product information in one link of the supply chain, but also information for each link. Standardized lists for data elements which can be included in data models have been acknowledged as a key technology for resolving semantic heterogeneity and are important in knowledge management in large organizations (FAO AGROVOC, 2006; Haverkort, 2007; Haverkort, 2006; Stuckenschmidt, 2003).

5 Data management strategies

A wide range of systems are available for traceability in the food industry, ranging from paper-based systems to IT enabled systems (FSA, 2002). Several papers (Karkkainen, 2003; Bechini et al., 2008; Sahin et al., 2002) discuss the use of radio frequency identification (RFID) from a pure supply-chain management point-of-view presenting possibilities for maintaining chain traceability through automatic data capture and exchange/sharing through different suitable solution architectures, middleware and/or electronic product code information services (EPCIS) with discovery services added. The RFID technology is also used to develop traceability systems in food supply chains (Natsui and Kyowa, 2004). Jones et al. (2004) stated that the main reason for RFID diffusion is the capability of tags to provide more information about products than traditional barcodes. Prater et al. (2005) discussed the main benefits of RFID and the EPCglobal network adoption for supply-chain processes, for the specific case of the grocery retailing. The availability of real-time information is regarded as the main benefit, although additional outcomes can be found in increased inventory visibility, stock-out reduction, real-time access and update of current store inventory levels, automated proof of delivery (Fernie, 1994), availability of accurate points of sale data, reduction of labor associated with performing inventory counts of shelved goods, improved theft prevention and shrinkage, and better control of the whole supply chain (Bushnell, 2000). EPC and RFID seem to be a cost-effective way to enable control of flow of goods between the actors in the value chain thus complying with the EU Food law (Official Journal of European Communities, 2002). Bottani and Rizzi (2008) assessed the impact of RFID and EPC system on the main processes of the fast moving consumer goods supply chain. Senneset et al. (2007) claim, however, that to enable transparent electronic traceability through a company, it is necessary to provide records of all transformations within a company, i.e., internal traceability information.

Myhre et al. (2009) outlined the general idea of using EPCIS as a system for collecting traceability information and described how a relationship between one and many traceable items that are tightly connected (such as mixed or blended) can be described by recording every join of many items into a transaction event, and similarly recording each split into another transaction event. This enables both the traditional logistical flow and the transformations (mixing and splitting) of the products along the value chain. Information management and database management techniques are also used for developing traceability systems. Niederhauser et al. (2008) presented a conceptual information system for tracking specialty coffee. It has been shown that the efficiency of a traceability system depends on its ability to record and retrieve the requested lot-related information (Folinas et al., 2006).

5.1 Standardization of Information

One of the biggest challenges with supply chain traceability is the exchange of information in a standardized format between various links in the chain (Thakur and Donnelly, 2010). To facilitate electronic interchange of such product information, international, non-proprietary standards are required such as the ones highlighted by Jansen-Vullers et al. (2003). Folinas et al. (2006) stated that standards must describe how information can be constructed, sent and received and also how the data elements in the information should be identified, measured, interpreted and stored. Previous studies have shown that there is currently no standardized way of formatting information for exchange in traceability systems. Research suggested that structured data lists, vocabularies and ontology will be appropriate tools in achieving effective universal data exchange (Donnelly et al. 2009, Dreyer et al., 2004; TRACE 2, 2008). Individual companies have made great progress in proprietary technologies for automated data capture and electronic data coding. However the benefit of these is lost when the data element transmission is required for use outside the originating company as it is only effective when there is an identical software system at the receiving end (Donnelly, 2008).

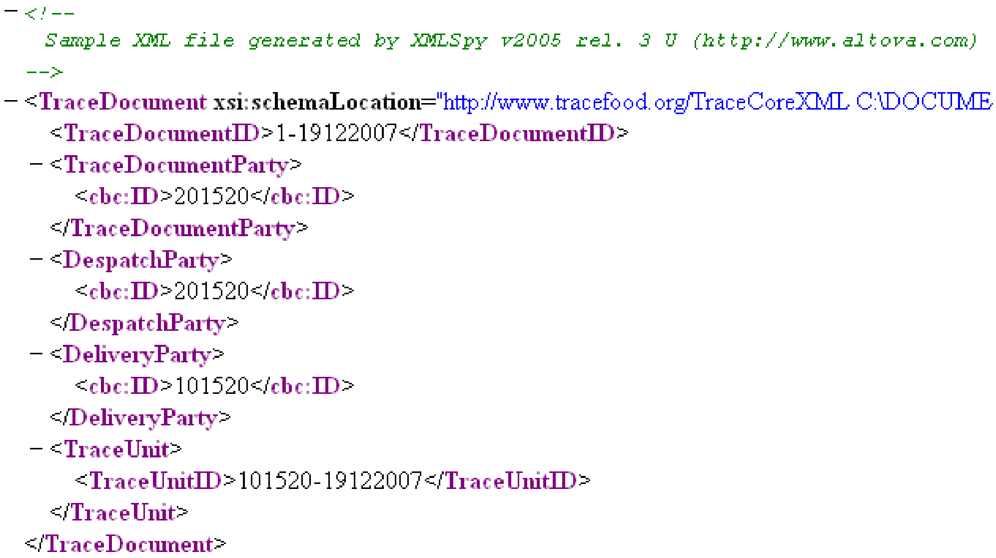
5.2 Traceability Information Exchange

Electronic Data Interchange (EDI) is commonly used in the B2B (Business-to-Business) environment as a reliable mode for electronic data exchange between business and trading partners and presents a set of standards for structuring information that is to be electronically exchanged between and within business organizations and other groups

(Electronic Data Interchange, 2009). EDI implies a sequence of messages between two parties, either of whom may serve as originator or recipient. The effectiveness of using EDI has been widely investigated and it is evident that the standard can be used efficiently by organizations with mature IT capabilities but that is generally not the case for all actors in the supply chain (Bechini, et al., 2008). On the other hand, the increasing popularity of XML (Extensible Markup Language) for information interchange has made it easy for businesses of any size to use this technology. The main purpose of XML is to facilitate the sharing of structured data across different information systems, particularly via the internet. Both EDI and XML formats are structured to describe the data they contain. The main difference is that the EDI structure has a record-field-like layout of data segments and elements; which makes the EDI file shorter, but not easily understandable. An XML document is a tree of nested elements, each of which can have zero or more attributes. There can only be one root element. Each element has a starting and ending tag, marked by angle brackets, with content in between, like: *<element>…content…</element>*. The content can contain other elements, or can consist entirely of other elements, or can be empty. Attributes are named values which are given in the start tag, with the values surrounded by single or double quotations, like: *<element attribute1="value1" attribute2="value2">* (Anderson, 2004).

5.3 TraceCore XML

The European Commission funded the TraceFood framework that is based on the work done in the EU projects TRACE, SEAFOODplus and TraceFish (TraceFood Wiki, 2009). TraceFood is a system for traceability and consists of principles, standards and methods for implementation of traceability in food industry. TraceCore eXtensible Markup Language (TCX) developed under this project is a standard way of exchanging traceability information electronically in the food industry. TCX makes it possible to exchange the information that is common for all food products, like the identifying number, the origin, how and when it was processed, transported and received, the joining and splitting of units, etc (TraceFood, 2007). The TraceCore XML standards can be adapted to various food supply chains where all actors can exchange information using this standard. Figure 3 shows a sample XML file used to exchange traceability information between dispatch party and a delivery party. The XML file identifies the document, parties involved and the trace units.



```
- <!--
    Sample XML file generated by XMLSpy v2005 rel. 3 U (http://www.altova.com)
  -->
- <TraceDocument xsi:schemaLocation="http://www.tracefood.org/TraceCoreXML C:\DOCUME
    <TraceDocumentID>1-19122007</TraceDocumentID>
  - <TraceDocumentParty>
      <cbc:ID>201520</cbc:ID>
    </TraceDocumentParty>
  - <DespatchParty>
      <cbc:ID>201520</cbc:ID>
    </DespatchParty>
  - <DeliveryParty>
      <cbc:ID>101520</cbc:ID>
    </DeliveryParty>
  - <TraceUnit>
      <TraceUnitID>101520-19122007</TraceUnitID>
    </TraceUnit>
  </TraceDocument>
```

Figure 3. Sample XML file for traceability (TraceFood, 2007)

6 Traceability optimization

One mechanism used to prevent the consumption of contaminated products is a product recall, implemented by the company that created the problem and tracked by the government and both the frequency and severity of food contamination are increasing (Skees et al., 2001). For the food industry, the emphasis is not only to decrease the food safety incidents (and recalls) but also limit the number of batches that constitute a given finished product in order to decrease the product quantities to be recalled (Dupuy, et al., 2005). Gattengo (2001) stated that after a recall of minced beef products due to BSE, a French producer not only improved the accuracy of their traceability system but also decreased the number of mixed batches of meat in one batch of minced beef. Dupuy et al. (2005) proposed a batch dispersion model to optimize traceability in food industry by minimizing the batch size and batch mixing. This model calculates the minimum batch dispersion which is given by the sum of links between the raw material batches and the finished product batches. However, the problem of incurring additional cost by minimizing batch dispersion has not been addressed in existing literature.

7 Sector specific traceability research

TraceFood framework states that there is a need to develop sector-specific traceability standards and information exchange guidelines (TraceFood Wiki, 2009). Several research studies have been conducted for developing sector specific traceability standards and implementing various data management and information exchange techniques in various product supply chains. Regattieri e al. (2007) proposed a general framework for a traceability system and showed its application for Parmigiano Reggiano cheese based on an integration of alphanumeric codes and RFID technology. Donnelly et al. (2008) presented a methodology for creating standardized data lists for traceability in honey processing industry by conducting multi-stage surveys in the honey processing chain. The resulting standardized list of data elements could be used by all honey processors. Randrup et al. (2008) studied the effectiveness and accuracy of chain traceability systems by conducting simulated recalls of fish products in retail shops in five Nordic countries. The study found that improved traceability practices in the whole chain can limit the batch sizes and minimize costs in case of food recalls. Shanahan et al. (2009) presented a system identify all aspects of beef traceability from farm to slaughter based on the European Union law and global standards. They proposed an integrated traceability system involving all of the stakeholders along the supply chain with the use of RFID for identification of individual cattle, and biometric identifiers for verification of cattle identity. Donnelly et al. (2009) conducted a study to track and trace lamb meat through a lamb meat processor where improvements to the current traceability system were suggested after identifying all critical traceability points.

References

Anderson T., 2004. Introducing XML, <http://www.itwriting.com/xmlintro.php>

Bechini A., Cimino M.G.C.A., Marcelloni F., Tomasi A., 2008. Patterns and technologies for enabling supply chain traceability through collaborative information e-business, *Information and Software Technology* 50, 342-359.

Becker, T., 2000. Consumer perception of fresh meat quality: a framework for analysis. *British Food Journal* 102(3), 158–176.

Bottani, E., Rizzi, A., 2008. Economical assessment of the impact of RFID technology and EPC system on the system on the fast-moving consumer goods supply chain. *International Journal of Production Economics* 112(2), 548-569.

Bushnell, R., 2000. RFID's wide range of possibilities, *Modern Materials Handling* 55(1), 37.

CEN14659, 2003. *CEN Workshop Agreement*. Traceability of Fishery products. Specification of the information to be recorded in caught fish distribution chains. European Committee for Standardization.

CEN14660, 2003. *CEN Workshop Agreement*. Traceability of Fishery products. Specification of the information to be recorded in farmed fish distribution chains. European Committee for standardization.

Denton, W., 2003. TraceFish: The development of a traceability scheme for the fish industry. In: *Luten, J.O., Olafsdottir, G. (eds.) Quality of fish from catch to consumer*, 75–91. Wageningen Academic Publishers, Wageningen.

Donnelly, K.A., Karlsen, K.M., Olsen, P., 2009. The importance of transformations of traceability – A case study of lamb and lamb products. *Meat Science* 83, 68-73.

Donnelly, K.A.-M., Karlsen, K.M., Olsen, P., van der Roest, J., 2008. Creating Standardized Data Lists for Traceability – A Study of Honey Processing. *International Journal of Metadata, Semantics and Ontologies* 3(4), 283-291.

Dreyer, C., Wahl, R., Storøy, J. and Forås, O.P., 2004. 'Traceability standards and supply chain relationships', in Aronsson, H. (Ed.): *Proceedings of the 16th Annual Conference for Nordic Researchers in Logistics, NOFOMA 2004*, Challenging Boundaries with Logistics, Linköping, Sweden, 155–170.

Dupuy, C., Botta-Genoulaz, V., Guinet, A., 2005. Batch dispersion model to optimize traceability in food industry. *Journal of Food Engineering* 70(3): 333-339.

Electronic Data Interchange, 2009. *Wikipedia*, <http://en.wikipedia.org/wiki/Electronic_Data_Interchange>

EVIRA, 2007. *Finnish Food Safety Authority Evira* (Elintarviketurvallisuusvirasto). <http://www.evira.fi/portal/fi/elintarvikkeet/>

FAO AGROVOC, 2006. < www.fao.org/agrovoc>

Fernie, J., 1994. Quick response: An international perspective, *International Journal of Physical Distribution & Logistics Management* 24(6), 38–46.

Fisk, G. Chandran, R., 1975. Tracing and recalling products. *Harvard Business Review* November–December, 90-96.

Florence, D., Queree, C., 1993. Traceability—Problem or Opportunity. *Logistics Information Management* 6 (4), 3-8.

Folinas, D., Manikas, I., Manos, B., 2006. Traceability data management for food chains. *British Food Journal* 108 (8), 622–633.

FSA, 2002. Traceability in the Food Chain – A Preliminary Study. *Food Chain Strategy Division*, Food Standards Agency.

Gattegno, I. Soviba, ce qui a change´ depuis le 20 octobre 2000. 2001 *RIA (Revue de l'Industrie Agroalimentaire)* 609: 46–47.

Golan, E., Krissoff, B., Kuchler, F., 2004. Food traceability: one ingredient in a safe and efficient food supply, economic research service. *Amber Waves* 2, 14–21.

Gunasekaran, A., Ngai, E.W.T., 2004. Information systems in supply chain integration and management. *European Journal of Operations Research* 159, 269–295

Haverkort, A., 2007. The Canon of Potato Science: 36. Potato Ontology. *Potato Research* 50(3), 357–361.

Haverkort, A., Top, J., Verdenius, F., 2006. Organizing Data in Arable Farming: Towards an Ontology of Processing Potato. *Potato Research* 49(3), 177–201.

International Organization for Standardization, 2007. New ISO Standard to Facilitate Traceability in Food Supply Chains. *ISO 22005:2007.*

Jansen-Vullers, M.H., van Dorp, C.A., Buelens, A.J.M., 2003. Managing traceability information in manufacture. *International Journal of Information Management* 23, 395–413.

Jones, C. Clarke-Hill, P. Shears, Comfort, D., Hillier, D., 2004. Radio frequency identification in the UK: Opportunities and challenges, *International Journal of Retail and Distribution Management* 32(3), 164–171.

Karkkainen, M., 2003. Increasing efficiency in the supply chain for short shelf life goods using RFID tagging. *International Journal of Retail and Distribution Management* 31, 529-536.

Kim, H.M., Fox, M.S., Grüninger, M., 1999. An ontology for quality management enabling quality problem identification and tracing. *BT Technology Journal*, 17(4), 131-140.

Martínez-Sala, A. S., Egea-López, E., García-Sánchez, F., García-Haro, J., 2009. Tracking of Returnable Packaging and Transport Units with active RFID in the grocery supply chain. *Computers in Industry* 60, 161-171.

McKean, J.D., 2001. The importance of traceability for public health and consumer protection. *Revue Scientifique Et Technique (International Office of Epizootics)* 20 (2), 363–371.

Meuwissen, M.P.M., Velthuis, A.G.J., Hogeveen, H., Huirne, R.B.M., 2003. Traceability and Certification in Meat Supply Chains, *Journal of Agribusiness* 21(2), 167-181.

Moe, T., 1998. Perspectives on traceability in food manufacture. *Trends in Food Science & Technology* 9(5), 211–214.

Myhre, B., Netland, T.H., Vevle, G., 2009. The footprint of food - A suggested traceability solution based on EPCIS. In the *5th European Workshop on RFID Systems and Technologies*(RFID SysTech 2009), Bremen, Germany.

Natsui, T., Kyowa, A., 2004. Traceability System using RFID and Legal Issues. *WHOLES*, A multiple view of individual privacy in a networked world, <www.sics.se/privacy/wholes2004/papers/takato.pdf>

Niederhauser, N., Oberthür, T., Kattnig, S., Cock, J., 2008. Information and its management for differentiation of agricultural products: The example of specialty coffee. *Computers and Electronics in Agriculture* 61(2), 241-253.

Official Journal of the European Communities. 2002. *Regulation (EC) No 178/2008* of the European Parliament and the Council of 28 January 2002.

Opara, L.U., Mazaud, F., 2001. Food traceability from field to plate. *Outlook on agriculture* 30, 239-247.

Peri, C., 2002. Rintracciabilita` della filiera dei prodotti agroalimentare: significato, metodi e strumenti. Atti del Convegno ''Rintracciabilita`di filiera per una provincia trasparente'', Cuneo. <http://www.think-quality.it>

Prater, E., Frazier, G.V., Reyes, P.M., 2005. Future impacts of RFID on e-supply chains in grocery retailing, *Supply Chain Management: An International Journal* 10(2), 134–142.

Randrup, M., Storøy, J., Lievonen, S., Margeirsson, S., Arnason, S.V., O´ lavsstovu, D., Møller, S.F., Frederiksen, M.T., 2008. Simulated recalls of fish products in five Nordic countries, *Food Control* 19, 1064–1069

Regattieri A., Gamberi M., Manzini R., 2007. Traceability of food products: General framework and experimental evidence, *Journal of Food Engineering*, 81, 374-356.

Sahin, E., Dallery, Y., Gershwin, S., 2002. Performance evaluation of a traceability system. An application to the radio frequency identification technology. In *Systems, Man and Cybernetics*, 2002 IEEE International Conference.

Senneset, G., Forås, E., Fremme, K.M., 2007. Challenges regarding implementation of electronic chain traceability. *British Food Journal*, 109(10), 805-818.

Shanahan, C., Kernan, B., Ayalew, G., McDonnell, K., Butler, F., Ward, S., 2009. A framework for beef traceability from farm to slaughter using global standards: An Irish perspective, *Computers and electronics in agriculture* 66(1) 62–69.

Skees, J.R., Botts, A., Zeuli, K.A., 2001. The potential for recall insurance to improve food safety. *International Food and Agribusiness Management Review* 4, 99-111.

Stuckenschmidt, H., 2003. Ontology based information in dynamic environments. In: *Twelfth IEEE Internation Worskshops on Enabling Technologies*: Infrastructures for Collaborative Enterprises(WETICE 2003).

Thakur, M., Donnelly, K.A.-M., 2010. Modeling traceability information in soybean value chains. *Journal of Food Engineering* 99(1), 98-105.

TRACE 2, 2008. *Annex I – TRACE – Tracing Food Commodities in Europe 'Description of Work'*, FP6-2003-FOOD-2-A Proposal No 006942, Sixth Framework Programme.

TraceFood Wiki, 2009. <http://www.tracefood.org>

TraceFood, 2007. TraceCore – XML Standard Guidelines, *TraceFood*, <http://193.156.107.66/ff/po/TraceFood/TraceCore%20XML.htm>

Wall, B., 1994. Quality management at Golden Wonder. *Industrial Management and Data Systems* 94(7), 24–28.

Williamson, E., Harrison, D.K., Jordan, M., 2004. Information systems development within supply chain management. *International Journal of Information Management* 24, 375–385.

.

CHAPTER 3. Framework for implementing traceability in the bulk grain supply chain

Published in *Journal of Food Engineering* (2009), 95(4):617-626.

Maitri Thakur[1,2,*] and Charles R. Hurburgh[1, 2]

[1]Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011
[2] Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA 50011
[2]Department of Food Science and Human Nutrition, Iowa State University, Ames, IA 50011
* Primary author, Corresponding author

Abstract

Implementation of a traceability system in the bulk grain supply chain is a complex task. Grain lots are often commingled to meet buyer specifications and the lot identity is not maintained. In this paper, a systems approach is used to develop methods for implementing bulk grain supply chain traceability in the United States, that includes both internal and chain traceability. First, the usage requirements of a traceability system are defined for all the actors in the supply chain. Second, a model is developed for implementing internal traceability system for a grain elevator that handles specialty grain. Then, we develop a model for information exchange between the supply chain actors. The model shows what grain lot information must be recorded and then passed on to the next actor. A sequence diagram is developed to show the information exchange in the grain supply chain when a user requests additional information about a suspect product. Finally, we discuss some suitable technologies to enable this information exchange. A few sample XML documents are shown for the transfer and sharing of information in the grain supply chain.

Keywords: Supply chain traceability; Internal traceability; Bulk grain; Information Exchange; Framework

1 Introduction

The agricultural sector has undergone considerable change during the past century. New farming practices as well as new handling and processing techniques have been developed to meet the increasing consumer demand for reliable and consistently safe supply of various food products. Furthermore, consumers are giving emphasis to safety, high quality and sustainability of food products. Consumer experiences with food safety and health issues combined with an increasing demand for high quality food and feed products have resulted in an increasing interest in developing systems to aid in food traceability efforts. Traceability in the food supply chains has gained considerable importance in the past few years (Carriquiry and Babcock, 2007; Folinas, et. al., 2006; Jansen-Vullers, et. al., 2003; Madec, et. al., 2001; McKean J.D., 2001). Various food safety and traceability laws exist in several countries. European Union's General Food Law entered into force on January 1, 2005. The law included important elements like rules on traceability and the withdrawal of dangerous food products from the market. Under the European Union Law, "Traceability" is defined as the ability to track any food, feed, food-producing animal or substance that will be used for consumption, through all the stages of production, processing and distribution (Official Journal of the European Communities, 2002). It is a risk-management tool that allows food business operators or authorities to withdraw or recall products which have been identified as unsafe.

In the United States, after the September 11 events, the US Public Health Security and Bioterrorism Preparedness and Response Act of 2002 (the Bioterrorism Act) was passed. The Bioterrorism Act requires that all companies involved in the food and feed industry to self-register with the Food and Drug Administration and maintain records and information for food traceability purposes (U.S. Food and Drug Administration, 2002). In Canada, federal, provincial, and territorial Ministries of Agriculture agreed on a landmark agreement, entitled the Agriculture Policy Agreement (APF) in 2003. APF has five objectives including food safety and food quality. Can-Trace was launched in July 2003 which is a collaborative and open initiative committed to the development of traceability standards for all food products sold in Canada (Can-Trace, 2003). The mission of Can-Trace is to define and develop minimum requirements for national whole-chain tracking and tracing standards based on the GS1 system. The GS1 Global Traceability Standard is a business process standard that describes the traceability process independently, in terms of key operations for any choice of enabling data management technologies.

Traceability is important for many reasons like responding to the food security threats, documenting chain of custody, documenting production practices, meeting regulatory compliance or analyzing logistics and production costs. USDA Economic Research Service states that besides ensuring a safe food supply, use of a traceability system results in lower cost distribution systems, reduced recall expenses, and expanded sales of products with attributes that are difficult to discern (Golan et. al., 2004). In every case, the benefits of traceability translate into larger net revenues for the firm. Thus, food traceability has become important for reasons other than just the legal obligations in several countries. The ISO 22005 Food Traceability Standard requires that each company know who their immediate supplier is and to whom the product is being sent, on the principle of one up and one down. It states that food safety is the joint responsibility of all the actors involved (International Organization for Standardization, 2007). Thus, all the actors involved in the food supply chain are required to store necessary information related to the food product that link inputs with outputs, so that when demanded, the information can be provided to the food inspection authorities on a timely basis. In order to achieve a fully traceable supply chain, it is important to develop systems for chain traceability as well internal traceability. This includes linking, to the best extent possible, units of output with specific units of input. Each supply chain actor should have an internal record keeping system that would enable them to trace back their ingredients and track forward the products so as to determine the cause of the problem or to efficiently recall the associated (or contaminated) food products. Each actor must be able to trace back and track forward the product information based on one-up and one-down basis.

Developing a traceability system is however, a complex undertaking as it involves all the stages of production, handling, storage, processing, transportation, and distribution. The next section describes the bulk grain supply chain in the United States.

1.1 **Bulk Grain Supply Chain in the United States**

Agricultural supply chains are unique in the sense that they include many different commodities that are grown in different regions at different time periods of the year, and are transported through different modes. Agricultural commodities have different end uses such as food, feed, industrial and energy and are relatively homogenous. They are transported and stored in bulk quantities which range from hundreds to several thousand metric tons (Nardi et. al., 2007). Figure 1 shows a typical bulk grain supply chain in the United States. A typical bulk grain supply chain in the United States starts from a seed company. The farmers buy seeds from a seed company and after harvesting, sell their crop to a grain elevator. The grain elevators handle bulk commodities marketed against generic grade standards that are based on physical attributes. Grain lots are commingled in order to meet buyer specifications and to maximize the profit. As a result of this commingling, lot identity is not maintained. Grain storage bins are extensively used to handle bulk grain and one storage bin can contain grain from many different sources. The elevators either sell the grain directly to a processor or ship it to a river terminal for overseas export. In case of an overseas export, the river terminal sells the grain to an export terminal which sells the grain to an overseas terminal. These terminals handle the grain in a similar fashion as an elevator. The grain lots are commingled to maximize profit and lot identity is not maintained. As shown in figure 1, an overseas export adds additional actors to the supply chain. The grain handlers (an elevator or an overseas importer) sell the grain to an ingredient processor. At the ingredient processing plant, the grain is processed into a final product with addition of other ingredients. Grain lots are commingled again and the finished product can contain grain from many different sources. The ingredient processor sells its product to the final processor where this product is used to manufacture the final product with addition of other products and ingredients while undergoing many processing steps. The final product is sold to the distributor and finally to the retailer for sale to the customer.

Figure 2 shows a typical scenario for grain aggregation and segregation that takes place at any stages in the supply chain. The figure also shows that how one contaminated lot can contaminate many other grain lots. Internal records are generally not maintained for the aggregation and segregation of grain lots. In case of a food related emergency, it would be almost impossible to isolate the source with the problem which would lead to a recall of all the finished goods that might have a chance of being contaminated. Many food recall incidents have taken place in the past that have affected the consumers and the producers alike. For instance, according to a news report, after the tomato-salmonella scare in June 2008, the Florida tomato industry could have potentially lost $40 Million because the producers could not sell their tomatoes until the source of salmonella outbreak was identified (Reuters, 2008). With fragile and quickly perishable items like tomatoes, the consequences on industry and growers/producers can be irreparable. The grain trade units must be tracked efficiently from the farm to the consumer to avoid such problems.

1.2 **Tracking and Tracing**

The terms “tracking” and “tracing” are very commonly used to describe traceability. Tracking (forward) is the ability to follow the downstream path of a particular trade unit in the supply chain, while, tracing (backward) is the ability to identify the origin of the products used in a particular trade unit. Thus, tracking is a top down approach and tracing is a bottom-up approach. Both, tracking and tracing play a very important role in the overall supply chain traceability. According to Van Dorp (2002), tracking and tracing provides the visibility to where work is at all times and its disposition and a tracking function creates a historical record by means of recorded identification that allows for the traceability of components and the usage of each end product. A good traceability system should have the capability of performing both functions efficiently. Laux (2007) demonstrated that tracing (backward) was harder than tracking (forward) for an elevator handling commodity grain.

1.3 **Supply Chain Traceability**

Effective supply chain traceability can only be achieved with a combination of internal traceability and chain traceability. Each actor in the supply chain must not only know who their supplier is, but also to whom the trade units are being sold. Opara (2003) states that in order to implement traceable agricultural supply chains, technological innovations are needed for product identification, process and environmental characterization, information capture, analysis, storage and transformation, as well as overall system integration. Regattieri et. al. (2007) state that a food traceability system is fundamentally based on four pillars of product identification, data to trace, product routing and traceability tools. Determining the requirements of a grain supply chain traceability system is the most important step before data modeling tools can be used. The traceability literature lacks in research on developing methodology for implementation of internal and chain traceability in food supply chains. In this paper, we present a systematic approach for implementing traceability in a bulk grain supply chain by using the business process integration tools including system requirements planning, enterprise modeling and integration. The objective of this paper is to develop a framework for implementing traceability in the bulk grain supply chain in the United States that to facilitate both internal and chain traceability. First, we define the usage requirements of the traceability system from each actor involved in the grain supply chain. Next, we develop an IDEF0 model for developing and implementing an internal traceability system at a grain elevator. Then, we discuss how to implement chain traceability based on information exchange among supply chain actors. Finally, we provide some conclusions and directions for future work.

2 Usage requirements of the Traceability System

According to Folinas et. al. (2006), an integrated traceability system must be able to file and communicate information regarding product quality, origin, and consumer safety. In order to design an efficient grain traceability system, the first step is to define the usage requirements for the grain supply chain. A system-level approach is used to develop models for implementing the traceability system. The usage requirements of the traceability system are defined by the UML (Unified Modeling Language) Use Case diagram technique (Eriksson and Penker, 2000). The Use Case diagrams are closely connected to scenarios. A scenario is an example of what happens when someone interacts with the system. One of the most important goals of defining system requirements is to have synchronization among the requirements of all actors involved. A Use Case diagram depicts the following (Miller, 2003):

- **Use cases**: A use case describes actions that provide something of measurable value to an actor and is drawn as a horizontal ellipse.
- **Actors:** An actor is a person or organization that plays a role in one or more interactions with the system. The actors are drawn as stick figures.
- **Associations:** Associations between actors and use cases are indicated in use case diagrams by solid lines. An association exists whenever an actor is involved with an interaction described by a use case.
- **System boundary:** A rectangle can be drawn around the use cases, forming the boundary and is called the system boundary box. The boundary indicates the scope of the system.

Lee and Xue (1999) state that an important advantage of Use Case driven analysis is that it helps manage complexity, since it focuses on one specific usage at a time. Figure 3 shows the Use Case diagram for the grain supply chain traceability system. The following use case examples are defined and different actors are associated with each use case:

- **Record breeding practices:** The seed company would record the seed development practices used in the traceability system. For example: genetically modified, organic practices, etc.
- **Record farming practices:** The farmer would record the farming practices used for a specific crop in the system. The data such as the seed variety used, date of planting, chemical application, harvesting, etc. would be recorded. The information such as organic practices would be recorded for specialty crops.

- **Record handling and storage practices:** The supply chain actors should be able to record the handling and storage practices used by them in the system.
- **Record processing practices:** The processor should be able to record the processing practices used in the system. Depending on the process and final product, this may include the cooking temperature, holding time, ingredients added, etc.
- **Authenticate claims:** The system users (supply chain actors) should be able to authenticate their claims based on the data stored in the system. For example, on request, the system should be able to provide data to support organic farming or processing practices.
- **Comply with food safety regulations:** Using the traceability system, within the time requirements provided, the users should be able to provide data to show that their production or processing practices comply with the food safety regulations. For example, a processor must be able to show that the processing conditions used to manufacture a product (temperature, holding time, etc) are in compliance with the food safety regulations. This data must be recorded in the traceability system and provided on demand by regulatory authorities.
- **Protect integrity of brand name:** The system users should be able to protect the integrity of their brand name by using the data stored in the traceability system. If the processor claims that their products are organic, there must be data recorded and available to back that claim.
- **Document chain of custody:** On request, the traceability system should be able to provide information about a specific trade unit that would document the chain of custody of that unit. In case of a food safety emergency, it is very important to know where a particular trade unit is in the supply chain at a given time.

3 Internal Traceability

Internal traceability plays a very important role in supply chain traceability. In order to develop systems for internal traceability, the Integrated Definition Modeling (IDEF0) technique is used in this work. IDEF0 is a common modeling technique for the analysis, development, re-engineering, and integration of information systems, business processes, or software engineering analysis. IDEF0 is capable of graphically representing a wide variety of business, manufacturing and other types of enterprise operations to any level of detail (Department of Defense, 2001). IDEF0 is a method designed to model the decisions, actions, and activities of an organization or system (IDEF0, 1993). The model consists of inputs, outputs, controls, and mechanisms for a process or function. IDEF0 is a hierarchical model with a tree structure where the parent process consists of many sub-processes. The first step in the IDEF0 process is identification of the prime function or process to be decomposed. Figure 4 shows a generic IDEF0 model. Figure 5 shows an IDEF0 model for developing an internal traceability system at a grain elevator. The necessity of developing a traceability system originates from the regulatory need. As discussed before, several traceability laws and regulations exist in different countries. So, the regulatory need is a driving force for development of a traceability system. Similarly, the food industry has to constantly adapt according to their business needs. If the elevator company deals with specialty grain, then it is a business requirement for them to segregate the specialty grain from other grains. The business need in turn stems from the customer needs or preferences. Thus, the regulatory need, business need and the customer preferences are categorized as the model inputs. The traceability system should be developed in compliance with any regulatory requirements. So, the regulatory compliance is also a control for this model. Various mechanisms are needed to develop this traceability system, such as industry standards, personnel and procedures. The desired outputs would depend on the type of product and the supply chain actor. In general, various documentations such as production practices, validation certificates, safety and quality assurance would be the desired outputs of the traceability system. The system must also be able to authenticate a company’s claims such as organic products, and also provide a measure for customer satisfaction. These would be the desired outputs of a traceability system.

The model is decomposed to show all the steps involved. The model is adapted for a grain elevator that handles specialty grain and is looking to obtain food safety management systems certification, such as ISO 22000. Obtaining an ISO certification becomes an input for the traceability system in this case. Figure 6 shows this decomposed IDEF0 model. Different steps involved in the development of a traceability system are represented in a sequence. Inputs, outputs, controls and mechanisms at each stage are shown.

(1) Determine traceability plan: The first step in developing an internal traceability system is the determination of the traceability plan by the grain elevator. The inputs of this step are the regulatory need, which is obtaining the ISO 22005 compliance;

segregation of specialty grain since the elevator handles specialty grain; and the consumer demand for specialty grain. The traceability plan is to be determined based on these requirements. The ISO 22005 standard is the control for this step and various mechanisms are needed to determine the traceability plan, such as industry standards, personnel and procedures. The personnel for the traceability team should be selected from a variety of different backgrounds and departments within the elevator company. The traceability plan should be clearly defined in a consistent format and should include information such as what data needs to be recorded and shared with other actors in the supply chain. It should also define the measures of success and the precision required. The output of this process is a traceability system manual that defines the procedure for implementing the traceability plan.

(2) Implement traceability plan: The output from process 1 is the input for this step. The traceability system manual is be used to implement the plan. This process has the same control and mechanisms as process 1. A relational database management system is used to implement the traceability plan. There is only one database for all the grain related information. The users can enter the relevant grain data in the database system. Both lot quality and lot activity data corresponding to a grain lot must be recorded. The relational database system connects the data about incoming grain lots, the internal lot activities and the outgoing grain lots. Since, grain acts like a fluid; it is very difficult to define the lot sizes. Traceability in terms of grain movements within the elevator and blending for customer shipments is more important than identification of lots. After this step is complete, an implementation report would be generated. This report would consist of a detailed description of the database system and its use.

(3) Evaluate system performance: The performance of the traceability system would be evaluated in this process. This would consist of evaluating the performance of the traceability database in terms of the efficiency of the system to react rapidly in a food safety crisis. The performance reports and audit reports are the output of this step. This step has the same control and mechanisms as the previous steps.

(4) System validation: Validation is required to ensure that the system is performing as defined by the traceability plan. The performance reports and audit reports from step 3 are used to validate the traceability system using the same ISO 22005 standard as the control and the same mechanisms that are used in the previous processes. The system validation would generate various documentations for this process. After the traceability system has been validated, the ISO 22005 compliance can be achieved. Other documentations for production practices, Quality Management Systems and system validation certificates can be generated. Proof of customer satisfaction would also be a desired output of the traceability system development process.

(5) System maintenance: Maintenance of the traceability system is a crucial step in the whole process. Maintenance is required to keep the system functional and for continuous improvement. This is a continuous process and the traceability plan should be modified according to the changes in regulations, customer demands or any other factors that cause a change in the business process. The subsequent steps would need to be carried out again every time there is a change in the traceability plan.

Developing such models can give the organization an overview of various steps that are required to accomplish the task of developing and implementing a traceability system.

4 Chain Traceability through Information Exchange

Although IDEF0 models are good at providing an initial view of activity decomposition, it is incapable of modeling information process flows which is due to the lack of time dependency input (Dorador and Young, 2000). So, there is a need for models to capture the sequence of processes and information flows in a system. Many lot activities take place at various points in the grain supply chain, as described below:

- **Movement:** Grain is moved from one actor in the supply chain to another. For example, farmer sells the grain to an elevator. In an elevator, grain is often moved internally from one storage bin to another due to storage space or other quality constraints.
- **Aggregation:** A grain lot is aggregated with other lots. For example, when an elevator ships the grain to a river terminal, depending on the buyer specification, the outgoing grain lot might come from several different storage bins. So, an outgoing grain lot may contain grain from several storage bins at the elevator.
- **Segregation:** An incoming grain lot is divided into many different grain lots. Incoming grain at an elevator purchased from a farmer is considered as one lot. This grain lot might be divided and assigned to a several different storage bins rather than one bin. This leads to segregation of an incoming grain lot.

- **Storage:** A grain lot can be stored for a certain period of time causing a change in its physical or chemical properties. For example, moisture content could change during storage.
- **Transformation:** A grain lot or a part of it can be used as an ingredient to produce another product, for example, livestock feed.
- **Destruction:** A grain lot or a part of it can be destroyed during a processing operation for various reasons.

It is important to record these activities accurately and pass on the information to the next actor in the supply chain. Figure 7 shows the grain supply chain and the information that should be recorded and passed onto the next link in the supply chain by each actor. It also shows that which information about a grain lot should be passed on to the next actor in the chain. The superscripts link the information that is passed on between supply chain actors. When all the relevant information is recorded and passed on to the next actor, the grain lots and their properties used in the final product can be traced back to the origin. Also, the grain lot from the farm can be tracked forward to the retailer. It can be seen from figure 7 that not all of the information is passed to the next link in the supply chain. However, it is important that all the relevant lot-information is passed to the next link. This information should be sufficient to obtain any additional information as required. As discussed before, there are many lot activities that take place throughout the supply chain. The goal is to achieve supply chain traceability, so it is important that each actor maintains an internal traceability system using a relational database management system. As long as all the lot information is recorded in an RDBMS (Relational Database Management System) form by each actor, retrieval of all necessary information linking individual lots at different points in the supply chain becomes easier. One such internal traceability database has been developed for a grain elevator as a part of this work.

Figure 8 shows a UML sequence diagram for information exchange between supply chain actors. A sequence diagram is used to show the interactions between objects in the sequential order in which the interactions occur. An organization can find sequence diagrams useful to communicate how the business works by showing how various objects interact. The main purpose of this diagram is to define event sequences that result in some desired outcome. The diagram shows what messages are sent between the system’s objects as well as the order in which they occur. It conveys this information along the horizontal and vertical dimensions: the vertical dimension shows, top down, the time sequence of messages as they occur, and the horizontal dimension shows, left to right, the object instances that the messages are sent to (Bell, 2004). The supply chain actors are the object instances for the grain supply chain case.

Figure 7 shows the information that should be shared between the actors in the supply chain, while Figure 8 shows the sequence of this information exchange. It also shows the sequence of events if any additional information is requested about a suspect product. The user can be a regulatory agency in this case. When additional information is requested in case about a product; the companies should provide this information in a timely manner to comply with the regulations. In the United States grain industry, a company has 24 hours to provide this information from the time it is requested.

5 Mode of information exchange

Electronic Data Interchange (EDI) is commonly used in the B2B (Business-to-Business) environment as a reliable mode for electronic data exchange between business and trading partners. EDI is a set of standards for structuring information that is to be electronically exchanged between and within business organizations and other groups. EDI implies a sequence of messages between two parties, either of whom may serve as originator or recipient. The effectiveness of using EDI has been widely investigated and it is evident that the standard can be used efficiently by organizations with mature IT capabilities. This is generally not the case for all actors in the supply chain (Bechini, et al., 2008). On the other hand, the increasing popularity of XML (Extensible Markup Language) for information interchange has made it easy for businesses of any size to use this technology. The main purpose of XML is to facilitate the sharing of structured data across different information systems, particularly via the internet. Both EDI and XML formats are structured to describe the data they contain. The main difference is that the EDI structure has a record-field-like layout of data segments and elements; which makes the EDI file shorter, but not easily understandable. The XML format has tags, which are more easily understood, but make the file bigger and verbose (Electronic Data Interchange Development, 2008). An XML document is a tree of nested elements, each of which can have zero or more attributes. There can only be one root element. Each element has a starting and ending tag, marked by angle brackets, with content in between, like: *<element>…content…</element>*. The content can contain other elements, or can consist entirely of other elements, or can be empty. Attributes are named values which are given in the start tag, with the values surrounded by single or double quotations, like: *<element attribute1="value1" attribute2="value2">* (Anderson, 2004).

The European Commission funded the TraceFood framework that is based on the work done in the EU projects TRACE, SEAFOODplus and TraceFish (TraceFood Wiki, 2009). TraceFood is a system for traceability and consists of principles, standards and methods for implementation of traceability in food industry. TraceCore eXtensible Markup Language (TCX) developed under this project is a standard way of exchanging traceability information electronically in the food industry. TCX makes it possible to exchange the information that is common for all food products, like the identifying number, the origin, how and when it was processed, transported and received, the joining and splitting of units, etc (TraceFood, 2007). The TraceCore XML standards can be adapted to grain supply chain where all actors can exchange information using this standard.

6 TraceCore XML and United States Grain Supply Chain

Figure 9 shows a part of an entity-relationship model developed for implementing internal traceability for a grain elevator in section 4. An XML document is created for every action relating to the grain. The basic elements in the TraceCore XML standard include documentation identification, sender and receiver information, traceability unit identification and traceability relations (TraceFood, 2007). Figure 10 shows the basic structure of an XML document for acquisition of grain by the elevator from the farmer. The entities used here are from the elevator database model shown in figure 9. Figure 10 also shows the tree format of this acquisition notification generated within the elevator system when grain is purchased from the farmer. The schema shows the sender information (farmer in this case), product and origin information, activity information and other quality attributes related to grain. Grain activity in this case refers to receiving grain from the farmer, which is identified by the *scale ticket* number as a unique identifier. The document also includes information regarding storage bin assignment to the grain received.

Sometimes, grain is moved internally in an elevator from one bin to another. Figure 11 shows the basic structure of an XML document for movement notification of grain in the elevator. The tree format of movement notification is also shown. Grain movement from one storage bin to another can be viewed as a transformation or splitting of different lots (one bin being considered as one lot). The *origin* and *destination* bins, *weight* of grain moved as well as *start* and *end time* of the internal movement is included in this document. Quality attributes of the grain lot are also captured similar to the acquisition notification document. These XML traceability documents contain both the lot and activity data. As mentioned before, grain aggregation and segregation takes place at many different stages in the supply chain. Thus, it is very important to record the grain quality data (moisture, test weight, damaged material and foreign material) for each activity type. This data can then used to calculate the quality parameters of the aggregated lots.

7 Conclusions

Implementation of a traceability system in the bulk grain supply chain in the United States is a complex task. Several problems exist at different stages throughout the supply chain. Grain lots are often commingled to meet buyer specifications and lot identity is not maintained. The internal grain movements at grain handling and processing facilities often go unrecorded. In order to achieve traceability goals along the grain supply chain, businesses should focus both on internal and chain traceability. Determination of the usage requirements of the traceability system is the first step in implementing the system. Each supply chain actor should determine their traceability plan based on the driving factors like the regulatory need, business need and the customer preferences. Relational database management system could be used to implement internal traceability system by each actor in the supply chain. All grain lot information should be recorded in a centralized database system and only relevant lot/batch information should be passed on to the next link in the supply chain. Additional information can be requested by the authorized users (such as regulatory agencies) in case of a suspect product. This additional information should be provided in a timely manner. The use of new technologies like XML can be a very powerful tool for e-information exchange between supply chain actors. The use of XML can have several benefits, like reduction of time and effort required for exchanging information. Use of a relational database management system to record information (internal traceability) and XML for exchange of this information (supply chain traceability) between different parties can simplify the record keeping and information exchange, and in turn, the traceability efforts in the grain supply chain.

Application of this framework for developing and implementing internal and supply chain traceability is the next step. The actual implementation for different supply chain actors would provide a better insight into the limitations of this framework and how it can be modified for traceability of different food products.

References

Anderson T., 2004. Introducing XML, <http://www.itwriting.com/xmlintro.php>

Bechini A., Cimino M.G.C.A., Marcelloni F., Tomasi A., 2008. Patterns and technologies for enabling supply chain traceability through collaborative information e-business, *Information and Software Technology*, 50, 342-359.

Bell D., 2004. UML's Sequence Diagram, *IBM*, <http://www.ibm.com/developerworks/rational/library/3101.html>

Can-Trace, 2003. Agriculture and Agri-Food Canada, <http://www.can-trace.org>

Carriquiry M., Babcock B.A., 2007. Reputations, market structure and the choice of quality assurance systems in the food industry, *American Journal of Agricultural Economics*, 89, 12-23.

Department of Defense, 2001. Systems Engineering Fundamentals, *Supplementary text prepared by the Defense Acquisition University Press*, Fort Belvoir, Virgina.

Dorador J.M., Young R.I.M., 2000. Application of IDEF0, IDEF3 and UML methodologies in the creation of information models, *International Journal of Computer Integrated Manufacturing*, 13(5), 430-445.

EDI vs. XML, 2008. Electronic Data Interchange Development, <http://www.edidev.com/XMLvsEDI.html>

Eriksson H., Penker M., 2000. UML Primer, *Business modeling with UML: Business patterns at work*, John Wiley & Sons, Inc., New York, 17-57.

Folinas D., Manikas I., Manos B., 2006. Traceability data management for food chains, *British Food Journal*, 108 (8), 622-633.

Golan E., Krissoff B., Kuchler F., 2004. Food Traceability: One Ingredient in a Safe and Efficient Food Supply, Economic Research Service, *Amber Waves 2*, 14-21.

IDEF0 Function Modeling Method, 1993. Integrated Definition Methods, *IDEF Family of methods, Knowledge Based Systems, Inc.*<http://www.idef.com/idef0.html>

International Organization for Standardization, 2007. New ISO standard to facilitate traceability in food supply chains, *ISO 22005:2007*.

Jansen-Vullers M.H., van Dorp C.A., Buelens A.J.M., 2003. Managing traceability information in manufacture, *International Journal of Information Management*, 23, 395-413.

Laux C.M., 2007. The impacts of a formal quality management system: a case study of implementing ISO 9001 at Farmers Cooperative Co., IA, *Ph.D. Thesis*, Iowa State University.

Lee J., Xue N.L., 1999. Analyzing user requirements by use cases: A goal-driven approach, *IEEE Software*, 16(4), 92-101.

Madec F., Geers R., Vesseur P., Kjeldsen N., Blaha T., 2001. Traceability in the pig production chain. Revue Scientifique Et Technique (International Office of Epizootics) 20 (2), 523–537.

Miller R., 2003. Practical UML: A Hands-on Introduction for Developers, *Embarcadero Technologies*,<http://dn.codegear.com/article/31863>

McKean J.D., 2001. The importance of traceability for public health and consumer protection, *Revue Scientifique Et Technique (International Office of Epizootics)*, 20(2), 363-371.

Nardi M.G., Sperry S.E., Davis T.D., 2007. Grain Supply Chain Management Optimization Using ArcGIS in Argentina, Environmental Systems Research Institute, *ESRI- Professional Papers, 2007.*

Official Journal of the European Communities, 2002. Regulation (EC) No 178/2002 of the European Parliament and the Council of 28 January 2002.

Opara L.U., 2003. Traceability in agriculture and food supply chain: a review of basic concepts, technological implications, and future prospects, *Food, Agriculture & Environment*, 1(1), 101-106.

Regattieri A., Gamberi M., Manzini R., 2007. Traceability of food products: General framework and experimental evidence, *Journal of Food Engineering*, 81, 374-356.

Reuters, 2008. North America tomato industry reeling: growers, *Reuters*, <http://www.reuters.com/article/wtMostRead/idUSN6A33595920080610>

TraceFood, 2007. TraceCore – XML Standard Guidelines, *TraceFood*, <http://193.156.107.66/ff/po/TraceFood/TraceCore%20XML.htm>

TraceFood Wiki, 2009. <http://www.tracefood.org>

U.S. Food and Drug Administration, 2002. The Bioterrorism Act of 2002.

Van Dorp K.J., 2002. Tracking and tracing: a structure for development and contemporary practices, *Logistics Information Management*, 15(1), 24-33.

Figure 1. The Bulk Grain Supply Chain in United States

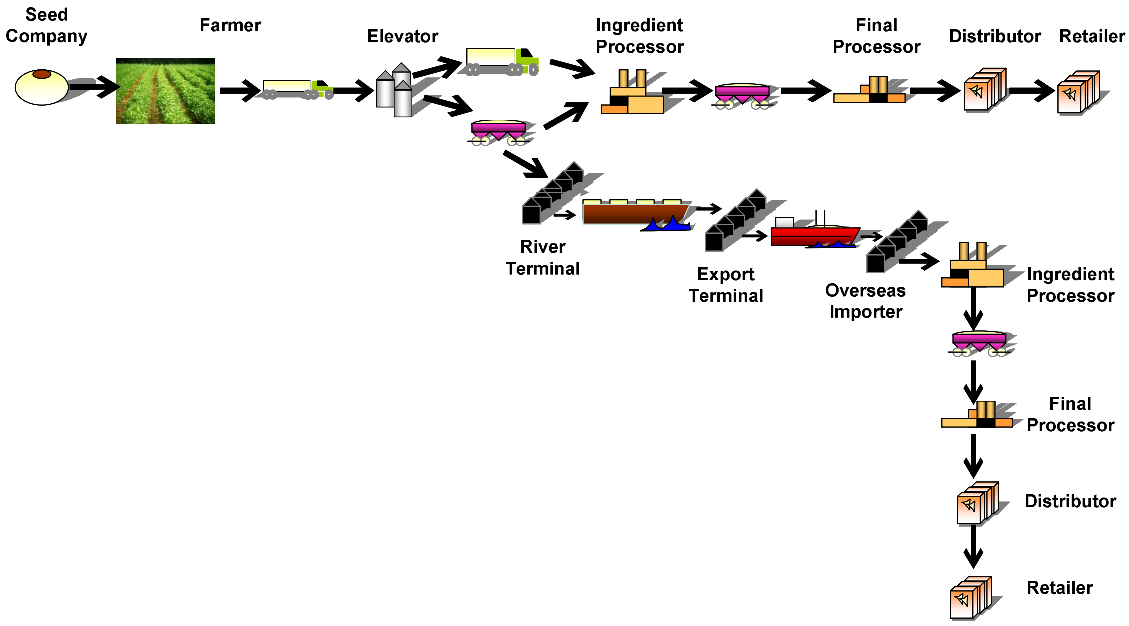
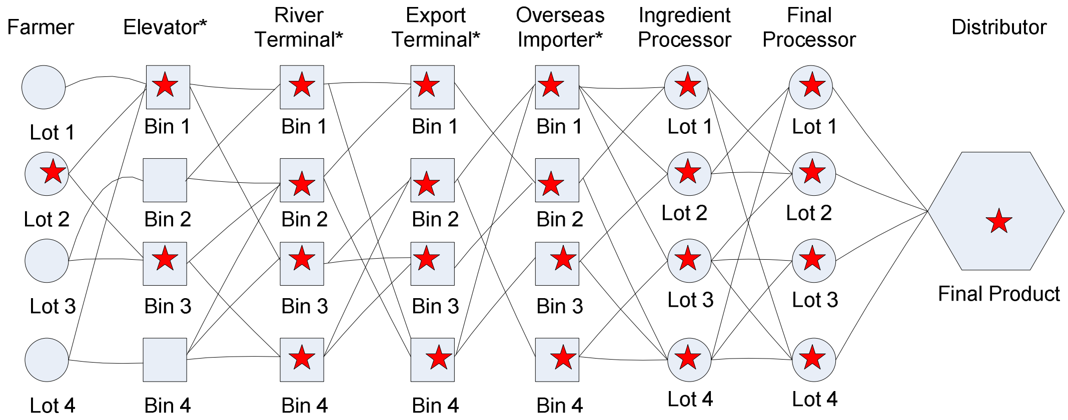


Figure 2. A typical grain lot aggregation and segregation scenario



*Bin- Storage bin, each bin is considered as one lot

★ - Contaminated Lot

Figure 3. Grain Supply Chain Traceability System Use Case diagram

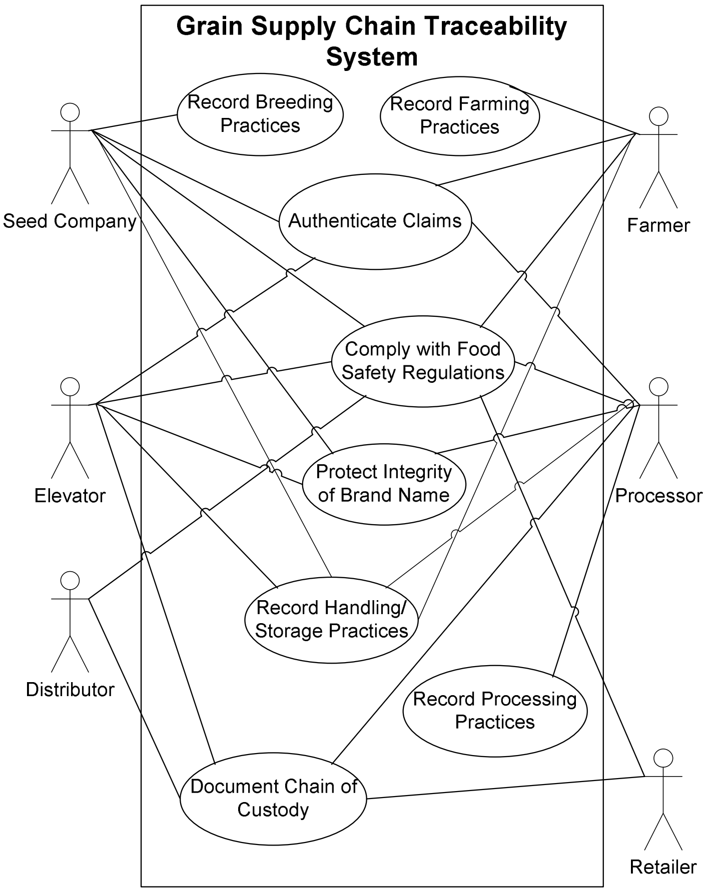


Figure 4. IDEF0 model

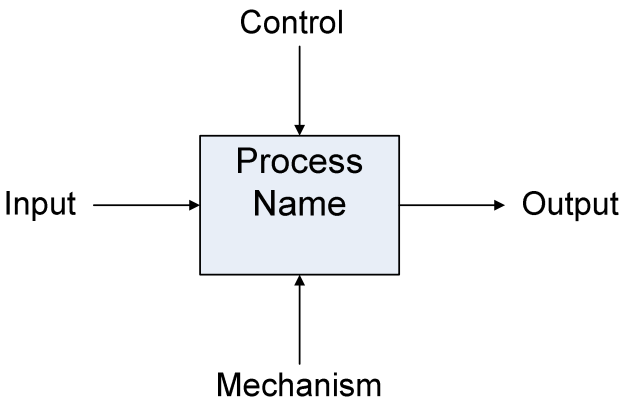


Figure 5. IDEF0 model for developing an internal traceability system at a grain elevator

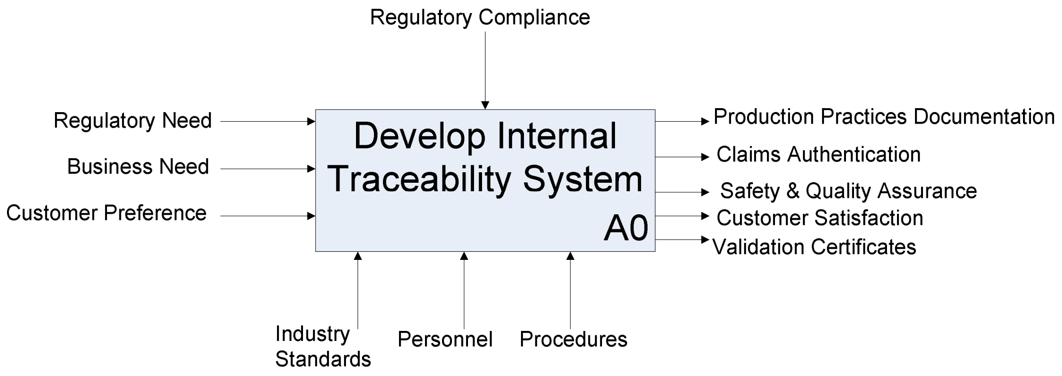


Figure 6. IDEF0 model for developing and implementing a traceability system at an elevator handling specialty grain

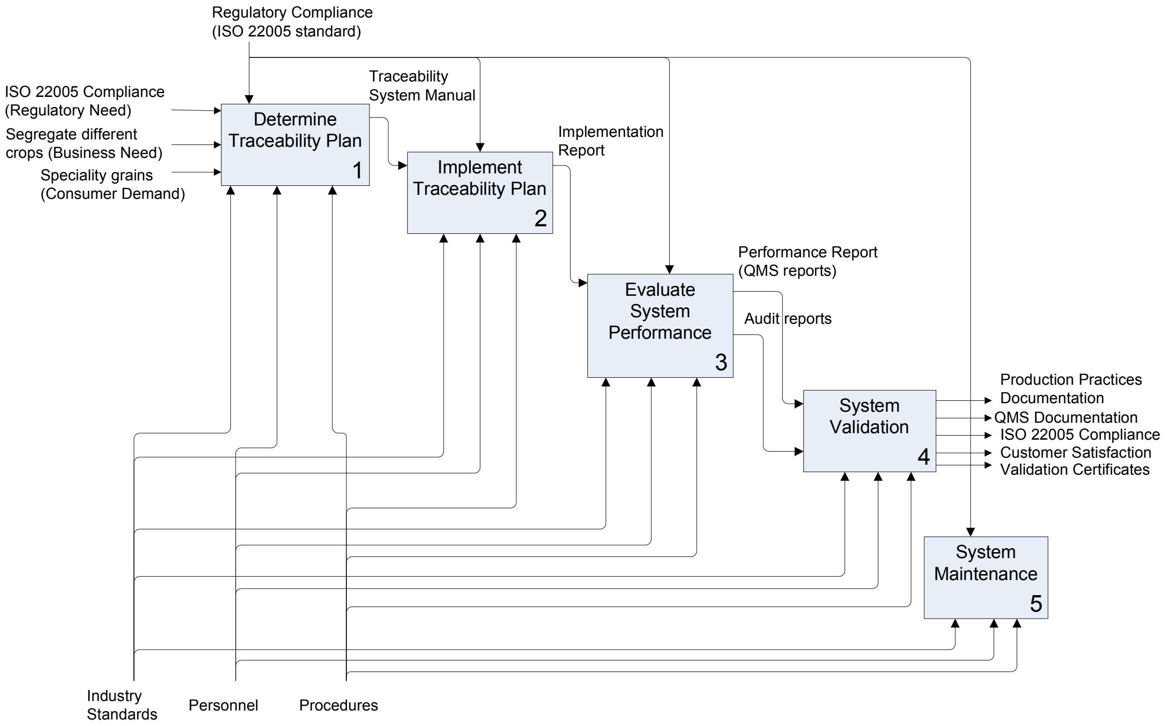
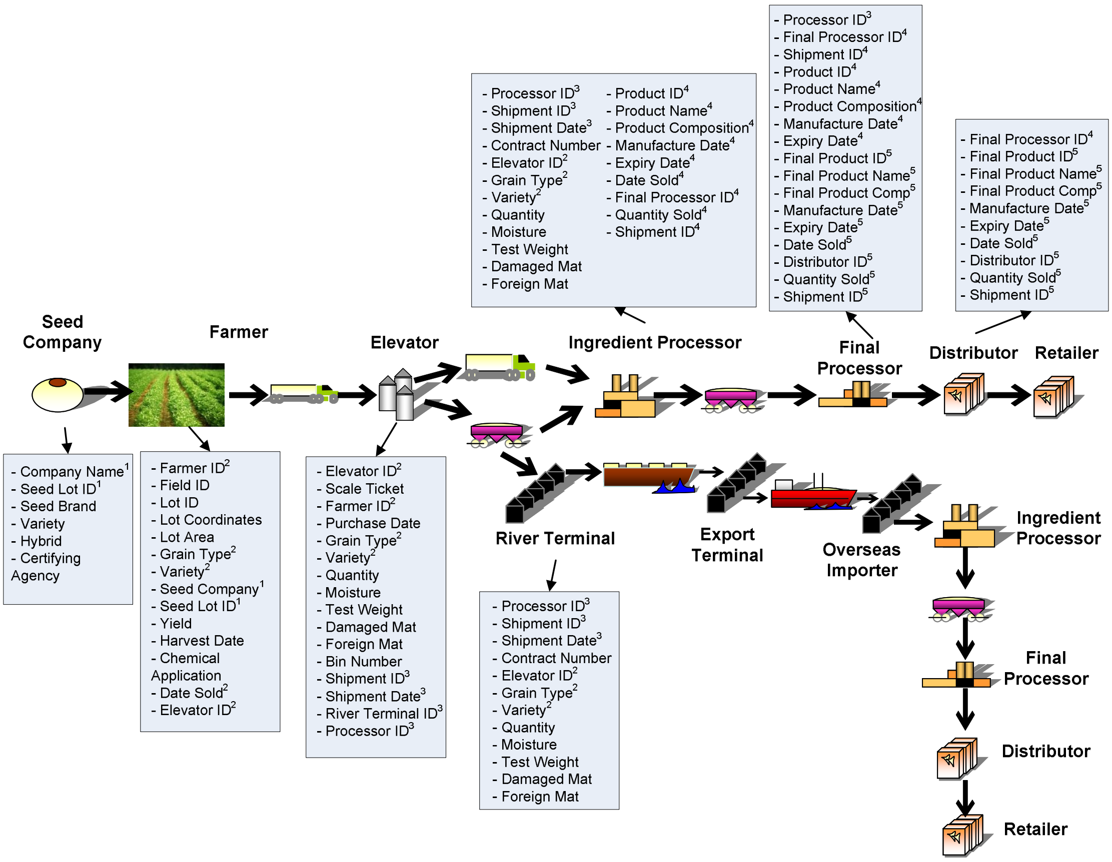


Figure 7. Possible information exchange between different actors in the grain supply chain



[1,2,3,4,5] Information that is passed from one actor to the next in the supply chain

Figure 8. Sequence diagram for information exchange in bulk grain supply chain when additional information about a suspect product is requested

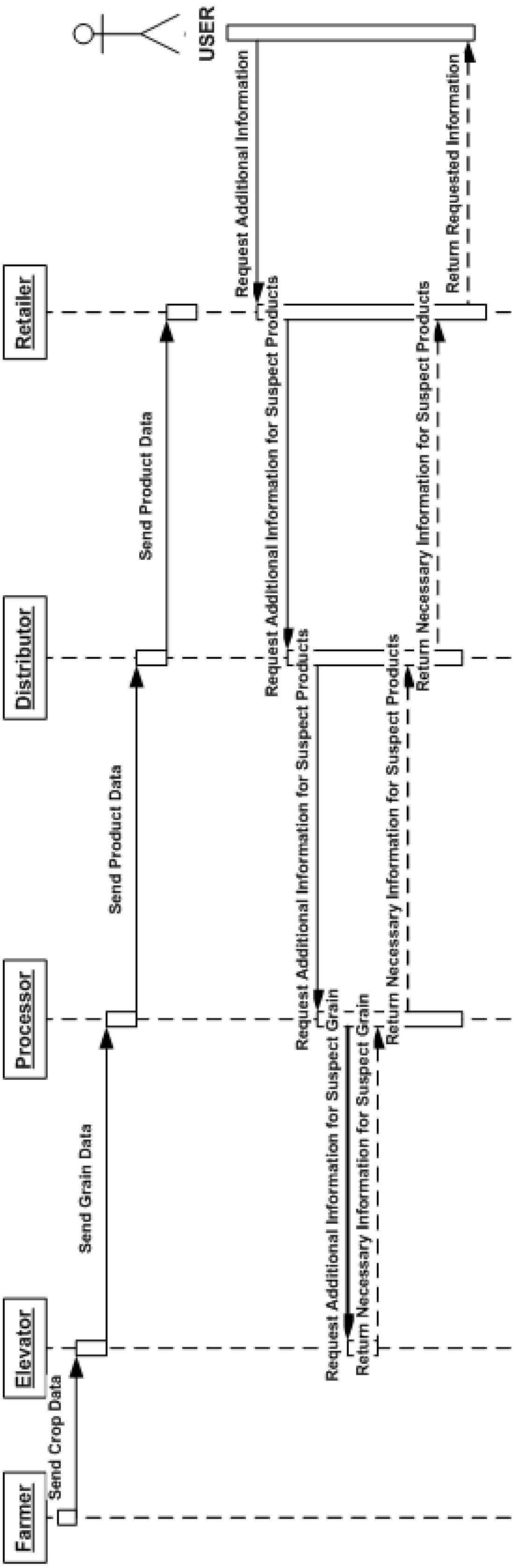


Figure 9. Partial Entity-Relationship diagram of internal traceability database for a grain elevator

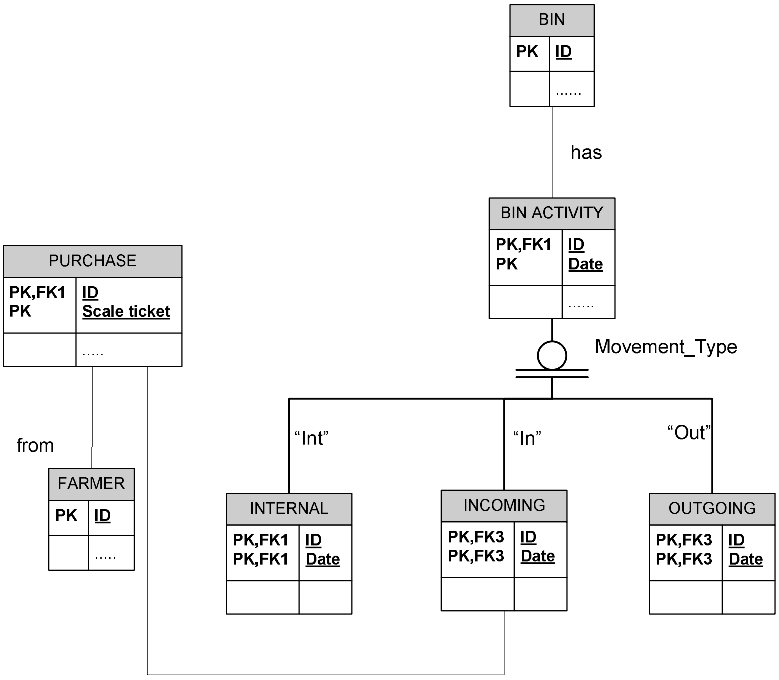
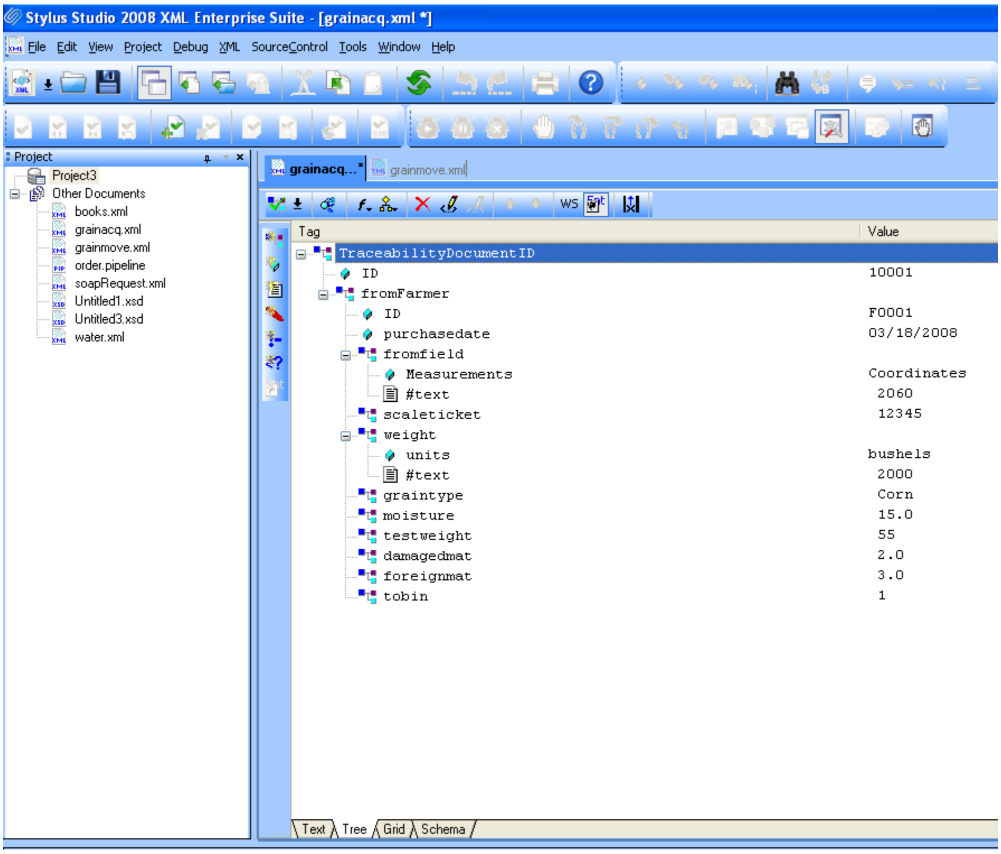


Figure 10. XML document and tree format for Acquisition Notification

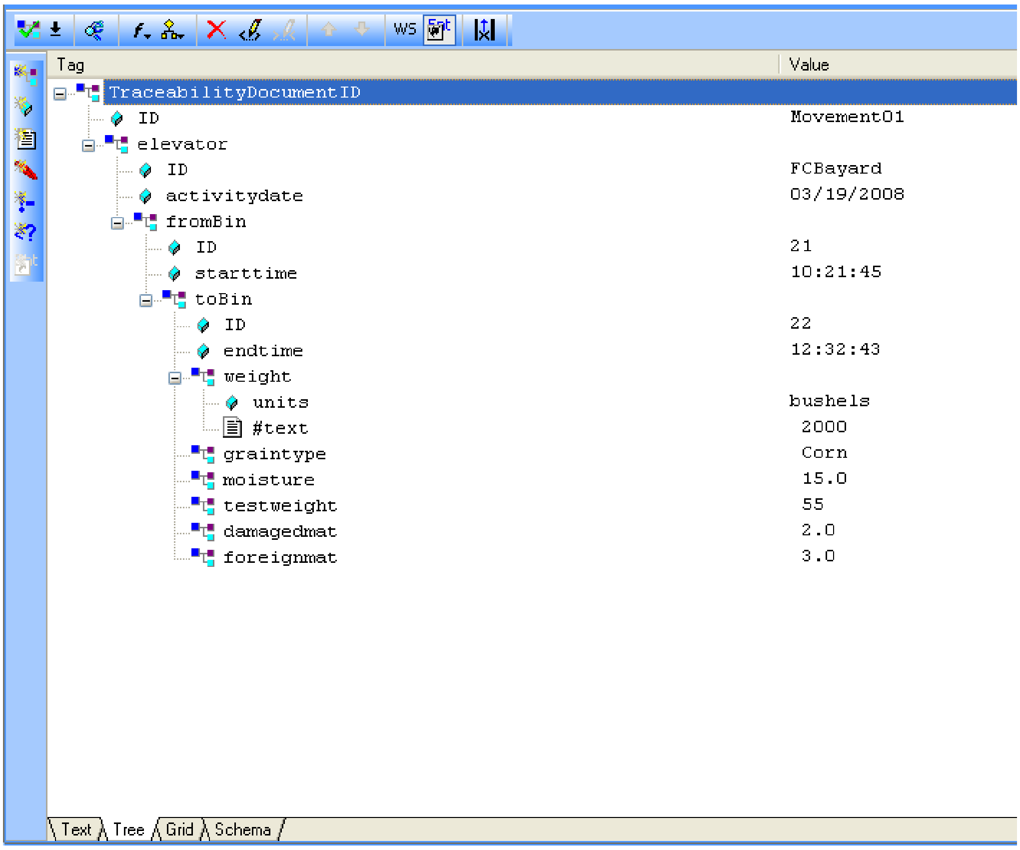
```
<TraceabilityDocumentID ID="10001">
 <fromFarmer ID="F0001" purchasedate="03/18/2008">
  <fromfield Measurements="Coordinates"> 2060 </fromfield>
  <scaleticket> 12345 </scaleticket>
   <weight units="bushels"> 2000 </weight>
   <graintype> Corn </graintype>
   <moisture> 15.0</moisture>
   <testweight> 55 </testweight>
   <damagedmat> 2.0</damagedmat>
   <foreignmat> 3.0 </foreignmat>
   <tobin> 1 </tobin>
   </fromFarmer>
   </TraceabilityDocumentID>
```



Tag	Value
TraceabilityDocumentID	
ID	10001
fromFarmer	
ID	F0001
purchasedate	03/18/2008
fromfield	
Measurements	Coordinates
#text	2060
scaleticket	12345
weight	
units	bushels
#text	2000
graintype	Corn
moisture	15.0
testweight	55
damagedmat	2.0
foreignmat	3.0
tobin	1

Figure 11. XML document and tree format for Movement Notification

```
<TraceabilityDocumentID ID="Movement01">
  <elevator ID = "FCBayard" activitydate = "03/19/2008">
  <fromBin ID="21" starttime="10:21:45">
    <toBin ID="22" endtime="12:32:43">
       <weight units="bushels"> 2000 </weight>
        <graintype> Corn </graintype>
        <moisture> 15.0</moisture>
        <testweight> 55 </testweight>
        <damagedmat> 2.0</damagedmat>
        <foreignmat> 3.0 </foreignmat>
         </toBin>
         </fromBin>
         </elevator>
         </TraceabilityDocumentID>
```



```
Tag                                         Value
⊟ TraceabilityDocumentID
    ID                                      Movement01
  ⊟ elevator
      ID                                    FCBayard
      activitydate                          03/19/2008
    ⊟ fromBin
        ID                                  21
        starttime                           10:21:45
      ⊟ toBin
          ID                                22
          endtime                           12:32:43
        ⊟ weight
            units                           bushels
            #text                            2000
          graintype                          Corn
          moisture                           15.0
          testweight                         55
          damagedmat                         2.0
          foreignmat                         3.0
```

CHAPTER 4. Data modeling to facilitate internal traceability at a grain elevator

Manuscript to be submitted to the *Journal of Food Engineering*

Maitri Thakur[1, 2,] *, Bobby J. Martens[3] and Charles R. Hurburgh[1, 4]

[1] Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011
[2] Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA 50011
[3] Department of Logistics, Operations, and Management Information Science, Iowa State University, Ames, IA 50011
[4] Department of Food Science and Human Nutrition, Iowa State University, Ames, IA 50011
* Primary author, Corresponding author.

Abstract

Data management in food supply chains to facilitate product traceability has gained importance in the past years. This paper presents a relational database model to facilitate internal traceability at a grain elevator, which is one of the first nodes in a food supply chain. At an elevator, grain lots (inbound deliveries) are blended to meet buyer specifications, and individual lot identity is not maintained. As a result, an outbound shipment to a customer likely contains grain from many different sources. In a food safety related emergency, tracing the source of a problem or tracking other affected shipments would be nearly impossible. An efficient internal data management system could mitigate these problems by recording all grain lot transformations/activities, including movement, aggregation, segregation, and destruction as well as supplier and customer information. In this paper, a relational database management system is proposed that stores all necessary information, including product and quality information, related to the grain lots in order to enable product traceability. The system can be queried to retrieve information related to incoming, internal and outgoing lots and to retrieve information that connects the individual incoming grain lots to an outgoing shipment. Furthermore, this system can be used both to trace back to the source of a given lot and to track information about previously shipped lots forward.

Keywords: Internal traceability, Bulk grain handling, Elevator, Data modeling, ER model

1 Introduction

Tracking and tracing food products throughout the supply chains has gained considerable importance over the last few years (Carriquiry and Babcock, 2007; Jansen-Vullers et al., 2003; Madec et al., 2001; McKean, 2001; Thakur and Hurburgh, 2009). Consumers all over the world have experienced various food safety and health issues. In addition, consumer demand for high quality food and feed products, non-GMO (genetically modified organisms) foods and other specialty products such as organic food has grown in the past years. These factors have led to a growing interest in developing systems for food supply chain traceability, and, as a result, a number of food safety and traceability laws exist in different countries.

The European Union law describes “Traceability” as an ability to track any food, feed, food-producing animal or substance that will be used for consumption, through all stages of production, processing and distribution (Official Journal of the European Unions, 2002). Considering this definition, traceability is important for many reasons, such as responding to food security threats, documenting chain of custody, documenting production practices, meeting regulatory compliance, and even analyzing logistics and production costs. Besides ensuring a safe food supply, the USDA Economic Research Service states that use of a traceability system results in lower cost distribution systems, reduced recall expenses, and expanded sales of products with attributes that are difficult to discern (Golan et al., 2004). Thus, in several countries food traceability has become important for reasons other than just the legal obligations.

Three examples demonstrate how traceability standards are being developed and implemented. The ISO 22005 Food Safety Standard requires that each company know their immediate suppliers and customers based on the principle of one up and one down (International Organization for Standardization, 2007). It also states that one weak link in the supply chain can result in unsafe food, which can present a serious danger to consumers and have costly repercussions for the suppliers. Food safety is therefore the joint responsibility of all the actors involved. Next, the Bioterrorism Preparedness and response Act of 2002 (the Bioterrorism Act) requires all food and feed companies to self-register with the Food and Drug Administration and maintain records and information for food traceability purposes (US Food and Drug Administration, 2002). Finally, the GS1 Traceability Standard states that traceability across the supply chain involves the association of flow of information with the physical flow of traceable items. It also states that in order to achieve traceability across the supply chain, all traceability partners must achieve internal and external traceability (GS1 Global Traceability Standard, 2007). Therefore, all the actors involved in the food supply chain are required to store necessary information related to the food product that link inputs with outputs, so that when demanded, the information can be provided to the food inspection authorities on a timely basis.

Previous research has emphasized the importance of internal traceability systems. Moe (1998) states that many advantages can accrue from having an internal traceability system from being able to trace the raw material that went into a final product to possibility of improved process control, correlating product data with raw material characteristics and processing data as well as optimization of the use of raw materials for each product type. In order to achieve a fully traceable supply chain, it is important to develop systems for both external supply chain traceability as well as internal traceability. This includes linking, to the best extent possible, units of output with specific units of input. First, each actor must have the ability to externally trace back and track forward product information using the one-up and one-down basis. Then, in order to determine the cause of the problem or to efficiently recall the associated (or contaminated) food products, each supply chain actor should have an internal record-keeping system enabling them to trace back to the input ingredients and track forward to the output products. Therefore, each actor in the supply chain must not only know their immediate suppliers and customers but also maintain accurate records of their internal processes.

Still, traceability in the food industry is lacking. This is especially a concern when evaluating supply chains related to bulk grain. In this paper, we present a traceability system for a bulk grain handling scenario. Because of the complexities associated with receiving, storing, and blending bulk grains, a bulk grain handling scenario serves as a good example of how a traceability system can be developed for complex product flows. In this paper, we first describe the functions of a grain elevator, including the complications related to implementing a bulk grain traceability system. Next, traceability literature is highlighted and data management systems are reviewed. Finally, our methodology is discussed and the results of our relational database model, which can be used to facilitate internal traceability at a grain elevator, are offered.

1.1 **Bulk grain handling**

Various lot-activities (transformations) take place as grain moves through the supply chain from the farm to the consumer. These transformations include aggregation, segregation, storage, transfer and destruction (Thakur and Hurburgh, 2009). It is important to be aware of the type and location of each transformation as it is necessary to be able to track and trace the food product through a firm or processing facility (Donnelly et al., 2009; Schwägele, 2005). Grain elevators, which handle bulk commodities like corn and soybeans, are important nodes in the bulk grain supply chan. The elevators buy grain from farmers and store the grain in storage bins (i.e., grain bins or silos) before selling it to the customers. Figure 1 shows a typical bulk grain handling scenario.

The incoming grain lots from farmers are assigned a unique scale-ticket number, weighed and graded based on quality parameters. These quality parameters include moisture, test weight, damaged material and foreign material. A quality grade is determined based on these parameters and the lot is assigned and transferred to one or more storage bins based on space and quality constraints. Grain is kept in storage bins until it is shipped to a customer. However, while in storage, all or part of the contents of a bin can be transferred to other bins in order to avoid spoilage due to environmental conditions (usually related on increasing temperature inside a bin). This internal movement often goes unrecorded and complicates the lot dynamics due to mixing of previously defined grain lots. In the absence of these internal records, it is impossible to link the incoming and the outgoing lots. Again, just before shipment, grain from different storage bins (i.e., different quality) is blended to meet the customer specifications for quality and to maximize the elevator’s profit.

As a result of this grain elevator blending process, one storage bin likely contains grain from many different sources (i.e., original farmer lots), and a specific grain lot shipped to a customer (i.e., food processor or manufacturing plant) may contain grain from multiple sources. Any number of original farmer lots might ultimately comprise a finished food product. If a food related emergency occurred, isolating the source of the problem would be nearly impossible, so a recall of all the finished goods that might possibly have been contaminated would be the only method to ensure the consumer’s safety. Such a recall would be time intensive and complex, result in high cost, be damaging to brand names, and add risk to consumers’ safety. The following section reviews relevant literature related to traceability and database management systems.

1.2 **Traceability and data management systems**

A data model is defined as a coherent representation of objects from a part of reality (Elmasri and Navathe, 2000). A wide range of systems are available for traceability in the food industry, ranging from paper-based systems to IT enabled systems (Food Standards Agency, 2002). Radio Frequency Identification (RFID) technology is also used to develop traceability systems in food supply chains (Natsui and Kyowa, 2004). RFID tags can be used for identification of individual product lots as they move through the supply chain. Information management and database management techniques are also used for developing traceability systems. Niederhauser et al. (2008) presents a conceptual information system for tracking specialty coffee while Jansen-Vullers et al. (2003) present a reference model designed to accommodate support for the registration of operations on lots or batches and support for the registration of associated operation variables and values. This model displays the functionality for traceability in manufacturing when production lots or batches are defined. Relational databases are widely used by corporations for operational management programs. The use of these databases for traceability in agricultural industry other than food manufacturing is, however, unheard of by the authors. Support for strategic decisions through analytical databases in the sense of data warehouses, as used and implemented intensively in the industrial sector has thus far not been given serious consideration in the agricultural sector (Schulze et al., 2007). It has been shown that the efficiency of a traceability system depends on its ability to record and retrieve the requested lot-related information (Folinas et al., 2006).

Senneset et al. (2007) state that one of the basic prerequisites of both internal and external supply chain traceability is the unique identification of all raw materials, semi-finished products and finished products. The authors offer three types of operations necessary for obtaining internal traceability:

(1) Recording the unique identities of traceable units. These usually refer to inputs to a process.

(2) Assigning unique identities to new traceable units. These usually refer to outputs from a process.

(3) Linking a set of input unit identities to one or more sets of output identities. These usually refer to transformation of raw materials to finished products.

Based on the concept of unique identification, a Traceable Unit (TU) is defined as any item with predefined information which may need to be retrieved and which may be priced, or ordered, or invoiced at any point in any supply chain. In practice, a TU refers to the smallest unit that is exchanged between two parties in the supply chain (TraceFood Wiki, 2009). In order to achieve chain traceability and meet the three traceability conditions offered above, efficient internal traceability systems must be in place at each food enterprise (node) in a supply chain. Therefore, it is important to develop systems which record both information related to traceable units and associated transformations occurring internally within each node. Such traceability systems can become complex, especially when TU are not well defined.

Since bulk grain is traded according to grade standards based on quality parameters of the grain lots, it is important to integrate the relevant quality data with the traceable units. Moe (1998) states that traceability can be used in four distinct contexts: product (origin, processing history, distribution and location after delivery), data generated throughout the quality loop, calibration (standards, physical properties, etc.), and IT and programming related to system design and implementation. Jansen-Vullers et al. (2003) suggest the following four elements for traceability:

(1) Physical lot integrity: this includes the lot size and how well the lot integrity is maintained.
(2) Data collection: this includes two types of data; lot tracing data and process data.
(3) Product identification and process linking: to determine product composition.
(4) Reporting: to retrieve data from the system.

Based on these principles, identification of data capture points and the data elements to be recorded at these points is the first step in developing a database management system for traceability.

For efficient grain supply chain traceability, the elevator has a responsibility to maintain data that links inputs (inbound deliveries) and outputs (outbound shipments). When needed, management should be able to retrieve the necessary information from this recorded data. In this paper, we propose the use of a relational database management system (RDBMS) for internal traceability at a grain elevator. The purpose of this database model is to record all the transformations related to incoming and outgoing grain lots as well as the transformations that take place internally at an elevator. Therefore, the objective of this database model is to track and trace individual grain lots through the bulk grain supply chain. The database can be queried to retrieve the relevant information when necessary. However, there are certain factors that create problems in modeling of the bulk grain handling data. The “fluid-like” characteristics of bulk grain distinguish it from other food products and make it very difficult to define a fixed lot-size (or traceable unit) for traceability purposes. The following section describes how these factors were modelled.

2 Methodology

2.1 **Traceable Units**

Defining a lot or a traceable unit (TU) by breaking product flows into discrete units is a way to achieve product differentiation for tracking (Golan, et al., 2004; Moe, 1998). However, the definition of a grain lot changes throughout the bulk handling process. In this database model, we use various definitions of a lot of bulk grain at different stages of handling within the elevator and each lot is uniquely identified. The following definitions of a grain lot are used:

1. At the time of purchase, a truckload of grain purchased from a farmer that is identified by a unique scale ticket number is considered a lot. This lot can be assigned to one or more storage bins depending on quality of grain and bin capacities available at that time.
2. In storage, the quantity of grain contained in one bin is considered as one lot. This lot can have multiple sub-lots (different incoming lots identified by unique scale ticket numbers). In storage, each lot is uniquely identified by the storage bin number.
3. For shipment to a customer, one truckload or the shipment load in one railcar is considered as one lot. This outgoing lot might come from several lots (in storage, each bin is a lot) blended together to meet the customer specifications. Each outgoing shipment has a corresponding customer contract and is uniquely identified by a shipment ID.

2.2 **Lot Transformations**

Figure 1 provides an overview of the lot dynamics at a grain elevator. Three types of activities related to incoming, internal and outgoing grain lots take place at an elevator. Each activity type can be defined by a set of transformations summarized in Table 1. Each lot transformation has a storage bin number associated with it because: 1) incoming grain is assigned to one or more bins, 2) grain can be moved internally from one bin to another and finally, 3) outgoing shipments are prepared by blending grain from different bins in order to meet customer specifications. So, this data model maintains information about lot transformations related to each bin in addition to activity date and time, farmer and customer information, and various grain quality parameters.

2.3 **Entity- Relationship Model (ER model)**

The entity-relationship (E-R) modeling technique was used to develop the internal traceability grain handling database model. An E-R model is a detailed, logical representation of data for an organization or for a business area. The E-R model is represented in terms of entities in the business environment, the relationships among those entities, and the attributes of both the entities and their relationships (Hoffer et al., 2006). The benefits to using a relational database management system (RDBMS) come from its ability to store data in a ‘‘normalized’’ format. This concept was originally presented by Codd (1970), who mathematically developed the relational model to provide a better structure for databases. Data normalization is simply a way of organizing data so that it allows for increased efficiency of data storage and retrieval. While spreadsheets can store data in a normalized format, it is very difficult to retrieve in a simple and timely manner. We developed a database designed to facilitate the storage, retrieval and analysis of grain handling data at an elevator. The internal traceability grain handling model was developed using Oracle Database 10g software. The rationale and principles used to develop this database are directly applicable to other commercially available RDBMS software. The design of the relational database adheres to the principles of normalization focusing on data handling efficiency and flexibility.

Figure 2 shows the symbols used in an ER model, which will be used in the later modeling steps. An entity stands for things that can be uniquely identified and characterized by their attributes; whereas relationships represent associations among different entities. Attributes represent information about an entity and relationship types by mapping them into value sets (Patig, 2006). A primary key is an attribute or combination of attributes that uniquely identify an instance in a database while a foreign key is used to link two tables (entities). Typically, a primary key from one table (entity) is inserted into another table (entity), and it then becomes a foreign key. Relationships between two entities work by matching the key columns in two tables. This is usually done by matching a primary key (that provides a unique row/instance) from one table to a foreign key instance in another table. Table 2 describes the different kind of relationships.

Such relationships were developed for the grain lot activities/transformations and associated quality characteristics.

Figure 2 also represents supertype and subtype entities. A supertype entity is used to represent two or more entities when they are viewed as the same entity by other entities. A subtype entity is an entity that is a special case of another entity, created when attributes or relationships apply to only some instances of an entity. The subsets of instances to which the attributes or relationships apply are separated into entity subtypes. When an attribute applies only to some occurrences of an entity, the subset of occurrences to which it applies should be separated into entity subtypes.

The common data elements are put in the supertype entity and the specific data elements are placed with the subtype to which they apply. All attributes of the supertype must apply to all subtypes. Each subtype contains the same key as the supertype. Database triggers can be used to automatically transfer data from supertype tables to subtype tables. A database trigger is a procedural code that is automatically executed in response to certain events on a particular table in a database (Hoffer et al., 2006). The Structured Query Language (SQL) was used to develop a functional model that can be implemented in a real elevator setting. Some sample reports and queries are discussed in the following sections.

3 Results

Figure 3 shows the E-R model for the internal traceability database at a grain elevator. Table 3 provides a description of each entity and the related attributes. Every time a transformation (aggregation, segregation, storage, transfer, etc.) takes place, the quality factors of moisture, test weight, foreign material and damaged material are recorded. A scale ticket number is assigned to the grain lots purchased from the farmers. Each incoming lot is tested for quality and transferred to one or more storage bins (that may already contain previous lots) depending on grain type (corn or soybeans), space availability and grain quality. The information related to the farmer and the activity dates are also recorded. Similar information is recorded when grain is moved internally at the elevator and for shipments to the customers (see Figure 3 for details). The *bin_activity* entity has three sub-types, one each for the *internal*, *incoming*, and *outgoing* grain movement corresponding to every storage bin. Similarly, the *shipment_info* entity has two sub-types, *truck* and *rail*. The data is recorded in each table depending on the mode of transportation of the outgoing shipment. Database triggers were created for automatic data transfer to the sub-type tables.

By utilizing the relational database design, the proposed model can store, manage, retrieve all grain handling data and run calculations for aggregated quality of the blended products. The integration of all these functions makes this model unique from the existing spreadsheet based inventory control programs for grain elevators. This model combines inventory information, grain handling and grain quality information as well as the grain blending process in one centralized location.

3.1 **Database Triggers**

A trigger is a named set of SQL statements that are considered (triggered) when a data modification (such as INSERT, UPDATE, and DELETE) occurs. If a condition stated within the trigger is met, then a prescribed action is taken (Hoffer et al., 2006). Triggers are commonly defined as ***On*** event ***If*** condition ***Then*** action (Dayal et al, 1988; Hanson, 1989; Kotz et al, 1988; Widom and Finkelstein, 1990). Triggers were used for two entities, namely, b*in_activity* and *shipment_info* to automatically transfer data from the supertype entity to the respective subtype entities based on the response (i.e. the type of activity). SQL code for these database triggers is shown in Figure 4. It can be noted that data is added to the respective subtype entities using the triggers based on the type of movement and the type of shipment mode, respectively, for the two supertype entities.

3.2 **Queries and Reports**

Once the data is stored in the database, the manipulation is accomplished through the use of queries written using the Structured Query Language (SQL). SQL allows in recreating the original spreadsheet file formats as well as subsets and data comparisons. The set of queries presented in this section act as a start for basic data retrieval, but the WHERE clauses should all be changed to match specific data requirements. Once written these queries can be saved and easily executed at a later date but would return varying results based on the changes made to the data set during that time. Some sample reports are shown in this section of the paper. The main purpose of this database is to be able to connect the incoming grain lots with the outgoing grain lots. This information is vital in case of a food safety related emergency. Reports can be generated from the database to answer queries such as:

- Which farmers supplied the grain contained in a specific storage bin?

- Which bins were used to blend grain for a specific outgoing shipment?
- Which incoming lots contributed to a specific outgoing shipment?

Figure 5 shows the SQL code and sample report generated to display the farmer information, purchase date, grain type and quantity purchased that was transferred to storage bin number 9.

Figure 6 shows the SQL code and sample report generated to display the outgoing shipments using truck as transportation mode. The report includes the activity date (shipment date), contract number, customer ID, the bin number/s from where the grain is drawn for blending, truck ID and the quantity shipped on each truck in bushels. Similarly, Figure 7 shows the code and report generated to display the outgoing shipments using rail as transportation mode.

The ability to connect the outgoing lot (shipment) information to the incoming lots is important to trace back the source of problem in case of a food safety emergency. Figure 8 shows the SQL code and sample report generated to display the incoming grain lot information corresponding to outgoing shipments to Company A. The query is created so that the report includes the scale ticket number of the incoming lots, purchase date, farmer name, quantity purchased in bushels, bin number assigned to the incoming lot, activity date (shipment date), contract number, bin number/s from where the grain is drawn, and the quantity shipped on each railcar in bushels. This report displays the incoming lots that are present in an outgoing shipment. The grain lots are divisible so a part or an entire incoming grain lot may be present in an outgoing lot. This information can be used to trace back the origin of grain (back to a farmer or a group of farmers) present in an outgoing shipment.

4 Conclusions

Development of data management systems to facilitate product traceability in food supply chains has gained importance in the past years. The ability to track and trace individual product units depends on an efficient supply chain traceability system which in turn depends on both internal data management systems and information exchange between supply chain actors. In this paper, we present a relational database model to facilitate internal traceability at a grain elevator.

Grain elevators handle bulk commodities marketed against generic grade standards that are based on physical attributes. Different lot-activities take place as the grain moves through the supply chain from the farm to the consumer. At an elevator, grain lots (inbound deliveries) are commingled to meet buyer specifications, and lot identity is not maintained. As a result, an outbound shipment to a customer can contain grain from many sources. In a food safety related emergency, it would be almost impossible to trace back the source of problem and to track (forward) other affected lots. This process is very time intensive, increases the recall costs, and can lead to a tainted brand name for the company. The problem can be mitigated by an efficient internal record keeping system that would document all grain activities (transformations). The proposed database system stores product identity and transformation information related to grain lots (traceable units) and can be queried to retrieve information related to all incoming, internal and outgoing lots.

Definition of a lot size or a traceable unit was an important step in developing a data management system since all the information has to be linked to a unique entity, which in general is a specific lot size. But, grain is handled in bulk and defining a lot size is a complex task. So, instead of a strict definition of a lot, we use several definitions and explain how the lot size changes as grain moves through an elevator. Each receipt from a farmer (usually, a truckload) is assigned a unique scale ticket number and considered as one lot. When in storage, a grain bin is considered as one lot which in turn can contain grain from different farmer deliveries (scale tickets). This implies that a storage bin can contain many sub-lots. Again, when the grain is shipped to a customer, an outgoing shipment is prepared by blending grain from different storage bins in order to meet customer specifications. For an outgoing shipment, a railcar or a truckload (depending on the transportation mode) is considered as one lot.

The entity-relationship modeling technique was used to develop the database management system for internal traceability. All the information related to the grain lot activities/transformations and associated quality characteristics were recorded in this database. An important feature of the ER model is the use of supertype and subtype entities. Two entities, the type of grain lot movement and the mode of transportation were modeled as supertype entities. This feature simplified the database design and information retrieval. Depending on the type of movement; whether it is an incoming grain activity, internal activity or an outgoing activity, the information is stored in the corresponding tables. This design was used because these entities (different movement types) share some common attributes. The common attributes such as the quality parameters are placed in the supertype entity *bin_activity* while the specific attributes *scale_ticket, shipment_ID* etc. are placed in the subtype entity to which they apply. Another feature of this model is the use of database triggers. Triggers were used to automatically transfer data from the supertype entity to the subtype entities.

The database can be queried to retrieve information related to any grain lot activity (transformation). It can be used to trace back the source of a given lot or track forward the information related to the shipped lots. The information that connects the individual incoming grain lots to an outgoing lot can also be retrieved using this system as is shown by some sample queries in the results section. This paper demonstrates that using a relational database management approach for recording all lot activities (transformations) is an effective way to link the incoming and outgoing grain lots at an elevator.

The next steps in this work include the development of a graphical user interface to enable the users to enter data in the database. The model also needs to be implemented in a real elevator setting and tested for performance based on the response time of information retrieval in case of a product recall. In future, this system can be used to meet both operational and analytical requirements of the business. The operational requirements of an enterprise's business processes generally include short-term decision making while analytical requirements refer to long-term decision making based on historical and aggregated data. The historical data recorded over long term using a relational database system could be analyzed to study the grain handling practices of the elevator. Elevators move grain from one bin to another and between different elevator locations based on space and quality constraints. Availability of historical data would allow the elevator management to analyze their grain handling practices and to define new procedures in order to optimize the logistics costs and to minimize the food safety risk by optimizing their blending practices.

Acknowledgement

The authors would like to thank Mr. Butch Hemphill of Farmers Cooperative Company, Iowa for his help in understanding the elevator operations.

References

Carriquiry, M., Babcock, B.A., 2007. Reputations, market structure and the choice of quality assurance systems in the food industry. American Journal of Agricultural Economics 89, 12–23.

Codd, E.F., 1970. A relational model of data for large shared data banks. Communications of the Association of Computing Machinery 13 (6), 377–387.

Dayal, U., Blaustein, B., Buchmann, A., Chakravarthy, U., Hsu, M., Ledin, R., McCarthy, D., Rosenthal, A., Sarin, S., Carey, M.J., Livny, M., Jauhari, R., 1988. The HiPAC project: combining active databases and timing constraints. ACM SIGMOD Record, 17(1), 51-70.

Donnelly, K.A., Karlsen, K.M., Olsen, P., 2009. The importance of transformations of traceability – A case study of lamb and lamb products. Meat Science 83, 68-73.

Elmasri, R., Navathe, S. B., 2000. Fundamentals of database systems 3rd edition, Reading, MA, USA: Addison-Wesley.

Folinas, D., Manikas, I., Manos, B., 2006. Traceability data management for food chains. British Food Journal 108 (8), 622–633.

Food Standards Agency, 2002. Traceability in the Food Chain; a Preliminary Study. Food Chain Strategy Division, Food Standards Agency.

Golan, E., Krissoff, B., Kuchler, F., 2004. Food traceability: one ingredient in a safe and efficient food supply, economic research service. Amber Waves 2, 14–21.

GS1 Global Traceability Standard, 2007. Business Process and System Requirements for Full Chain Traceability. <http://www.gs1.org/traceability/gts>

International Organization for Standardization, 2007. New ISO Standard to Facilitate Traceability in Food Supply Chains. ISO 22005:2007.

Hanson, E.N., 1989. An initial report on the design of Ariel: a DBMS with an integrated production rule system. SIGMOD Record, 18(3):12-19.

Hoffer, J.A., Prescott, M., McFadden, F., 2006. Modern Database Management. Prentice Hall, New Jersey.

Jansen-Vullers, M.H., van Dorp, C.A., Buelens, A.J.M., 2003. Managing traceability information in manufacture. International Journal of Information Management 23, 395–413.

Kotz, A.M., Dittrich, K.R., Mülle, J.A., 1988. Supporting Semantic Rules by a Generalized Event/Trigger Mechanism, Proceedings of the International Conference on Extending Database Technology: Advances in Database Technology, March 14-18, 76-91.

Madec, F., Geers, R., Vesseur, P., Kjeldsen, N., Blaha T., 2001. Traceability in the pig production chain. Revue Scientifique Et Technique (International Office of Epizootics) 20 (2), 523–537.

McKean, J.D., 2001. The importance of traceability for public health and consumer protection. Revue Scientifique Et Technique (International Office of Epizootics) 20 (2), 363–371.

Moe, T., 1998. Perspectives on traceability in food manufacture. Trends in Food Science and Technology 9, 211-214.

Natsui, T., Kyowa, A., 2004. Traceability System using RFID and Legal Issues. WHOLES, A multiple view of individual privacy in a networked world, <www.sics.se/privacy/wholes2004/papers/takato.pdf>

Niederhauser, N**.,** Oberthür**,** T., Kattnig, S., Cock, J., 2008. Information and its management for differentiation of agricultural products: The example of specialty coffee. Computers and Electronics in Agriculture 61(2), 241-253.

Official Journal of the European Communities, 2002. Regulation (EC) No. 178/2002 of the European Parliament and the Council of 28 January 2002.
Patig, S., 2006. Evolution of entity-relationship modeling. Data and Knowledge Engineering 56(2), 122-138.
Schulze, C., Spilke, J., Lehner, W., 2007. Data modeling for precision dairy farming within the competitive field of operational and analytical tasks. Computers and Electronics in Agriculture 59(1-2), 39-55.
Senneset, G., Forås, E., Fremme, K.M., 2007. Challenges regarding implementation of electronic chain traceability. British Food Journal 109(10), 805-818.
Schwägele, F., Traceability from a European perspective. Meat science 71(1), 164-173.
Thakur, M., Hurburgh, C.R., 2009. Framework for implementing traceability in the bulk grain supply chain. Journal of Food Engineering 95(4), 617-626.
TraceFood Wiki, 2009. <http://www.tracefood.org>
US Food and Drug Administration, 2002. The Bioterrorism Act of 2002.
Widom, J., Finkelstein, S.J., 1990. Set-oriented production rules in relational database systems, Proceedings of the 1990 ACM SIGMOD international conference on Management of data, May 23-26, Atlantic City, New Jersey, 259-270.

Figure 1. A typical bulk grain handling scenario

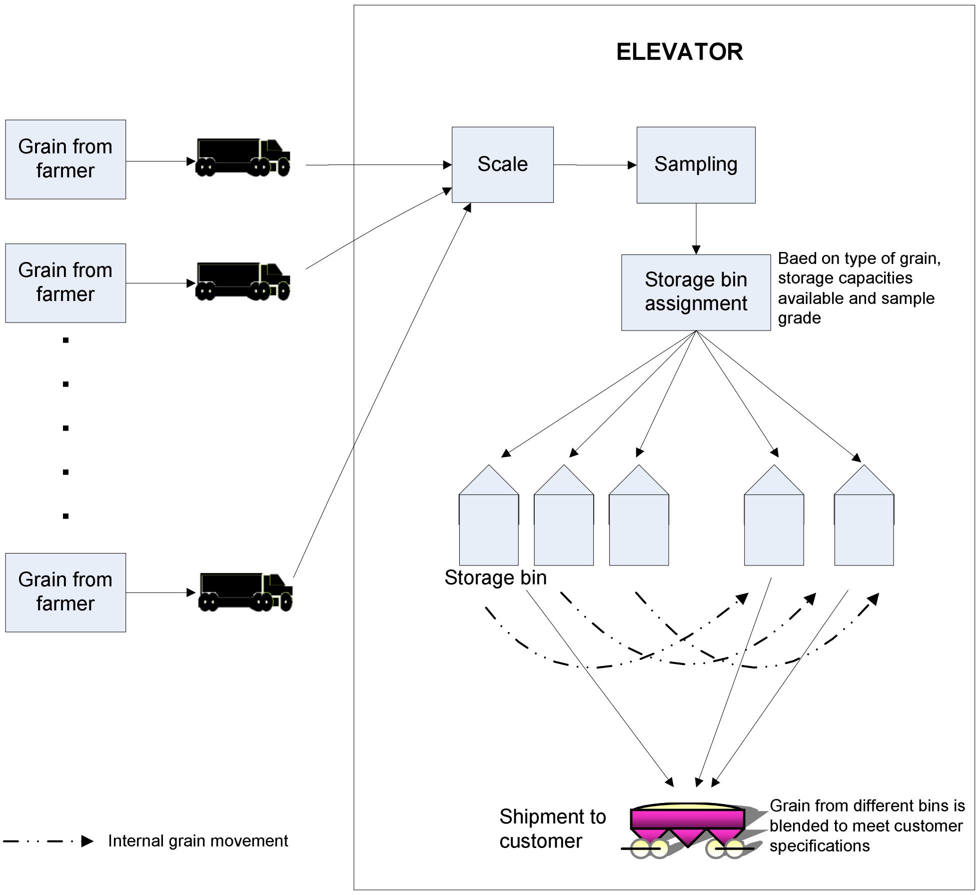


Figure 2. Symbols used in an E-R model

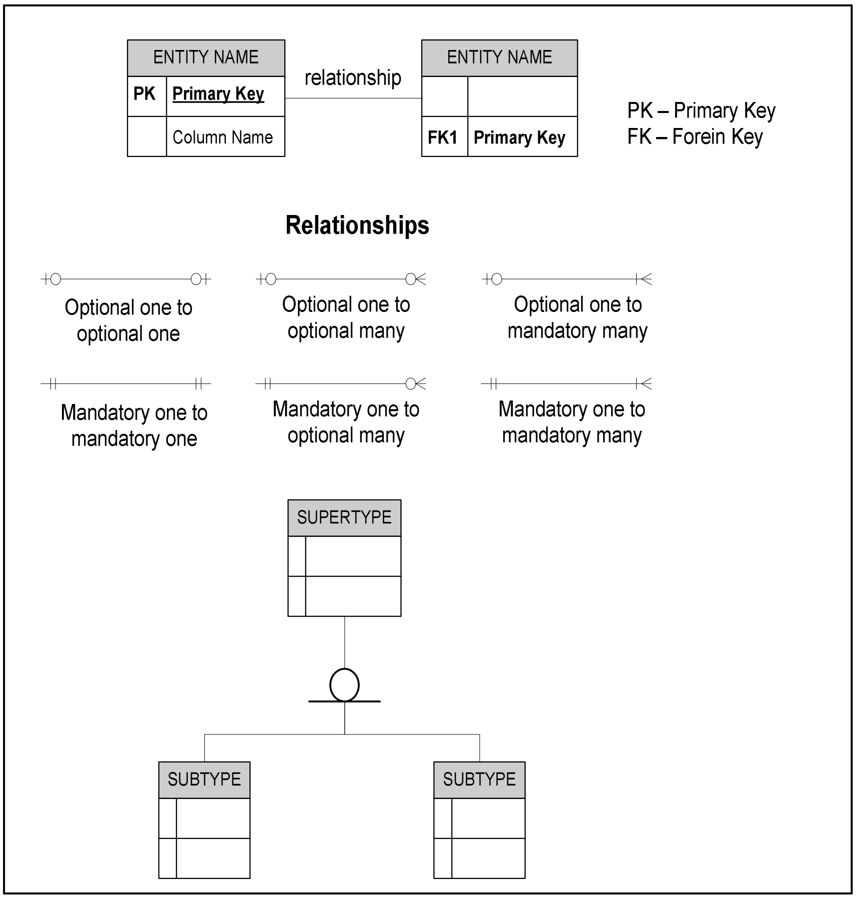


Figure 3. Entity-Relationship Diagram for internal traceability at a grain elevator

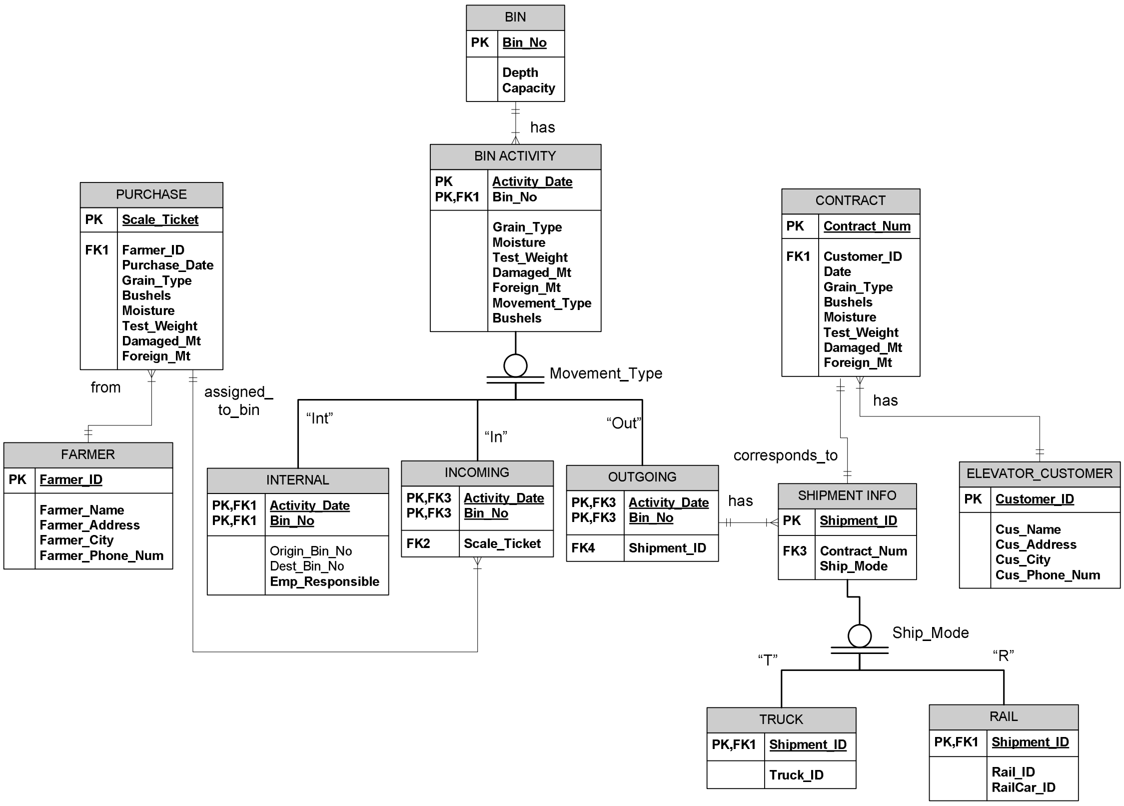
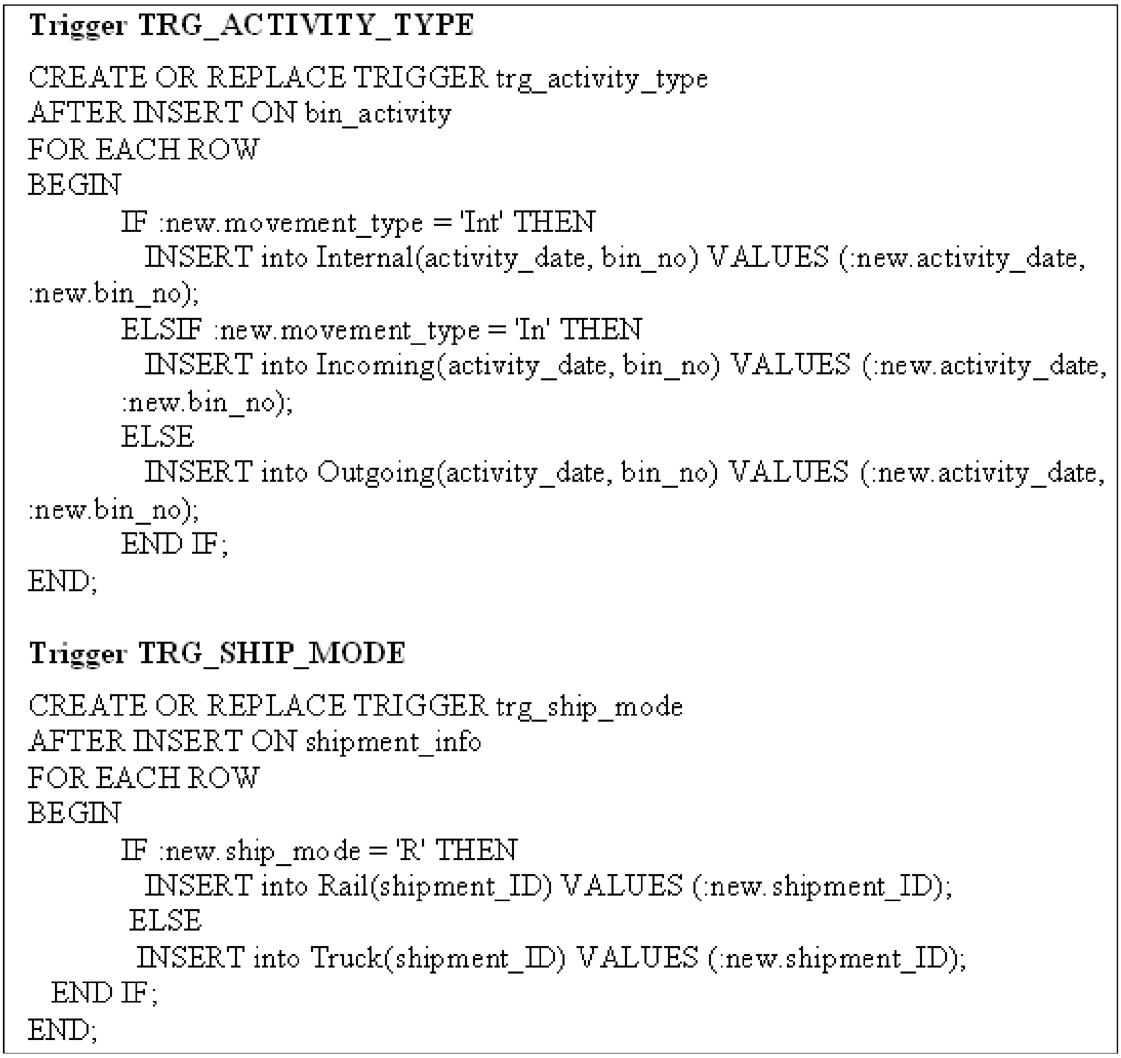


Figure 4. Database triggers used for entities *bin_activity* and *shipment_info*



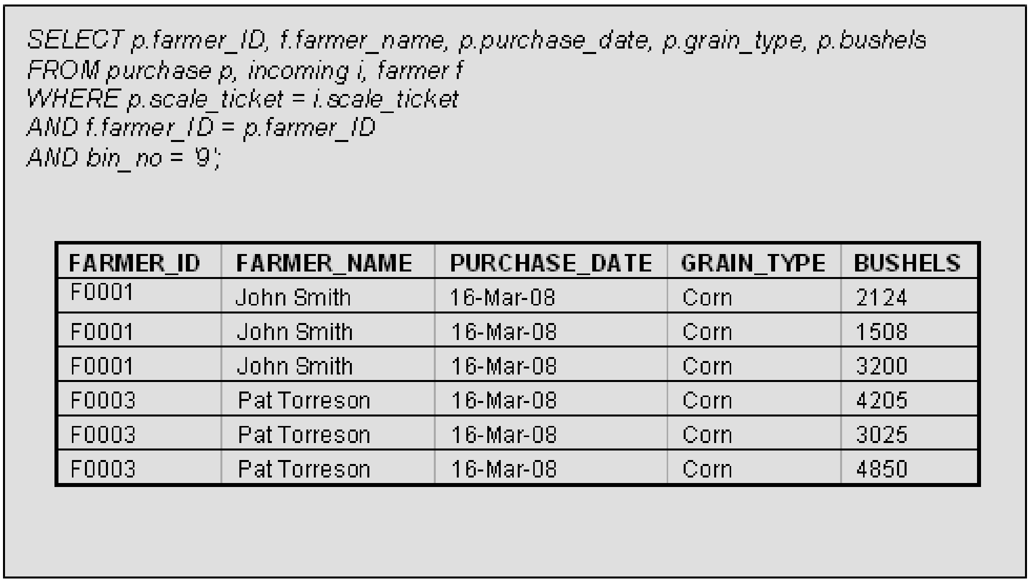
Trigger TRG_ACTIVITY_TYPE

```
CREATE OR REPLACE TRIGGER trg_activity_type
AFTER INSERT ON bin_activity
FOR EACH ROW
BEGIN
        IF :new.movement_type = 'Int' THEN
          INSERT into Internal(activity_date, bin_no) VALUES (:new.activity_date,
:new.bin_no);
        ELSIF :new.movement_type = 'In' THEN
          INSERT into Incoming(activity_date, bin_no) VALUES (:new.activity_date,
        :new.bin_no);
        ELSE
          INSERT into Outgoing(activity_date, bin_no) VALUES (:new.activity_date,
:new.bin_no);
        END IF;
END;
```

Trigger TRG_SHIP_MODE

```
CREATE OR REPLACE TRIGGER trg_ship_mode
AFTER INSERT ON shipment_info
FOR EACH ROW
BEGIN
        IF :new.ship_mode = 'R' THEN
          INSERT into Rail(shipment_ID) VALUES (:new.shipment_ID);
         ELSE
         INSERT into Truck(shipment_ID) VALUES (:new.shipment_ID);
  END IF;
END;
```

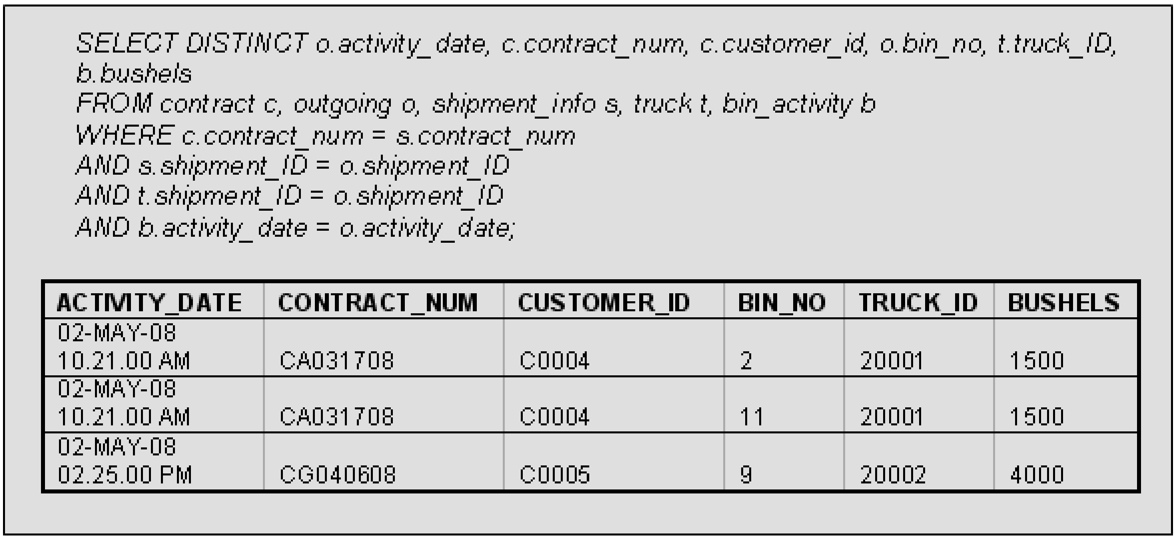
Figure 5. Sample query and report generated for incoming lot information



```
SELECT p.farmer_ID, f.farmer_name, p.purchase_date, p.grain_type, p.bushels
FROM purchase p, incoming i, farmer f
WHERE p.scale_ticket = i.scale_ticket
AND f.farmer_ID = p.farmer_ID
AND bin_no = '9';
```

FARMER_ID	FARMER_NAME	PURCHASE_DATE	GRAIN_TYPE	BUSHELS
F0001	John Smith	16-Mar-08	Corn	2124
F0001	John Smith	16-Mar-08	Corn	1508
F0001	John Smith	16-Mar-08	Corn	3200
F0003	Pat Torreson	16-Mar-08	Corn	4205
F0003	Pat Torreson	16-Mar-08	Corn	3025
F0003	Pat Torreson	16-Mar-08	Corn	4850

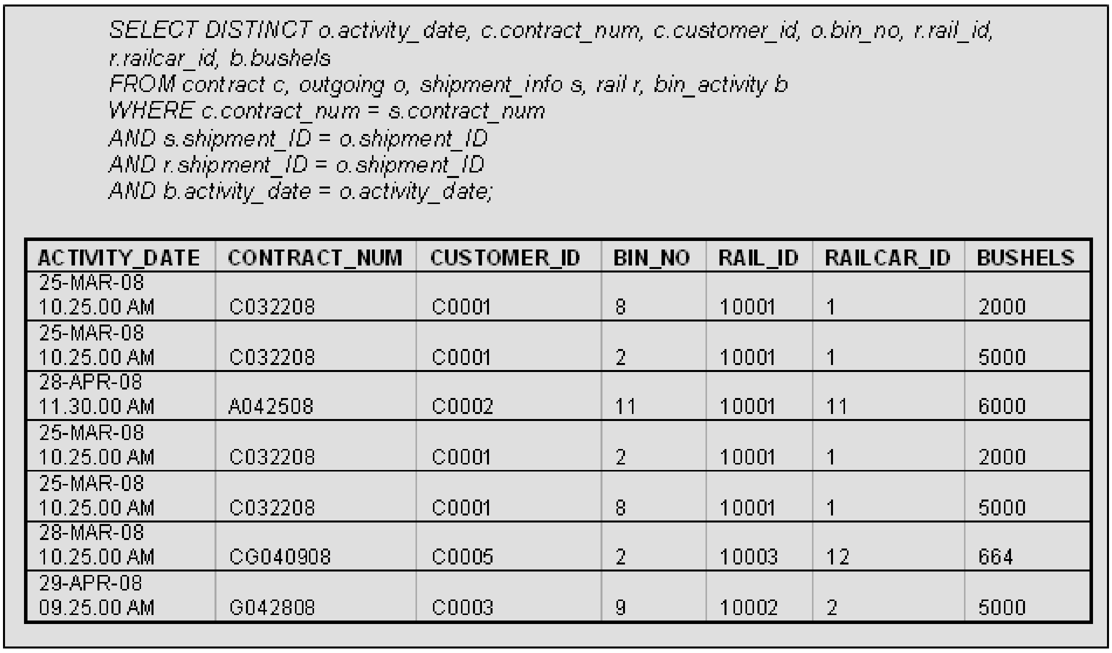
Figure 6. Sample query and report generated for outgoing lot information using truck as transportation mode



```
SELECT DISTINCT o.activity_date, c.contract_num, c.customer_id, o.bin_no, t.truck_ID,
b.bushels
FROM contract c, outgoing o, shipment_info s, truck t, bin_activity b
WHERE c.contract_num = s.contract_num
AND s.shipment_ID = o.shipment_ID
AND t.shipment_ID = o.shipment_ID
AND b.activity_date = o.activity_date;
```

ACTIVITY_DATE	CONTRACT_NUM	CUSTOMER_ID	BIN_NO	TRUCK_ID	BUSHELS
02-MAY-08 10.21.00 AM	CA031708	C0004	2	20001	1500
02-MAY-08 10.21.00 AM	CA031708	C0004	11	20001	1500
02-MAY-08 02.25.00 PM	CG040608	C0005	9	20002	4000

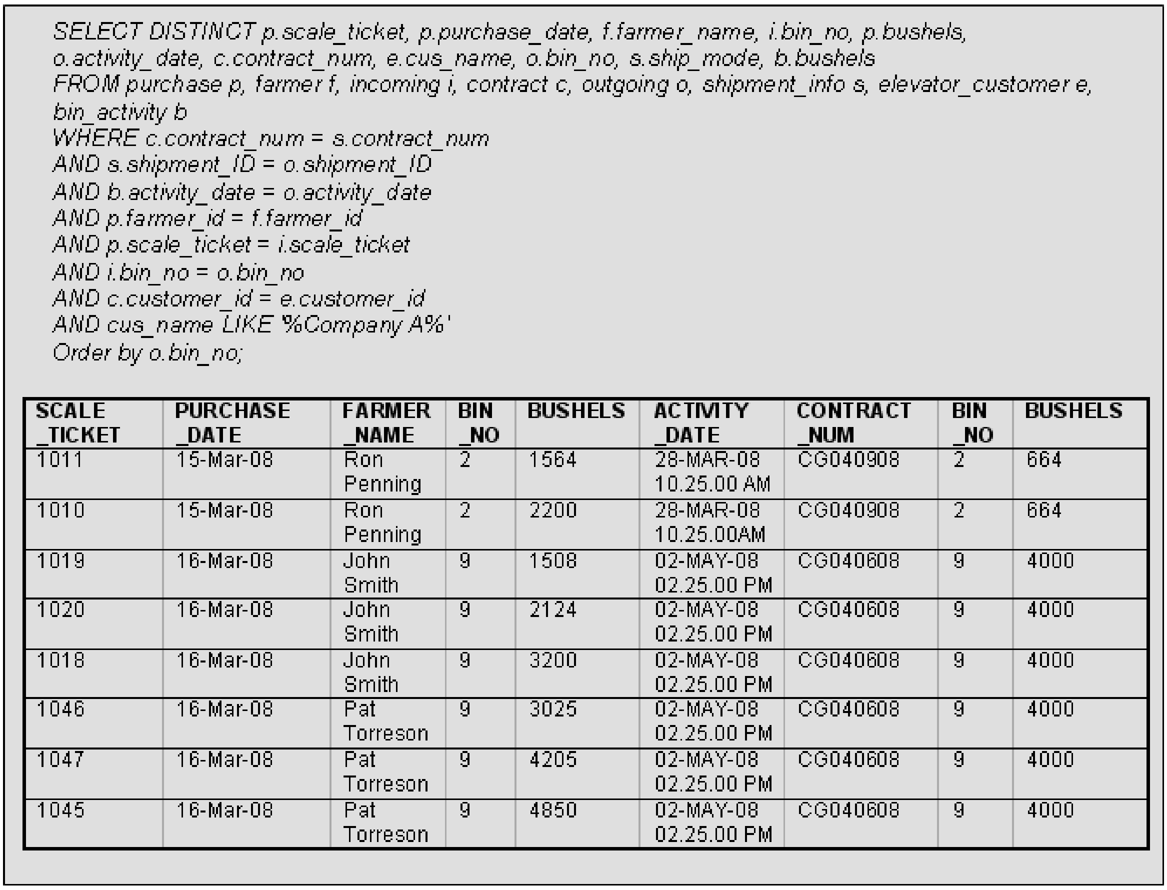
Figure 7. Sample query and report generated for outgoing lot information using railcars as transportation mode



```
SELECT DISTINCT o.activity_date, c.contract_num, c.customer_id, o.bin_no, r.rail_id,
r.railcar_id, b.bushels
FROM contract c, outgoing o, shipment_info s, rail r, bin_activity b
WHERE c.contract_num = s.contract_num
AND s.shipment_ID = o.shipment_ID
AND r.shipment_ID = o.shipment_ID
AND b.activity_date = o.activity_date;
```

ACTIVITY_DATE	CONTRACT_NUM	CUSTOMER_ID	BIN_NO	RAIL_ID	RAILCAR_ID	BUSHELS
25-MAR-08 10.25.00 AM	C032208	C0001	8	10001	1	2000
25-MAR-08 10.25.00 AM	C032208	C0001	2	10001	1	5000
28-APR-08 11.30.00 AM	A042508	C0002	11	10001	11	6000
25-MAR-08 10.25.00 AM	C032208	C0001	2	10001	1	2000
25-MAR-08 10.25.00 AM	C032208	C0001	8	10001	1	5000
28-MAR-08 10.25.00 AM	CG040908	C0005	2	10003	12	664
29-APR-08 09.25.00 AM	G042808	C0003	9	10002	2	5000

Figure 8. Sample query and report generated to connect incoming and outgoing lot information



```
SELECT DISTINCT p.scale_ticket, p.purchase_date, f.farmer_name, i.bin_no, p.bushels,
o.activity_date, c.contract_num, e.cus_name, o.bin_no, s.ship_mode, b.bushels
FROM purchase p, farmer f, incoming i, contract c, outgoing o, shipment_info s, elevator_customer e,
bin_activity b
WHERE c.contract_num = s.contract_num
AND s.shipment_ID = o.shipment_ID
AND b.activity_date = o.activity_date
AND p.farmer_id = f.farmer_id
AND p.scale_ticket = i.scale_ticket
AND i.bin_no = o.bin_no
AND c.customer_id = e.customer_id
AND cus_name LIKE '%Company A%'
Order by o.bin_no;
```

SCALE _TICKET	PURCHASE _DATE	FARMER _NAME	BIN _NO	BUSHELS	ACTIVITY _DATE	CONTRACT _NUM	BIN _NO	BUSHELS
1011	15-Mar-08	Ron Penning	2	1564	28-MAR-08 10.25.00 AM	CG040908	2	664
1010	15-Mar-08	Ron Penning	2	2200	28-MAR-08 10.25.00AM	CG040908	2	664
1019	16-Mar-08	John Smith	9	1508	02-MAY-08 02.25.00 PM	CG040608	9	4000
1020	16-Mar-08	John Smith	9	2124	02-MAY-08 02.25.00 PM	CG040608	9	4000
1018	16-Mar-08	John Smith	9	3200	02-MAY-08 02.25.00 PM	CG040608	9	4000
1046	16-Mar-08	Pat Torreson	9	3025	02-MAY-08 02.25.00 PM	CG040608	9	4000
1047	16-Mar-08	Pat Torreson	9	4205	02-MAY-08 02.25.00 PM	CG040608	9	4000
1045	16-Mar-08	Pat Torreson	9	4850	02-MAY-08 02.25.00 PM	CG040608	9	4000

Table 1. Transformations associated with each grain lot activity

Activity type	*Transformation*
Incoming grain purchased from farmer and transferred to a storage bin	1. Transfer: Incoming grain lot is transferred to one or more storage bins 2. Aggregation: Incoming lot is mixed with grain present in the assigned bin/s 3. Storage: Incoming lot is stored in assigned bin/s until next transformation occurs
Grain is transferred internally from one bin to another	1. Transfer: Internal grain lot is transferred to one or more storage bins 2. Segregation: A part of an internal lot (storage bin) is transferred to other bin/s 3. Aggregation: The transferred lot is mixed with grain present in the assigned bin/s 4. Storage: The transferred lot is stored in assigned bin/s until next transformation occurs
Grain lots from different storage bins are blended and shipped to the customer	1. Transfer: A part or entire internal lot (storage bin) is transferred from a bin 2. Segregation: A part of an internal lot (storage bin) is drawn from a bin for blending 3. Aggregation: The grain from different bins is blended together

Table 2. Relationship types in an Entity-Relationship model

Relationship type	*Description*
One-to-One	There is exactly one instance in table A that corresponds to exactly one instance in related table B
One-to-Many	There is exactly one instance in table A that corresponds to many instances in related table B
Many-to-One	There are many instances in table A that correspond to exactly one instance in related table B

Table 3. Description of entities in the ER model

Table Name (Entity)	*Attribute Name*	*Contents*
BIN	Bin_No	Grain storage bin number
	Depth	Bin depth (ft)
	Capacity	Bin capacity (Bushels)
BIN_ACTIVITY	Activity_Date	Bin activity date
	Bin_No	Grain storage bin number
	Grain_Type	Type of grain moved (Corn or Soybeans)
	Moisture	Average Moisture content of grain in the bin (%)
	Test_Weight	Average Test weight of grain in the bin (lb/Bu)
	Damaged_Mt	Average Percentage of damaged grain in the bin (%)
	Foreign_Mt	Average Percentage of foreign material in the bin (%)
	Movement_Type	Type of movement (Internal, Inbound or Outbound)
	Bushels	Quantity of grain moved in Bushels
INTERNAL	Activity_Date	Bin activity date
	Bin_No	Grain storage bin number
	Origin_Bin_No	Grain origin bin number
	Dest_Bin_No	Grain destination bin number
	Emp_Responsible	Name of employee responsible for moving grain
INCOMING	Activity_Date	Bin activity date
	Bin_No	Grain storage bin number
	Scale_Ticket	Scale ticket number of inbound grain in elevator
OUTGOING	Activity_Date	Bin activity date

Table Name (Entity)	*Attribute Name*	*Contents*
	Bin_No	Grain storage bin number
	Shipment_ID	ID of outbound shipment
SHIPMENT_INFO	Shipment_ID	ID of outbound shipment
	Contract_Num	Contract number of shipment
	Ship_Mode	Shipment mode (Truck or Rail)
TRUCK	Shipment_ID	ID of outbound shipment
	Truck_ID	ID of truck for outbound shipment
RAIL	Shipment_ID	ID of outbound shipment
	Rail_ID	ID of rail for outbound shipment
	Railcar_ID	ID of railcar for outbound shipment
ELEVATOR_CUSTOMER	Customer_ID	Customer ID
	Cus_Name	Customer name
	Cus_Address	Customer address
	Cus_City	Customer city
	Cus_Phone_Num	Customer phone number
CONTRACT	Contract_Num	Contract number -outbound shipment
	Customer_ID	Customer ID for shipment
	Contract_Date	Date of contract
	Grain_Type	Type of grain
	Bushels	Quantity of grain required in Bushels
	Moisture	Max. Moisture content of grain required on contract (%)
	Test_Weight	Min. test weight of grain required on contract (lb/Bu)
	Damaged_Mt	Max. allowable damaged grain on contract (%)
	Foreign_Mt	Max. allowable foreign material on contract (%)
FARMER	Farmer_ID	Farmer ID
	Farmer_Name	Farmer name
	Farmer_Address	Farmer address
	Farmer_City	Farmer city
	Farmer_Phone_Num	Farmer phone number
PURCHASE	Scale_Ticket	Scale ticket number of inbound grain in elevator
	Farmer_ID	Farmer ID
	Purchase_Date	Date of purchase
	Grain_Type	Type of grain purchased (Corn or Soybeans)
	Bushels	Quantity of grain purchased in Bushels
	Moisture	Moisture content of grain purchased (%)
	Test_Weight	Test Weight of grain purchased (lb/Bu)
	Damaged_Mt	Damaged matter in grain purchased (%)
	Foreign_Mt	Foreign matter in grain purchased (%)

CHAPTER 5. A multi-objective optimization approach to balancing cost and traceability in bulk grain handling

Manuscript to submitted to the *Journal of Food Engineering*

Maitri Thakur[1, 2,] *, Lizhi Wang[2] and Charles R. Hurburgh[1, 3]

[1] Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011
[2] Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA 50011
[3] Department of Food Science and Human Nutrition, Iowa State University, Ames, IA 50011
* Primary author, Corresponding author.

Abstract

This paper proposes a multi-objective optimization model to minimize lot aggregation at a grain elevator. The problem involves blending of bulk grain to meet customer specifications while reducing the food safety risk by minimizing the aggregation of different grain lots. A mathematical multi-objective mixed integer programming (MIP) model is proposed with two objective functions. The objective functions allow in calculating the minimum levels of lot aggregation and minimum total cost of blending grain to meet the customer contract specifications. Constraints on the system include customer contract specifications, availability of grain at the shipping elevator location as well as other locations and the blending requirements. The solutions include the quantities of grain from different storage bins to be used for blending for a shipment while using the minimum number of storage bins and the total cost. The total cost includes transportation cost between elevator locations, blending cost and the discount applied to the shipment when customer specifications are not met. The numerical results are presented for a corn shipment scenario to demonstrate the application of this model to bulk grain blending. Pareto optimal front is computed for the problem for simultaneous optimization of lot aggregation and cost of blending. The Pareto front provides a set of optimal solutions for different blending options for the elevator management to choose from. Sensitivity analysis is conducted to analyze the application of the model under different operating conditions. This model provides an effective method for minimizing the traceability effort by minimizing the food safety risk caused by lot aggregation. Besides minimizing the lot aggregation, the model also allows in using the maximum volume of grain present in a given storage bin which leads to emptying of the bins and the extent of aggregation of old grain lots with the new incoming lots can decrease considerably. Use of fewer bins for blending shipments is also

easier logistically and can lead to additional savings in terms of grain handling cost and time.

Keywords: Lot aggregation, multi-objective optimization, traceability, bulk grain handling, food safety risk

1 Introduction

Food safety and food control continue to gain significant attention as our food supply chains and production practices become increasingly complex. Food safety is in fact a very important part of public health, and although several advanced surveillance and monitoring systems exist in developed countries, outbreaks of foodborne diseases continue to be commonplace. Such foodborne diseases are caused by consumption of contaminated foods or beverages. There are many different types of foodborne infections as many disease-causing microbes or pathogens can contaminate foods. In addition to these, several poisonous chemicals can also cause foodborne diseases if present in food (CDC, 2005). According to the Center for Disease Control and Prevention (CDC), an outbreak of foodborne illness occurs when a group of people consume the same contaminated food and two or more of them come down with the same illness. CDC (2005) estimates that foodborne diseases cause 76 million illnesses, 325,000 hospitalizations, and 5000 deaths in the United States every year.

Consumers all over the world have faced various food safety and health issues in the recent years. This has led to a growing interest in developing systems for food supply chain traceability (Carriquiry and Babcock, 2007; Folinas et al., 2006; Jansen-Vullers et al., 2003; Madec et al., 2001; McKean, 2001). Various food safety and traceability guidelines and regulations exist in several countries. Under the European Union Law, ‘‘traceability” is defined as the ability to track any food, feed, food-producing animal or substance that will be used for consumption, through all the stages of production, processing and distribution (Official Journal of the European Communities, 2002). It is a risk-management tool that allows food business operators or authorities to withdraw or recall products which have been identified as unsafe. In the United States, the Bioterrorism Act of 2002 requires that all companies involved in the food and feed industry to self-register with the Food and Drug Administration and maintain records and information for food traceability purposes (US Food and Drug Administration, 2002). In Canada, Can-Trace was launched in July 2003 which is a collaborative and open initiative committed to the development of traceability standards for all food products sold in Canada (Can-Trace, 2003).

Traceability is important for many reasons such as responding to the food security threats to documenting chain of custody, documenting production practices, meeting regulatory compliance, and analyzing logistics and production costs. Moe (1998) defines traceability as the ability to track a product batch and its history through the whole, or part, of a production chain from harvest through transport, storage, processing, distribution and sales or internally in one of the steps in the chain (for example the production step). The General Food Law (Official Journal of the European Communities, 2002) requires traceability throughout the food supply chain. In order to be able to track and trace products throughout the supply chain, food business operators must maintain relevant information from the suppliers and keep track of all products and their transformation through all stages of production and then pass this information to the next link in the supply chain (Donnelly et al., 2009; Schwägele, 2005; Thakur and Hurburgh, 2009). Senneset et al. (2007) state that in order to achieve chain traceability, the identities of traceable units must be recorded at reception and shipping, and that internal traceability requires recording of all transformations during the production process.

1.1 **Concept of lot aggregation**

Many papers have addressed the concept of traceability in terms of ensuring food safety and quality by implementation of information systems in food supply chains (Donnelly et al., 2009; Schwägele, 2005; Thakur and Hurburgh, 2009, Senneset et al., 2007). Laux (2007) presented a quality management systems approach for ensuring product quality and traceability at a grain elevator. Little research has been conducted on the cost and benefits of such systems. While consumers demand more in terms of food safety and quality, for food industry, a thorough investigation into the cost of such systems is very important. Food production involves blending or mixing of several ingredients and batches that constitute the final product. Several product transformations take place in food production, including, splitting, mixing, cooking, destruction, etc. of product or ingredient lots. Lot aggregation occurs when several product batches or lots are used to produce the finished product. It is common in food industry to utilize a proportion of a product lot in one batch of the finished product and the remaining portion can be used for subsequent production batches. So, a contaminated ingredient lot can in turn contaminate several production batches.

For the food industry, the emphasis is not only to decrease the food safety incidents (and recalls) but also limit the number of batches that constitute a given finished product in order to decrease the product quantities to be recalled (Dupuy et al., 2005). For instance, after a recall of minced beef products due to BSE, a French producer not only improved the accuracy of their traceability system but also decreased the number of mixed batches of meat in one batch of minced beef (Gattengo, 2001). Dupuy et al. (2005) proposed a batch dispersion model to optimize traceability in food industry by minimizing the batch size and batch mixing. This model calculates the minimum batch dispersion which is given by the sum of links between the raw material batches and the finished product batches. This model, however, does not take into account the additional cost that might be incurred in trying to minimize the number of batches used in production. Furthermore, certain food products like bulk grain need to be blended in order to meet the trade specifications.

1.2 **Mathematical programming for blending problems**

The mathematical programming approach has been extensively used for many blending problems. Shih and Frey (1995) proposed a coal blending optimization model to minimize the expected costs of coal blending while minimizing the expected sulphur emissions. Singh et. al. (2000) proposed a gasoline blend optimization model that could provide competitive benefit for oil refiners. While mathematical models have been used for blending optimization of bulk products like coal, wine, and gasoline, the application to grain blending is limited to minimizing discounts. Sivaraman et al. (2002) presents a general mathematical model to determine the optimal grain blending and segregation strategies to maximize the sale premiums based on protein content of wheat. Bilgen and Ozkarahan (2007) addresses the blending and shipping problem faced by a company that manages a wheat supply chain by formulating the problem as a mixed-integer linear programming model. A mixed-integer program (MIP) is a linear program with additional constraints that some of the variables must take on integer values. A multi-objective optimization models simultaneous optimizes several conflicting objectives. Such models have the advantage of accurately representing the real multi-criteria nature of certain situations (Benayoun, et. al. 1971). In order to address the food traceability concerns, there is a need to develop techniques to solve two aspects; to minimize the number of batches that are used to produce a finished product and to maximize the profits at the same time.

In this paper, we use a multi-objective optimization model to control the aggregation of different lots or batches of bulk grain product while minimizing the total cost of blending grain. The next section provides a description of bulk grain handling scenario.

1.3 **Bulk grain handling**

Grain elevators handle bulk commodities marketed against generic grade standards that are based on physical attributes. Grain lots are commingled in order to meet buyer specifications and to maximize the profit. As a result of this commingling, lot identity is not maintained. Grain storage bins are extensively used to handle bulk grain and one storage bin can contain grain from many different sources. The elevator buys grain with different quality characteristics in terms of moisture, test weight, damaged material and foreign material from the farmers. These incoming grain lots are assigned to one or more storage bins depending on the quality and space constraints. As a result, one storage bin can contain grain from many different sources.

Figure 1 shows a typical bulk grain handling scenario. The incoming grain lots from the farmers are assigned a unique scale-ticket number, weighed and graded based on quality parameters. These quality parameters include moisture, test weight, damaged material and foreign material. A quality grade is determined based on these parameters and the lot is assigned and transferred to one or more storage bins based on space and quality constraints. Grain is kept in storage until it is shipped to a customer. When in storage, a part or entire contents of a bin can be transferred to other bins in order to avoid spoilage due to environmental conditions (usually related on increasing temperature inside a bin). Finally, grain for the outgoing shipments is blended from several bins in order to meet the customer specifications for quality, shown in Figure 2. As a result of this process, one storage bin can contain grain from many different sources. A specific grain lot shipped to a manufacturing plant in turn can contain grain from all these sources that can end up in the finished product. In case of a food related emergency, it would be almost impossible to isolate the source with the problem which would lead to a recall of all the finished goods that might have a chance of being contaminated. This process is very time intensive, increases the recall costs, and can lead to a tainted brand name for the company. Many food recall incidents have taken place in the past that have affected the consumers and the producers alike. For instance, according to a news report, after the tomato-salmonella scare in June 2008, the Florida tomato industry could have potentially lost $40 Million because the producers could not sell their tomatoes until the source of salmonella outbreak was identified (Reuters, 2008). With fragile and quickly perishable items like tomatoes, the consequences on industry and growers/producers can be irreparable. The grain trade units must be tracked efficiently from the farm to the consumer to avoid such problems.

In addition to keeping track of all the product transformation in the food supply chain, it is important to develop operational techniques that can help in reducing the food safety risk. Of all the product transformation, mixing or blending of different lots or batches is the most difficult to track in bulk grain handling industry (Thakur and Hurburgh, 2009). As grain is drawn from different storage bins for blending and shipping to the customers, most of the bins are not emptied and more incoming grain (bought from the farmers) is transferred to these bins. This practice leads to a state of continuous lot aggregation and several individual grain lots get mixed while in storage at the grain elevators. In case of a contamination, the problem can spread very rapidly because of the mixing leading to an increased food safety risk. We study the problem of lot aggregation and propose a model for minimizing the lot aggregation which in turn would reduce the food safety risk due to mixing of lots keeping with the business model of minimizing the total cost of blending the grain for shipment.

2 Problem description

The problem under study is taken from an Iowa co-op, Farmers Cooperative (FC) Company that handles bulk commodities including corn and soybeans. The elevator blends and sells the bulk grain to its customers. Different grain lots from various bins are blended to meet the customer contract specifications. A discount is applied if the given shipment does not meet the specifications. There are no premiums if the quality is better than what is required. So, the objective while blending different lots is to be as close to the specifications as possible. While the elevator blends grain to meet the specifications, there are no restrictions on the number of bins that can be used. A specific grain load shipped to a customer can contain grain from all available. In case of a food related emergency, it would be almost impossible to connect the source with the problem, which would lead to a recall of all the finished goods that might have a chance of being contaminated. This process is very time consuming, increases recall costs, and can lead to a tainted brand name. So, the risk in case of a food safety increases. Currently, the FC Company uses blending optimization software with a goal of minimizing the discounts (in turn, maximizing net profit). Minimization of food safety risk is not considered in this model. In most cases, all bins contribute to an outgoing shipment. Only a fraction of the total volume of grain present in a bin is used for blending, so the bins are not emptied. New incoming lots are constantly added to bins already containing grain. This causes a continual aggregation state and many grain lots get commingled even before they are blended for shipment. Food safety risk is not considered by the elevator.

FC has several elevator locations throughout the state of Iowa as shown in Figure 3. Since, the goal is to meet the customer specifications, in an event when the required volume of grain is not available at the shipping location, the remaining amount can be transported from other locations (Hemphill, 2009). The blending optimization technique currently used by FC focuses only on minimizing the discount and does not take into consideration the transportation and blending cost or the food safety risk that can occur when grain from several storage bins is used to blend the product for a single shipment.

3 Multi-objective optimization

Due to multiple objective nature of this problem, we propose a multi-objective mixed integer program for simultaneous improvement of the blending practices of the elevator and the total cost of blending and loading the railcars for grain shipment to the customers.

A general form of the multi-objective linear problem with two objectives can be expressed as:

$$max\ \{\varsigma c = c^T x\ and\ \varsigma d = d^T x : Ax \leq b, x \geq 0\} \qquad (1)$$

In multi-objective problems, a single solution that optimizes both objectives may not exist. In such cases, a group of trade-off solutions can be computed by Pareto optimization technique (Deb, 2001).

In Pareto optimization, each of the solutions x in the decision space has a vector $z(x) = \{z1(x), z2(x),..., zk(x)\}$ of objective values that represents the trade-off between the objectives. The Pareto optimal front is the set of solutions that contains all solutions that are not dominated by any other solution in the entire feasible search space. A solution x_1 dominates x_2 if none of the components in x_1 is worse than the corresponding value in x_2 and at least one of the components in x_1 is strictly better than its corresponding value in x_2 (Deb, 2005). In the context of our work, the Pareto optimal front represents the set of blending options (quantity of grain drawn from specific storage bins) with an optimal trade-off between total blending cost and level of lot aggregation. The following factors further define the problem:

- FC has several elevator locations throughout the state of Iowa. Each location has multiple bins that store grain bought from the farmers.
- Grain may be sold months in advance but the customer normally notifies the elevator one or two days in advance before railcars arrive for loading.
- There is not always enough grain available at the elevators for shipment.
- In an event when the required volume of grain is not available at the shipping location, the remaining amount can be transported from other locations.
- While determining the location from where the remaining volume is transported, the elevator considers factors such as product availability.

3.1 **Mathematical model**

This section presents the mathematical model for grain blending and cost optimization. We describe the parameter notations and definitions used in the model followed by the description of the objective functions and constraints.

The blending and cost optimization problem is presented as a multi-objective mixed integer model with two objectives:

1. Minimize the number of storage bins used to blend grain for a given shipment. This includes the stoarge bins from all elevator locations from where additional grain can be transported in an event when sufficient volume is not available at the shipping location.
2. Minimize the total cost for blending and shipping grain. The total cost includes the discount given to customer when contract specifications are not met, the cost of transporting grain between different locations and the blending cost.

Shipment discount

A discount is applied to the shipment when the blended grain does not meet customer contract specifications for quality. This discount is calculated as dollars per bushel. The shipment discount is expressed by the following equation:

$$\text{Discount (\$)} = Cb(DmZm + DtZt + DdZd + DfZf) \tag{2}$$

Transportation cost

As explained earlier, when the shipping location does not have the required volume of grain available for shipment, additional grain is transported from other elevator locations and a transportation cost is incurred which is expressed by the following equation:

$$\text{Transportation cost (\$)} = \sum_{i\epsilon I}(Cii \sum_{j\epsilon J} Xij) \tag{3}$$

Blending cost

A blending cost is incurred at the shipping location where grain from several storage bins is blended and loaded on railcars for shipment. The blending cost is expressed by the following equation:

$$\text{Blending cost (\$)} = Cblend \sum_{i\epsilon I} \sum_{j\epsilon J} Xij \tag{4}$$

3.2 **Objective functions**

The two objective functions of this model can be presented as:

Minimize:

$$\sum_{i\epsilon I} \sum_{j\epsilon J} Yij \tag{5}$$

Minimize:

$$Cb(DmZm + DtZt + DdZd + DfZf) + \sum_{i\epsilon I}(Cii \sum_{j\epsilon J} Xij) + Cblend \sum_{i\epsilon I} \sum_{j\epsilon J} Xij \tag{6}$$

Equation (4) minimizes the number of storage bins used to blend grain for a given shipment while equation (5) minimizes the total cost of blending and shipping grain that meets the customer contract specifications.

3.3 **Constraints**

The multi-objective mixed integer optimization model consists of the following constraints:

(1) Product availability

(2) Contract specifications and product discount schedule

3.3.1 Product availability

The product availability constraint corresponds to the availability of a specific quantity of grain required for a given contract. Also, the amount of grain that can be taken from any storage bin must be less than or equal to the quantity available in each bin, represented by equation (6). The definitions of all variables are provided in Table 1.

$$X\,ij \;\leq Bij\,Yij \tag{7}$$

3.3.2 Contract specifications and discount schedule

Each product shipment must meet the customer contract specifications for quantity as well as quality. Equation (7) specifies that the total quantity of grain drawn from all bins for blending must be equal to the customer shipment requirement.

$$\sum_{i\epsilon I} \sum_{j\epsilon J} Xij = Cb \tag{8}$$

The blended grain must meet the contract specifications for four quality factors; moisture, test weight, damaged material and foreign material. In case, the quality specifications are not met, a discount is applied to the shipment based on the product discount schedule. Equations (8) – (11) specify this requirement for each quality factor. The first term in each equation calculates the quality of the blended grain as an aggregate factor and the second term represents the discount penalty that would be incurred if the requirements are not met.

$$\frac{\sum_{i\epsilon I} \sum_{j\epsilon J} Xij\,Mij}{Cb} - Zm \;\leq Cm \tag{9}$$

$$\frac{\sum_{i\epsilon I} \sum_{j\epsilon J} Xij\,Tij}{Cb} + Zt \;\geq Ct \tag{10}$$

$$\frac{\sum_{i\epsilon I} \sum_{j\epsilon J} Xij\,Dij}{Cb} - Zd \;\leq Cd \tag{11}$$

$$\frac{\sum_{i\epsilon I} \sum_{j\epsilon J} Xij\,Fij}{Cb} - Zf \;\leq Cf \tag{12}$$

Equations (12) – (15) calculate the quality of the blended grain that is shipped to the customer.

$$Sm = \frac{\sum_{i\epsilon I} \sum_{j\epsilon J} Xij\ Mij}{Cb} \tag{13}$$

$$St = \frac{\sum_{i\epsilon I} \sum_{j\epsilon J} Xij\ Tij}{Cb} \tag{14}$$

$$Sd = \frac{\sum_{i\epsilon I} \sum_{j\epsilon J} Xij\ Dij}{Cb} \tag{15}$$

$$Sf = \frac{\sum_{i\epsilon I} \sum_{j\epsilon J} Xij\ Fij}{Cb} \tag{16}$$

Equation (16) defines the allowed values for all decision variables used in the optimization model.

$$Xij \geq 0, Yij \ \epsilon \ \{0,1\}, Zij \geq 0, Zm, Zt, Zd, Zf, Sm, St, Sd, Sf \geq 0 \tag{17}$$

The inputs and outputs of the multi-objective optimization model for grain blending are presented in Figure 4.

4 Computational study and results

The computational experiments carried out on a real application are presented in this section. The proposed multi-objective mixed integer optimization model for grain blending was applied to a real elevator situation that blends and ships bulk grain including corn and soybeans. Twenty elevator locations were selected where each location has between ten to fifteen grain storage bins. Corn was selected as the product and elevator location A was the shipping location for the computational study. Location A receives a customer order to ship one million bushels of corn. The quality factors included in the customer contract and the discount schedule for corn are presented in Table 2.

The GLPK (GNU Linear Programming Kit) was used to solve the optimization problem. GLPK is intended for solving large-scale linear programming (LP), mixed integer linear programming (MIP), and other related problems by means of the revised simplex method (GLPK, 2008).

The results obtained by solving the optimization problem for both objective functions separately are shown in Table 3. The total cost for blending and loading the grain for shipment when the objective is to minimize the blending by using the least number of storage bins is \$76,837. This is almost twice the total cost of \$40,157 computed when the objective is to minimize the cost of blending and loading grain. The quality of blended grain meets the customer contract specifications for each quality factor except moisture for Objective 1 and a total discount of $10,292 is applied to the shipment as shown in Table 4. While the total cost for Objective 2 does not contain any discount.

4.1 **Pareto optimal solutions**

The goal was to solve for the two objectives simultaneously by computing the Pareto optimal solutions. The Pareto optimal solutions are shown in Table 5 and the Pareto optimal frontier is shown in Figure 5. The quality of the blended grain for each Pareto optimal solution is also shown in Table 5. It can be noted from Table 5 that when the number of storage bins used for blending is low, the total cost of blending and preparing grain for shipment is higher. The grain storage bins are cleaned out only when they are emptied and in many cases they are not emptied for up to one year. New incoming grain lots are constantly added to the bins and the extent of aggregation can be immeasurable. A set of optimal solutions are calculated to provide the elevator management with various grain blending options so the blending decision can be made by considering the trade-off between cost and food safety risk.

4.2 **Sensitivity analysis**

To analyze the sensitivity of the grain blending model to different operating conditions, we studied the affect of changing the transportation cost and the contract specifications on the total cost and the level of lot aggregation. The transportation cost was increased in the increments of 10%. The percentage change in total cost, the resulting total cost as well as the number of bins used for blending grain for a shipment are shown in Table 6. The results are shown only for Objective 2 that minimizes the total cost of blending and preparing grain for shipment as Objective 1 does not contain the cost component. A 10% increase in transportation cost causes the total cost to increase by 7.2%. The cost of transporting grain between different elevator locations is an important component of total cost that includes three cost components, blending cost, transportation cost and discount. This shows that proper transportation planning between elevator locations can result in large monetary savings.

Next, we changed the customer contract specifications for moisture of blended corn and studied its affect on the blending results. The new moisture content required for the blended grain and percentage change in moisture content is shown in Table 7. The corresponding percentage change in cost of blending corn is also included. The cost is computed for the two objectives, one that minimizes the level of aggregation of grain lots and the second objective that minimizes the total cost of blending. It can be seen that the change in total cost when the objective is to minimize the cost is almost twice than the change in total cost when the objective is to minimize the level of aggregation.

5 Results and discussion

We present a comprehensive model for the bulk grain blending problem while addressing the problem of lot aggregation. The model allows simultaneous optimization of cost of blending and a control over the extent of mixing of individual grain lots. We compute the amount of grain to be taken from different storage bins to meet the customer contract specifications. In an event when the shipping elevator location does not have sufficient quantity available to meet the contract specifications, grain is transported from other locations. This paper makes two important contribution as we address the problem of grain blending to minimize the total cost that includes the blending cost, transportation cost and shipment discounts. Secondly, we incorporate the problem of minimizing the food safety risk (by controlling aggregation of lots) which in turn would minimize the traceability effort and the cost of recalls. The model integrates all of these factors simultaneously. Since the model has two objectives, we formulated the problem as a multi-objective mixed integer optimization. Pareto optimal front was also computed so that the elevator management has different blending options and they can consider the trade-offs between cost and food safety. Sensitivity analysis was conducted to study the application of the model under different operating conditions. We increased the transportation cost and changed the moisture specifications for blended grain and computed the blending options.

Usually, the grain storage bins are cleaned out only when they are emptied and in many cases they are not emptied for up to one year. New incoming grain lots are constantly added to the bins and the extent of aggregation can be immeasurable. Since this optimization model minimizes the number of bins used for blending a shipment; it in turn maximizes the proportion of grain drawn from these bins. This provides an opportunity for cleanouts and the aggregation with incoming lots can be reduced to a great extent. The use of this model would provide additional savings to the elevator company in terms of time and money used for handling the grain since the use of fewer numbers of bins is logistically easier. This model provides a good starting point for grain industry and can be used as an important strategic tool for decision making to meet two important requirements, minimizing the cost while simultaneously controlling the food safety risk.

Our future work will focus on developing models for optimal initial storage bin assignment policies for incoming grain at the elevator. We will focus on optimizing the storage assignment policies to minimize the level of lot aggregation at the incoming end of the elevator. The two models combined would provide an overall minimization of food safety risk caused by excessive lot aggregation.

Acknowledgements

The authors would like to thank Mr. Butch Hemphill of the Farmers Cooperative Company, Farnhamville, Iowa for providing the information about elevator operating strategies and providing the data for numerical examples.

References

Bilgen, B., Ozkarahan, I., 2007. A mixed-integer linear programming model for bulk grain blending and shipping. International Journal of Production Economics, 107: 555–571.

Benayoun, R., Montgolfier, J., Tergny, J., 1971. Linear programming with multiple objective functions: Step Method (STEM). Mathematical Programming, 1: 366-375.

Can-Trace, 2003. Agriculture and Agri-Food Canada. <http://www.can-trace.org>

Carriquiry, M., Babcock, B.A., 2007. Reputations, market structure and the choice of quality assurance systems in the food industry. American Journal of Agricultural Economics, 89: 12–23.

CDC, 2005. Foodborne Illness, Centers for Disease Control and Prevention. <http://www.cdc.gov/ncidod/dbmd/diseaseinfo/foodborneinfections_g.htm>.

Deb K. 2005. Multi-objective optimization. Berlin: Springer pp. 273–316.

Deb K., 2001. Multi-objective optimization using evolutionary algorithms. New York: Wiley.

Donnelly, K.A.-M., Karlsen, K.M., Olsen, P., 2009. The importance of transformations for traceability - A case study of lamb and lamb products. Meat Science, 83, 68-73.

Dupuy, C., Botta-Genoulaz, V., Guinet, A., 2005. Batch dispersion model to optimize traceability in food industry. Journal of Food Engineering, 70(3): 333-339.

Farmers Cooperative Company, 2009. <http://www.fccoop.com/>

Folinas, D., Manikas, I., Manos, B., 2006. Traceability data management for food chains. British Food Journal 108 (8), 622–633.

Gattegno, I. Soviba, ce qui a change´ depuis le 20 octobre 2000. 2001 RIA (Revue de l’Industrie Agroalimentaire). 609: 46–47.

GLPK (GNU Linear Programming Kit), 2008. <http://www.gnu.org/software/glpk/>

Hemphill, 2009. Personal Communication.

Jansen-Vullers, M.H., van Dorp, C.A., Buelens, A.J.M., 2003. Managing traceability information in manufacture. International Journal of Information Management 23, 395–413.

Laux, C.M., 2007. The Impacts of a Formal Quality Management System: A Case Study of implementing ISO 9001 at Farmers Cooperative Co., IA. Ph.D. Thesis, Iowa State University.

Madec, F., Geers, R., Vesseur, P., Kjeldsen, N., Blaha T., 2001. Traceability in the pig production chain. Revue Scientifique Et Technique (International Office of Epizootics) 20 (2), 523–537.

McKean, J.D., 2001. The importance of traceability for public health and consumer protection. Revue Scientifique Et Technique (International Office of Epizootics) 20 (2), 363–371.

Moe, T., 1998. Perspectives on traceability in food manufacture. Trends in Food Science & Technology 9(5), 211–214.

Official Journal of the European Communities. 2002. Regulation (EC) No 178/2008 of the European Parliament and the Council of 28 January 2002.

Reuters, 2008. North American tomato industry reeling growers, <http://www.reuters.com/article/idUSN6A33595920080610>

Schwägele, F., 2005. Traceability from a European perspective. Meat science 71(1), 164-173.

Senneset, G., Forås, E., Fremme, K.M., 2007. Challenges regarding implementation of electronic chain traceability. British Food Journal, 109(10), 805-818.

Shih, J., Frey, H.C., 1995. Coal blending optimization under uncertainty. Journal of Operations Research, 83(3): 452-465.

Singh, A., Forbes, J.F., Vermeer P.J., Woo, S.S., 2000. Model-based real-time optimization of automotive gasoline blending operations. Journal of Process Control, 10(1): 43-58.

Sivaraman, E., Lyford, C.P., Brorsen, B.W., 2002. A General Framework for Grain Blending and Segregation. Journal of Agribusiness, 20(2): 155-161.

Thakur, M., Hurburgh, C.R. ,2009. Framework for implementing traceability system in the bulk grain supply chain. Journal of Food Engineering, 95(4), 617-626.

US Food and Drug Administration, 2002. The Bioterrorism Act of 2002.

Figure 1. Grain handling process at an elevator

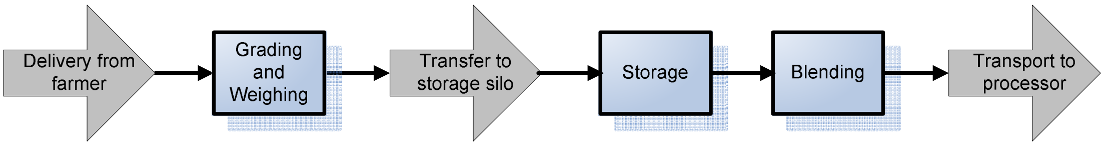


Figure 2. A typical bulk grain handling scenario

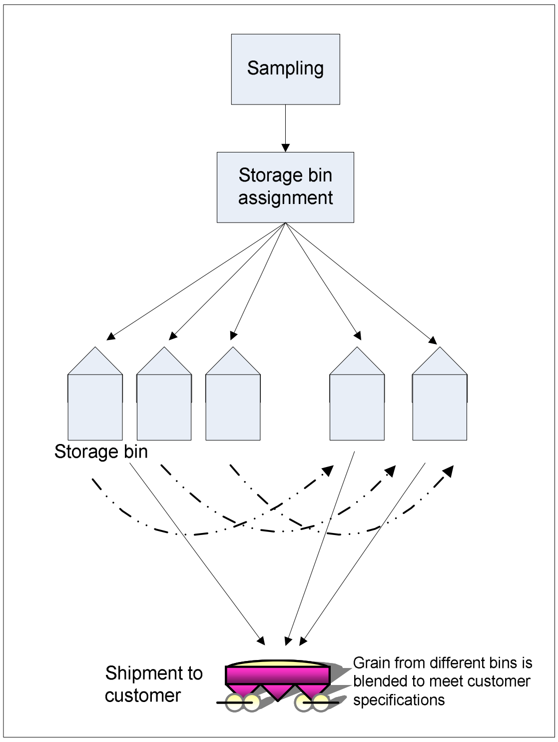


Figure 3. Farmers Cooperative Location Map (Farmers Cooperative Company, 2009)

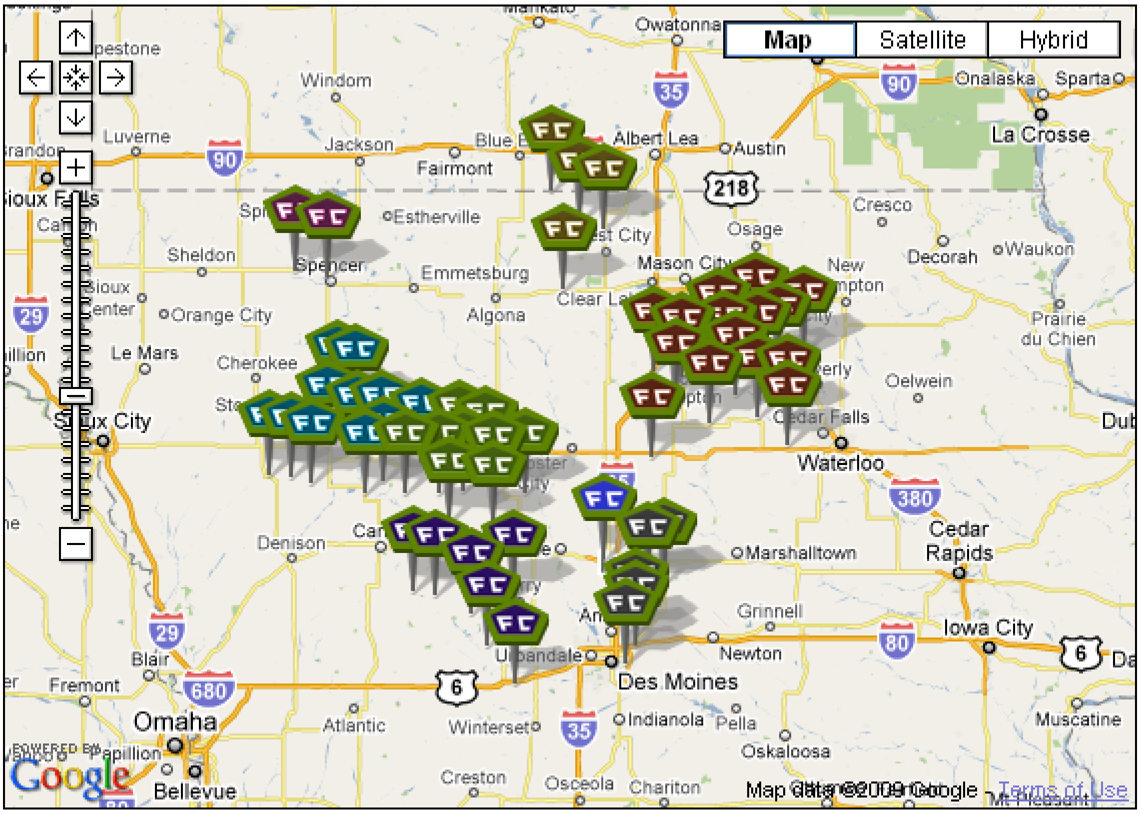


Figure 4. Optimization model inputs and outputs

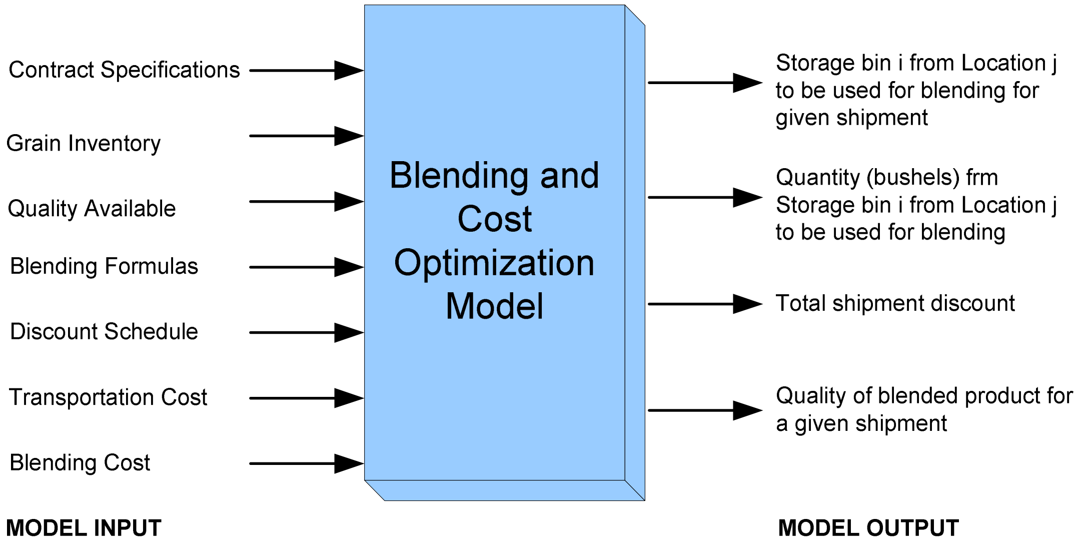


Figure 5. Pareto Optimal Front for Blending Optimization Model

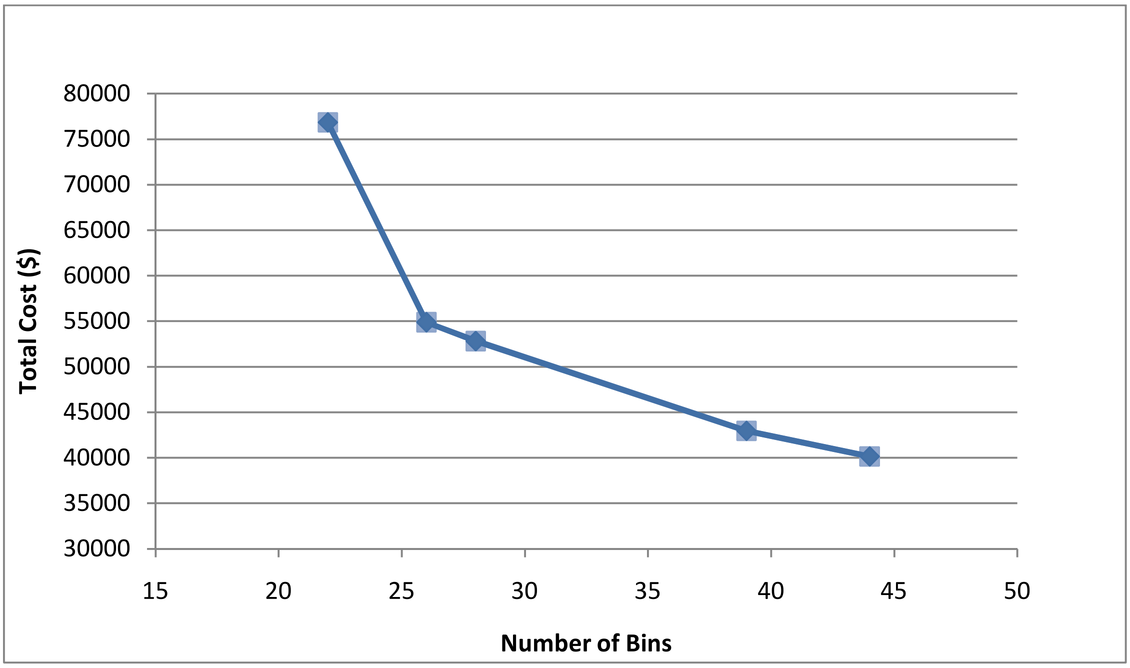


Table 1. Model notation

Notation	Description
Index sets	
J	Set of storage bins {1, 2, ……, J}
I	Set of elevator locations {1, 2, ……, I}
Input parameters	
Bij	Volume of grain in bin j at location i (Bushels)
Mij	Moisture content of grain in bin j at location i (%)
Tij	Test weight of grain in bin j at location i (lb/bu)
Dij	Damaged material content of grain in bin j at location i (%)
Fij	Foreign material content of grain in bin j at location i (%)
Cb	Contract specification for volume of grain (Bushels)
Cm	Contract specification for moisture content of grain (%)
Ct	Contract specification for test weight of grain (lb/bu)
Cd	Contract specification for damaged material content of grain (%)
Cf	Contract specification for foreign material content of grain (%)
Dm	Shipment discount for moisture ($/bu)
Dt	Shipment discount for test weight ($/bu)
Dd	Shipment discount for damaged material ($/bu)
Df	Shipment discount for foreign material ($/bu)
$Cblend$	Cost of blending grain ($/bu)
Decision variables	
Yij	Binary variable, equal to 1 if bin j at location i is used for blending grain for shipment, 0 otherwise
Xij	Volume of grain used for blending from bin j at location i
Sm	Moisture content of blended grain for shipment (%)
St	Test weight of blended grain for shipment (lb/bu)
Sd	Damaged material content of blended grain for shipment (%)
Sf	Foreign material content of blended grain for shipment (%)
Zm	Total shipment discount penalty for moisture (%)
Zt	Total shipment discount penalty for test weight (lb/bu)
Zd	Total shipment discount penalty for damaged material content (%)
Zf	Total shipment discount penalty for foreign material content (%)

Table 2. Quality factors and Discount Schedule for Corn

Quality Factor	Condition	Value	Discount ($/bu)
Moisture	≤	15%	0.02
Test Weight	≥	54 lb/bu	0.02
Damaged material	≤	5%	0.03
Foreign material	≤	3%	0.01

Table 3. Blending results from the optimization model

Objective	Bins used	Total Cost ($)	Moisture (%)	TW (lb/bu)	DM (%)	FM (%)
1	22	76,836.40	15.51	55.71	0.11	0.16
2	44	40,156.25	15.00	55.81	0.11	0.16

Table 4. Total quantity of grain transported to location A

Objective	Quantity transported (bu)	Transportation Cost ($)	Discount ($)
1	378,018	56,544.40	10,291.95
2	274,348	30,156.25	0

Table 5. Pareto optimal solutions

Pareto Optimal Solution	Bins used	Total Cost ($)	Moisture %	TW (lb/bu)	DM (%)	FM (%)
1	22	76,836.40	15.51	55.71	0.11	0.16
2	44	40,156.25	15.00	55.81	0.11	0.16
3	26	54,891.40	15.25	55.70	0.12	0.16
4	39	42,963.52	15.12	55.90	0.09	0.15
5	28	52,819.71	15.21	55.69	0.12	0.15

Table 6. Change in total cost of blending grain by changing the transportation cost

Change in transportation cost (%)	Total cost ($)	Change in total cost (%)
10	43,040.20	7.2
20	45,783.68	14.0
30	48,747.54	21.4
40	51,045.58	27.1
50	54,112.98	34.8

Table 7. Change in total cost of blending grain by changing the transportation cost

New moisture (%)	Change in moisture (%)	Change in total cost (%)	
		Objective 1	Objective 2
14.75	-1.7	6.5	12.3
14.50	-3.3	13.0	24.8
14.25	-5.0	19.5	37.2
14.00	-6.7	26.0	49.7

CHAPTER 6. Modeling traceability information using UML statecharts: Cases from pelagic fish and grain industries

Manuscript submitted to the *Journal of Food Engineering*

Maitri Thakur[1, 2,] *, Carl-Fredrik Sørensen[3], Eskil Forås[3], Charles R. Hurburgh[1, 4]

[1] Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011
[2] Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA 50011
[3] SINTEF Fisheries and Aquaculture, Brattørkaia 17C, 7010 Trondheim, Norway
[4] Department of Food Science and Human Nutrition, Iowa State University, Ames, IA 50011
* Primary author, Corresponding author

Abstract

This paper introduces a new methodology for modeling the traceability information using the UML statecharts following an event management approach in bulk food production. We follow the approach of defining states and events in food production rather than identification of traceable units. A generic model is presented and evaluated based on its practical application in bulk food production by providing illustrations from two supply chains; pelagic fish and grain. Food safety and quality issues generally occur due to incorrect processing and handling of food products. Monitoring the flow of products, their quality and the process parameters throughout production and linking them to each transition in the state of these products is an effective way of implementing and ensuring product safety and traceability. The statecharts are developed for frozen mackerel production and corn wet milling processes. All states and events for these processes as well as the information that needs to be captured for each transition are indentified that includes the product, process and quality information. The data capture points have been identified based on the various states and events that occur during food production and are connected to product, process as well as quality information.

Keywords: bulk product traceability; states and events in food production; UML statecharts; mackerel production; corn wet milling

1 Introduction

The use of electronic systems to implement traceability in food supply chains has been investigated in the recent years. The European Union law describes “Traceability” as an ability to track any food, feed, food-producing animal or substance that will be used for consumption, through all stages of production, processing and distribution (Official Journal of the European Communities, 2002). There has been an increasing interest in the use of systems such as radio frequency identification (RFID) and electronic product codes (EPC) to implement electronic traceability systems throughout the food product supply chains. The EPCglobal architecture framework is a collection of hardware, software, and data standards that can be operated by EPCglobal, its delegates and third party providers for enhancing the business flows and computer applications through the use of electronic product codes. The fundamental principle of this architecture is the assignment if a unique identity to physical objects, loads, locations, assets, and other entities whose use can be tracked (EPCglobal, 2007). Shanahan et al. (2009) proposed the use of RFID for the identification of individual cattle and biometric identifiers for verification of cattle identity. They also proposed a data structure for RFID tags and a middleware to convert animal identification data to the EPC (electronic product code) data structure. Bottani and Rizzi (2008) studied the impact of RFID technology and EPC system on the main processes of the fast moving consumer goods supply chain that composed of manufacturers, distributors and retailers. The outcomes of their study provided economical justifications for implementation of RFID and EPC in fast moving consumer goods supply chains. Myhre et al. (2009) provided a conceptual solution on how EPCIS (EPC Information Services) can be used to achieve both upstream and downstream traceability.

A food value chain consists of several actors such as farmers, producers, processors, distributors, retailers, etc. that trade goods among each other. The raw materials are transported from one actor to another where these raw materials may be processed into finished products while going through various transformations such as mixing, cooking, segregating, etc. The processed food products are then transported to distributors and retailers for sale to the customers for final consumption (Thakur and Hurburgh, 2009). In addition to the trade of goods and information between supply chain actors, several product transformations take place within an enterprise. The use of electronic systems such as EPCIS and RFID is limited to tracking product lots between actors and its use within an enterprise to record all product transformations has not been investigated. The GS1 Traceability Standard states that traceability across the supply chain involves the association of flow of information with the physical flow of traceable items. It also states that in order to achieve traceability across the supply chain, all traceability partners must achieve internal and external traceability (GS1 Global Traceability Standard, 2007). Therefore, all the actors involved in the food supply chain are required to store necessary information related to the food product that link inputs with outputs, so that when demanded, the information can be provided to the food inspection authorities on a timely basis.

One of the biggest challenges with supply chain traceability is the efficient exchange of information between various actors in the chain. The information exchanged between various actors is not complete when internal traceability systems do not exist within individual enterprises. Absence of such systems makes it impossible to connect the information related to incoming products to that of the outgoing products in any enterprise. This information needs to be captured in a precise, effective and electronic manner (FSA, 2002; Moe, 1998).

The TraceFood Framework developed under the European Commission sponsored TRACE project provides a toolbox with principles and guidelines for how to implement electronic chain traceability. The framework consists of several components: the principle of unique identifications and documentation of transformation (joining and splitting) of units being the most important requirements for implementing a traceability system (TraceFood Wiki, 2009). In order to capture and retrieve the product or process data for traceability, it is important that the data is linked to uniquely identified traceable units (TU). The TraceFood framework defines a Traceable Unit (TU) as any item upon which there is a need to retrieve predefined information and that may be priced, or ordered, or invoiced at any point in a supply chain. In practice, it refers to the smallest unit identifiable that is exchanged between two parties in the supply chain. Based on this framework, the implementation of chain traceability requires industry analysis to understand the material flow, information flow and information handling practices. Using this method, based on the industry analysis, recommendations can be provided for new sector-specific data terminology and what information needs to be recorded by each link and communicated to other links in the chain.

In this paper, we present the case of bulk food product traceability. Webster dictionary defines bulk products as “those that cannot be divided into parts or packaged in separate units”. Several food products like grain, milk, feed, pelagic fish, etc. are handled in bulk. Implementation of traceability systems in bulk product supply chains is a complex task. The two most important requirements of a traceability system are principle of unique identifications and documentation of joining and splitting of units. However, several additional challenges exist in bulk product management. For instance, bulk grain essentially has a “fluid-like” property which makes defining a fixed traceable unit (TU) practically impossible. Also, the definition of a lot or batch is not consistent throughout the supply chain. In addition, bulk product lots are often blended (mixed) and split throughout the chain. Documentation of these transformations is a challenge if the initial TUs are not well defined. Blending and splitting of individual batches complicates how information is tied to a specific entity (traceable unit) (Thakur and Hurburgh, 2009).

EPC provides a method for unique identification of all items in a supply chain. The use of EPC also makes it possible to register internal and external events electronically that are related to the movement of tagged items. EPCIS is proposed as a general, multipurpose software architecture that also has promising properties related to food traceability and thus food safety within and across enterprises (Sørensen, et al., 2010). Although, before such systems can be implemented it is crucial to identify the specific events that take place internally at an enterprise about which the product and process information needs to be recorded. In this paper, we develop supply chain models for these industries and develop a generic events diagram for bulk product processing using event management approach. This model is adapted to represent and mackerel production (packing) and corn wet milling processes. We identify the data capture points and what traceability data must be recorded at each stage. The traceability data includes product and process data as well as quality data that must be recorded whenever a transition takes place.

1.1 **Internal Traceability**

Previous research has emphasized the importance of internal traceability systems (Moe, 1998). In order to achieve a fully traceable supply chain, it is important to develop systems for chain traceability as well as internal traceability. This includes linking, to the best extent possible, units of output with specific units of input. Senneset et al. (2007) states that one of the basic prerequisites of both internal and chain traceability is the unique identification of all raw materials, semi-finished products and finished products.

They also state that there are three types of operations that are necessary for obtaining internal traceability:

(4) Recording the unique identities of traceable units. These usually refer to inputs to a process.

(5) Assigning unique identities to new traceable units. These usually refer to outputs from a process.

(6) Linking a set of input unit identities to one or more sets of output identities. These usually refer to transformation of raw materials to finished products.

It is very important to record all internal product and process data in order to link process inputs and outputs. Typical production processes to support within a company are the different transformations raw materials go through from step to step in a production into a finished product ready for shipment. The transformations may consist of many different processes where some are revocable (i.e., it is possible to go back to original state of the parts used), while others are irrevocable (i.e., it is not possible to go back to the original state).

1.2 **Bulk product traceability challenges**

Several challenges exist in implementation of traceability systems in bulk product chains. As mentioned in the previous section, the two most important requirements of a traceability system are principle of unique identifications and documentation of joining and splitting of traceable units. The concept of a traceable resource unit (TRU) was first introduced by Kim et al. (1999) where a TRU was defined as a batch of any resource. A Traceable Unit (TU) can be defined as any item upon which there is a need to retrieve predefined information and that may be priced, or ordered, or invoiced at any point in a supply chain. In practice, it refers to the smallest unit that is exchanged between two parties in the supply chain (TraceFood Wiki, 2009). Each traceable unit must be uniquely identified. In order to capture and retrieve traceability information when required, this information must be associated with a uniquely identified TU (Thakur and Donnelly, 2010). Bulk products, however, cannot be divided into parts or packaged in separate units. Several food products like grain, milk, feed, pelagic fish, etc. are handled in bulk. For instance, bulk grain essentially has a “fluid-like” property which makes defining a fixed traceable unit (TU) practically impossible. Also, the definition of a lot or batch is not consistent throughout the supply chain. In addition, bulk product lots are often blended (mixed) and split throughout the chain. Documentation of these transformations is a challenge if the initial TUs are not well defined. Blending and splitting of individual batches complicates how information is tied to a specific entity (traceable unit) (Thakur and Hurburgh, 2009). The definition of a TU would be different for each link in the bulk product chain. For example, at an elevator a truckload of grains delivered could be defined as a TU while for a processor, a TU could be a production batch.

Therefore, we present a novel technique for monitoring different states and events in bulk food production instead of defining traceable units. We present a methodology for recording the traceability data corresponding to different states and events. The traceability data consists of product data, process data and quality data. In the next section, we discuss the integration of product, process and quality data.

1.3 **Integrating product, process and quality data**

Besides the capability to track food products as they move through the supply chains, one of the most important objectives of any food traceability system is to ensure product safety and quality. Jansen-Vullers et al. (2003) suggest the following four elements for traceability: (i) physical lot integrity that includes the lot size and how well the lot integrity is maintained, (ii) data collection that includes two types of data; lot tracing data and process data, (iii) product identification and process linking: to determine product composition, and (iv) reporting to retrieve data from the system. Several product transformations and processing steps take place during industrial production of food. These transformations alter the food composition, and if not monitored properly, can affect the food quality as well as food safety.

Little research has been conducted where the information related to the food product, the processing techniques and their affect on the food quality and safety is recorded simultaneously. In order for a traceability system to meet its goal, there is need to integrate all this information into one system where a problem caused either due to processing or handling/logistics can be identified and traced back to the source. Food traceability should have an ability to indentify food safety issues linked to specific trade units and/or production batches efficiently so that necessary action can be taken in a timely manner.

Most of the research in this field presents traceability solutions where only the product packaging is tracked through the supply chains but fail to address the internal traceability issues linked to the production events within a food facility. In this paper, we present a novel solution for identification of different states and events in food production where either product or process information needs to be recorded that is essential for a traceability system to work as designed. Because we are dealing with bulk products, we follow the approach of defining states and events in food production rather than identification of traceable units.

2 Methodology

We develop a novel technique for monitoring different states and events in bulk food production and recording all product, process and quality information related to these states and events to ensure traceability. We integrate of product, process and quality data in one traceability model. We use the UML (Unified Modeling Language) statechart technique to develop the generic traceability model for bulk food production and demonstrate the application of this technique by presenting two bulk food production chains; pelagic fish and grain.

2.1 **Traceability and UML statecharts**

UML statecharts depict the various states that an object may be in and the transitions between those states. A state represents a stage in the behavior pattern of an object, and it is possible to have initial states and final states. An initial state, also called a creation state, is the one that an object is in when it is first created, whereas a final state is one in which no transitions lead out of. A transition is a progression from one state to another and will be triggered by an event that is either internal or external to the object. So, the statecharts depict the dynamic behavior of an entity based on its response to events, showing how the entity reacts to various events depending on the current state that it is in. A state is a stage in the behavior pattern of an entity. States are represented by the values of the attributes of an entity (Ambler, 2004).

A statechart is simply a network of states and events. A state is a condition during the life of an object or an interaction during which it satisfies some condition, performs some action, or waits for some event. A composite state is a state that, in contrast to a simple state, has a graphical decomposition. A composite state is decomposed into two or more concurrent substates or into mutually exclusive disjoint substates. A given state may only be refined in one of these two ways. Naturally, any substate of a composite state can also be a composite state of either type.

UML statecharts are extensively used in computer science and related fields for describing the behavior of classes, but the statecharts may also describe the behavior of other model entities such as use cases, subsystems, operations or methods. The use of statecharts in production and manufacturing systems has been limited to applications such as automated production control and planning and modeling of manufacturing systems (Köhler et al., 2000; Guojun et al., 2007; Francês et al., 2005; Vijaykumar et al., 2002). Köhler et al. (2000) present a modeling approach using UML statecharts for flexible, autonomous production agents that are used for the decentralized production systems while Guojun et al. (2007) use stochastic statecharts to describe a manufacturing system model and to obtain performance data from the system. Although, a variety of applications of statecharts exist, their application for modeling traceability events at a food production facility has not been studied.

In this paper, we present a generic model using UML statechart to represent states and events in food production where traceability information needs to be recorded. The traceability information includes product, process and quality information. We illustrate the use of the UML statechart developed by applying it to two different bulk food supply chains including pelagic fish and grain. The data capture points and the data to be recorded were identified in each chain corresponding to either an event or a state represented by the statecharts. The information to be captured includes product and process data as well as quality data that must be recorded whenever a transition takes place. The results are presented in the next section.

3 Results

3.1 **Modeling traceability events in food production**

Figure 1 shows an overview of generic states and events for general industrial production and/or processing of products. We identified 13 states and 26 generic events or transitions that may be used to provide traceability information based on data collection at specific points in the production process. The green states are typical logistics and production processes while the blue states show the use of production equipment and the gray states represent the transformation processes that take place in food production. The transformation processes may include treatments like heating, boiling, smoking, cooling, mixing, etc. The state diagram is agnostic to which kind of products that are managed. Further, the use of load carriers is not explicitly shown neither as states nor transitions, but is supposed managed by the transitions within the diagram. The same applies to other physical products that are used within the different states. Thus, the state model has emphasis on events that includes objects rather than the object themselves. Chain traceability is covered by registering events in *Product receiving/Product shipping* states, while the *Transit in/Transit out* states designate that goods are commissioned or in transit from one actor to another. As can be noted in Figure 1, only registering events related to these states, will not give a transparent view of the flow of goods between actors. In total, 12 different events are directly relevant to typical logistic processes while 14 additional events are relevant to achieve transparency related to production management and product quality and safety.

3.2 **Case studies**

In this section we present the two different bulk product supply chains and apply the statechart model presented in the previous section to these products. The states and events where traceability information needs to be recorded are identified are described for each product.

3.2.1 Pelagic fish supply chain (Mackerel)

Small pelagic fish species such as herring, mackerel, horse mackerel, etc. swim together in shoals. The fish is caught by trawling vessels in hauls and stored in one or more containers on board the fishing vessel. Pelagic fish is essentially handled as a bulk product until it arrives at the production facility. Figure 2 shows the mackerel supply chain from catch to consumption. In this case, we investigated the mackerel supply chain from Norway to Japan. The fish is caught by trawling vessels in hauls and stored in one or more containers on board the fishing vessel. The haul is a Traceable Unit (TU) that is recorded in the official log. Each haul is stored in one or multiple tanks onboard the vessel. When the trip ends, the vessel reports the catch as one or multiple TUs to NSS. This TU will be used through auction and sales. NSS enters catch data into auction and the sales report is sent to the buyer. At landing (at the production/ packing facility), fish is weighed and quality is verified. If disparity in quality is detected, the original TU may be separated into several new TUs. Each TU is identified with a unique ID. After packing the fish, the boxes are stacked on pallets are stored in freezers. The product can be in storage from two to three days and up to six months before it is shipped to the customer. Outgoing packed TU are pallets. The bill of lading is sent from the producer to transporters and Japanese importers through the Norwegian exporter. About 60% of the exported fish goes directly to the Japanese importer which is further sold to the mackerel processor. The remaining 40% arrives at the Chinese processor to be processed into the end product and then sent to Japan where it is sold by the importers to the Japanese customers.

3.2.1.1 Frozen mackerel production process

We focused on the frozen mackerel production process and developed the UML statecharts for three links in the supply chain: the fishing vessel, frozen mackerel producer and the shipper. The flow diagram for the mackerel production process is shown in Figure 3.

The frozen mackerel production can be described as following:

1. The fishing vessel is received at the production facility and the fish is pumped into the production plant.
2. The quantity of fish received from a vessel is determined by the flow rate during pumping.
3. When fish enters the production plant, it is graded and divided based on weight (size) using automatic graders. Manual checks are also performed to ensure the accuracy of graders and provide a visual quality control.
4. After grading, fish is packed in 20 kg boxes and labeled. The label identifies several product and process parameters described in the later sections.
5. After packing the fish, the boxes are stacked and refrigerated in freezing tunnels.
6. After refrigeration, the boxes are stored in cold storage. When in storage, the temperature measurements of the product are taken at fixed intervals. The boxes closer to the walls of the storage unit are retrieved for temperature measurements. The optimum temperature for storage of mackerel is -18° Celsius.
7. The boxes are palleted for shipment and stored in containers (temperature controlled) before shipping to the customers. The product can be in storage from two to three days and up to six months before it is shipped.

It was noted that a shipping container can carry one or more orders from one or several production batches. A production batch refers to one day of production.

3.2.1.2 UML statechart modeling

Based on the analysis of the production process, we developed the UML statechart for the frozen mackerel production process, the fishing vessel and shipper entities. Figure 4 represents the states and events for the frozen mackerel production process. Seventeen states consisting of three composite states and twenty-nine events were identified in the production process. The product, process and quality data collected during production can be linked to one of these states or events and can be used to provide traceability information. The different states and events are described in Table 1 and Table 2 respectively.

Three composite states were identified in the process. *Sorting* of fish as it enters the production plant comprises of three sub-states: *Weight control*, *Distribution to belt* and *Manual check*. As the fish is pumped into the production plant, it is sorted into three grades (A, B, C) based on the weight before transferring to the conveyor belts. After sorting, fish of each grade is handled separately and never mixed again during the entire production process. The sorted fish on conveyor belts is weighed manually as a quality control check. The second composite state *Packing* represents three concurrent states for packing of graded (sorted) fish separately. Similarly, the third composite state *Palleting* represents the three concurrent states for palleting of boxes of graded (sorted) fish separately. It must be noted that production of frozen mackerel is a continuous process and each state ends when there is no product available in the system. In addition, one day of production is considered as one product batch. Figures 5 and 6 represent the states and events for the fishing vessel and shipper entities. The various states and events for these entities are described in Tables 3 to 6.

3.2.2 Bulk grain supply chain (Corn)

Corn is the most widely produced feed grain in the United States, accounting for more than 90 percent of the total value and production of feed grains. Corn is processed into several food and industrial products including starch, sweeteners, corn oil, beverage and industrial alcohol and fuel ethanol. The United States is a major player in the world corn trade market, with approximately 20 percent of the corn crop exported to other countries (Economic Research Service, 2009).

Corn is handled as a bulk commodity as it moves from the farmer to the consumer. Three soybean chain stakeholders are presented in this paper; farmer, elevator and processor. Figure 7 shows a simple flowchart of the corn value chain. The farmer is the first link in the corn value chain. Farmers purchase seeds from a seed company and sell their crop to an elevator after harvesting. Several chemical compounds including fungicides and herbicides are used for soybean seed treatment to inhibit damage to the crop. Combines are commonly used for harvesting the corn crop. After harvest, corn can be stored on farm before selling to an elevator. An elevator is a very important link between the farmer and the processor. Elevators buy corn from the farmers, keep it in storage, and blend it before selling to the processors. Corn crops received at the elevator are sampled and graded based on moisture content, test weight, foreign material and damaged material. The farmers are paid according to the quality grade. The grain is then conveyed to the storage silos before shipping to the customers. One storage silo can contain grain from several farmers. The incoming lots from the farmers are blended before shipment in order to meet the buyer’s quality specifications. Thus, a specific lot shipped to the processor can contain grain from all different sources that may end up in the finished product. In this paper, we present the corn wet milling process and develop the UML statechart for defining the states and events for recoding traceability information.

3.2.2.1 Corn wet milling process

The corn wet milling is a process for separating corn into its component parts using a water sulphur dioxide system. The products of the corn wet milling process are: (1) Starch: used as starch or converted to syrup such as glucose, dextrose or high fructose corn syrup which can be further used in production of ethanol by fermentation, (2) Germ: pressed to remove corn oil and the fibrous residue is used as cattle feed, (3) Gluten: used for poultry feed enrichment, and (4) Fiber and steep water solids: used as livestock feed.

The corn wet milling process can be described as following (Corn wet milled feed products, 2006):

1. The processor receives corn from the elevator usually delivered by truck, barge or railcar.
2. The grain is cleaned and stored in large storage silos. The cleaned corn is transported to large tanks called steep where warm water (at about 130° F)

containing dissolved sulphur dioxide is circulated for approximately 40 hours to soften the corn kernels.

3. Next, the softened corn kernels pass through attrition mills that break them up, loosen the hull and free the germ from the endosperm. Centrifugal force is used to isolate the germ.
4. The clean germ is dried and crude corn oil is removed either by mechanical press or solvent extraction method. The extracted germ meal is used in animal feed.
5. The remaining mixture of hull and endosperm then passes through a series of grinding and screening operations. The hull particles are removed on screens, while the finer particles of protein and starch pass through. The hull is used as a constituent in animal feed or for production of refined corn fiber for food use.
6. The water slurry of starch and gluten is separated in centrifuges. The gluten is dried and sold as gluten meal or used as an ingredient in corn gluten feed.
7. The starch slurry is washed to remove small quantities of solubles. The starch slurry may be used to make sweeteners or further processed to make corn starch.

All constituents obtained from the corn wet milling process are used for further processing into several components that can be used for food, feed and fuel purposes.

3.2.2.2 UML statechart modeling

Based on the analysis of the production process, we developed the UML statechart for corn wet milling process, the elevator and the farmer entities. Figure 8 represents the states and events for the corn wet milling process. Thirty-one states thirty-three events were identified in the production process. The product, process and quality data collected during production can be linked to one of these states or events and can be used to provide traceability information. The different states and events are described in Table 7 and Table 8 respectively. It must be noted that corn wet milling is a continuous process that produces several products and each state ends when there is no product available in the system. Figures 9 and 10 represent the states and events for the farming and elevator operations. The various states and events for these entities are described in Tables 9 to 12.

3.3 **Discussion of results**

Detailed descriptions of the states and events for each entity in the two supply chains are provided. These descriptions include the start and end point of each state, the corresponding objects and the quality control parameters. The objects corresponding to each state are identified and these objects can either be an actor, a resource or a traceable item. The kind of object/s related to a given state allow in determining the information that needs to be recorded for a particular state. Similarly, the quality control parameters are identified for each state and can be linked to either the resource or the traceable item or both. In addition to the production states, events in food production for the two chosen products are also described. An event takes place when a traceability object transitions from one state to the next. It is important to link each event to the corresponding states. Identifying the events in food production helps in determining the transformations that occur so that appropriate information can be stored corresponding to these transitions. It must be noted that the product, process and quality information is integrated in this model and corresponds to a given state or event in food production. Technologies such as EPCIS can be used for implementing food traceability systems within and across enterprises once the specific events that take place during food production are identified.

4 Conclusions

In this paper, we have introduced a methodology for using the UML statecharts to model the states and events in bulk food production where traceability information needs to be recorded. Because we are dealing with bulk products, we follow the approach of defining states and events in food production rather than identification of traceable units. We presented a generic model and its practical application was demonstrated by adapting it for two different bulk food supply chains; pelagic fish and grain. Several challenges exist in implementation of traceability systems in bulk product supply chains including definition of traceable units and documentation of product transformations. Bulk products replicate the fluid-like properties and normally undergo a continuous production process which makes it impossible to define a fixed lot-size of traceable unit. To overcome this problem, we introduce the modeling technique to identify all the states and events that occur in food production and processing to cover internal traceability.

The statecharts are developed for frozen mackerel production process including the fishing vessel, producer and shipper entities and for corn wet milling process including the farmer, elevator and corn wet miller entities. All states and events for these processes as well as the information that needs to be captured for each transition are indentified. In order for any traceability system to meet one of its most important requirements of ensuring food quality and safety, there is need to integrate all this information into one system so that a problem caused either due to processing or handling/logistics can be identified and traced back to the source. Therefore, we integrate the product, process and quality information into the data that is recorded when transition takes place from one state to another.

Food safety and quality issues generally occur due to incorrect processing and handling of food products. Bulk food production also has other challenges including product transformations such as blending or splitting of batches. Monitoring the flow of products, their quality and the process parameters throughout production and linking them to each transition in state of the products is an effective way of implementing and ensuring product safety and traceability.

The model presented in this paper has been evaluated based on its practical application in bulk food production by providing illustrations from two supply chains; pelagic fish and grain. The data capture points have been identified based on the various states and events that occur during food production and are connected to product, process as well as quality information.

References

Ambler, S.W., 2004. The Object Primer: Agile Model Driven Development with UML 2, 3rd Edition, Cambridge University Press, New York.

Bottani, E., Rizzi, A., 2008. Economical assessment of the impact of RFID technology and EPC system on the system on the fast-moving consumer goods supply chain. International Journal of Production Economics, 112(2), 548-569.

Carriquiry, M., Babcock, B.A., 2007. Reputations, market structure and the choice of quality assurance systems in the food industry. American Journal of Agricultural Economics 89, 12–23.

Corn Refiners Association, 2006. Corn wet milled feed products, 4th Edition, Corn Refiners Association, Washington D.C.

Economic Research Service, 2009. <http://www.ers.usda.gov/Briefing/Corn>

EPCglobal, 2007. The EPCglobal Architecture Framework. EPCglobal Final Version 1.2 <http://www.epcglobalinc.org>

Francês, C.R.L., Oliveira, E.D.L., Costa, J.C.W.A., Santana, M.J., Santana, R.H.C., Bruschi, S.M., Vijaykumar, N.L., Carvalho, S.V.D., 2005. Performance evaluation based on system modeling using statecharts extensions. Simulation Modeling Practice and Theory, 13(3), 584-618.

FSA, 2002. Traceability in the Food Chain – A Preliminary Study.

Golan, E., Krissoff, B., Kuchler, F., 2004. Food traceability: one ingredient in a safe and efficient food supply, economic research service. Amber Waves 2, 14–21.

GS1 Global Traceability Standard, 2007. Business Process and System Requirements for Full Chain Traceability. <http://www.gs1.org/traceability/gts>

Guojon, Z., Jiabing, H., Haiping, Z., Xuan, C., 2007. Manufacturing system modeling and performance evaluation based on improved stochastic statechart. Frontiers of Mechanical Engineering in China, 2(4), 453-458.

International Organization for Standardization, 2007. New ISO Standard to Facilitate Traceability in Food Supply Chains. ISO 22005:2007.

Jansen-Vullers, M.H., van Dorp, C.A., Buelens, A.J.M., 2003. Managing traceability information in manufacture. International Journal of Information Management 23, 395–413.

Kim, H.M., Fox, M.S., Grüninger, M., 1999. An ontology for quality management enabling quality problem identification and tracing. BT Technology Journal, 17(4), 131-140.

Köhler, H.J., Nickel, U., Niere, J., Zündorf, A., 2000. Integrating UML diagrams for production control systems, in Proceedings of 22nd International Conference on Software Engineering (ICSE 2000), Limerick, Ireland, 241-251.

Madec, F., Geers, R., Vesseur, P., Kjeldsen, N., Blaha T., 2001. Traceability in the pig production chain. Revue Scientifique Et Technique (International Office of Epizootics) 20 (2), 523–537.

McKean, J.D., 2001. The importance of traceability for public health and consumer protection. Revue Scientifique Et Technique (International Office of Epizootics) 20 (2), 363–371.

Moe, T.,1998. Perspectives on traceability in food manufacture. Trends in Food Science & Technology 9(5), 211–214

Myhre, B., Netland, T.H., Vevle, G., 2009. The footprint of food - A suggested traceability solution based on EPCIS. In the 5th European Workshop on RFID Systems and Technologies (RFID SysTech 2009), Bremen, Germany.

Official Journal of the European Communities, 2002. Regulation (EC) No. 178/2002 of the European Parliament and the Council of 28 January 2002.

Senneset, G., Forås, E., Fremme, K.M., 2007. Challenges regarding implementation of electronic chain traceability. British Food Journal, 109(10), 805-818.

Shanahan, C., Kernan, B., Ayalew, G., McDonnell, K., Butler, F., Ward, S., 2009. A framework for beef traceability from farm to slaughter using global standards: An Irish perspective. Computers and Electronics in Agriculture, 66(1), 62-69.

Sørensen, C., Sødal, A., Reinertsen, B.J., Sivertsen, R., Vevle, G., Huseby, P., Braathe, E., Sørensen, T.E., Johanson, E., 2009. Architectural and system design of an electronic traceability system based on EPCIS. SINTEF Report, SINTEF Fisheries and Aquaculture.

Sørensen, C., Vevle, G., Gunnlaugsson, V.N., Bjørnson, F.O., Forås, E., Margeirsson, S., Thakur, M., 2010. EPCIS as an infrastructure for electronic traceability, manuscript under preparation.

Thakur, M., Hurburgh, C.R., 2009. Framework for implementing traceability system in the bulk grain supply chain. Journal of Food Engineering, 95(4), 617-626.

Thakur, M., Donnelly, K.A.-M., 2010. Modeling traceability information in soybean value chains. Journal of Food Engineering, doi:10.1016/j.jfoodeng.2010.02.004.

TraceFood Wiki, 2009. <http://www.tracefood.org>

US Food and Drug Administration, 2002. The Bioterrorism Act of 2002.

Vijaykumar, N.L., Carvalho, S.V.D., Abdurahiman, V., 2002. On proposing statecharts to specify performance models. International Transactions in Operations Research, 9(3), 312-336.

Figure 1. Generic events in food production and processing

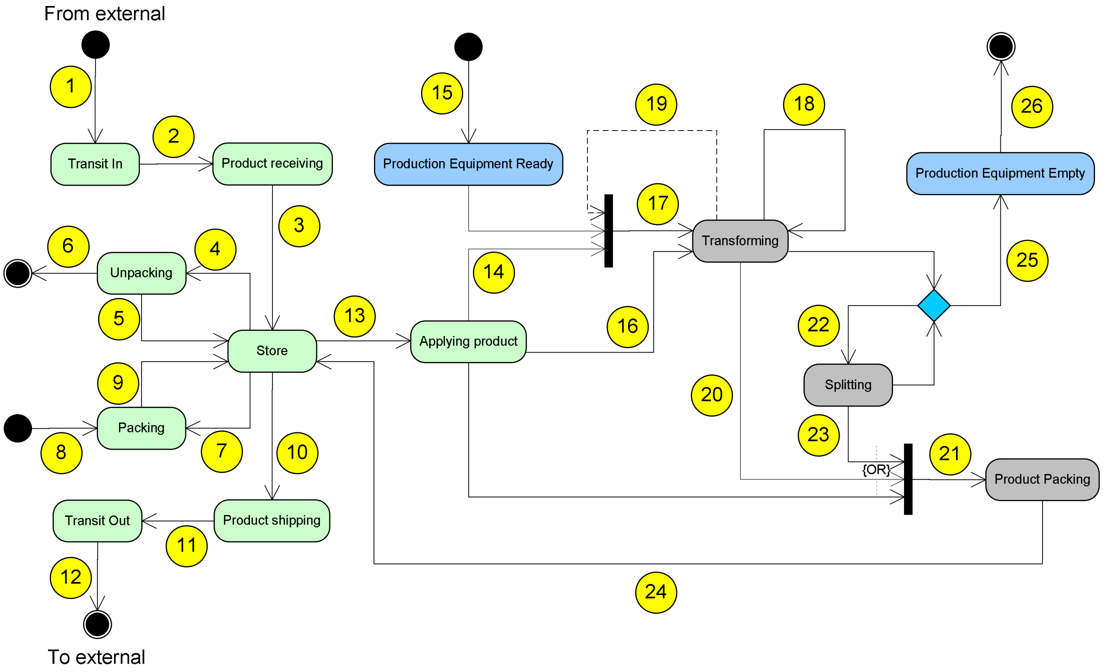


Figure 2. Flow of goods and information in the mackerel supply chain from Norway to Japan

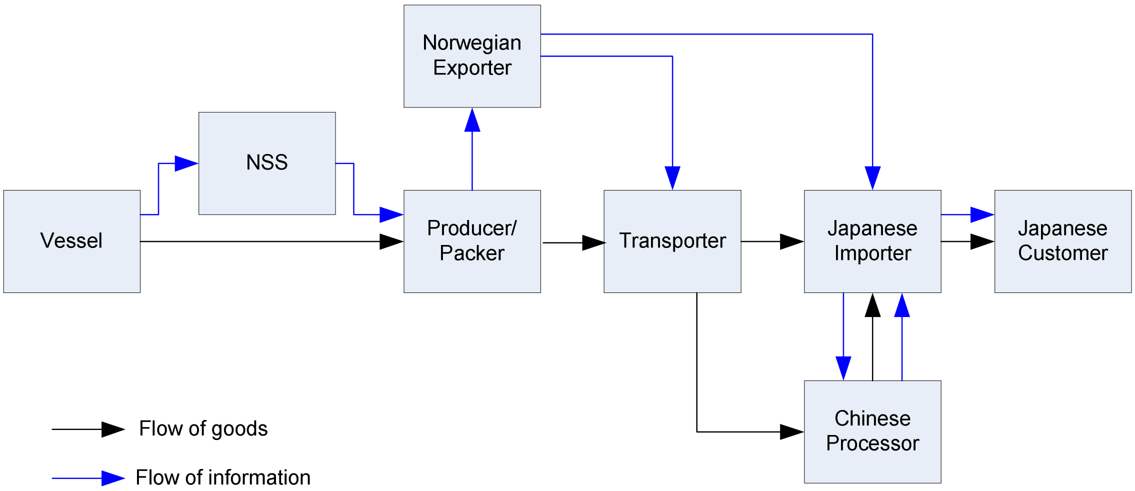


Figure 3. Flow diagram for mackerel production process

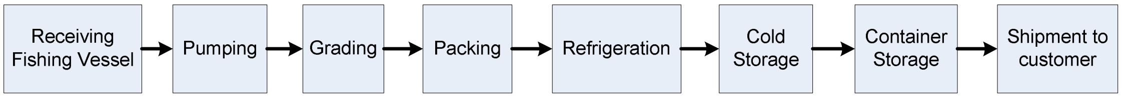


Figure 4. States and events in frozen mackerel production process

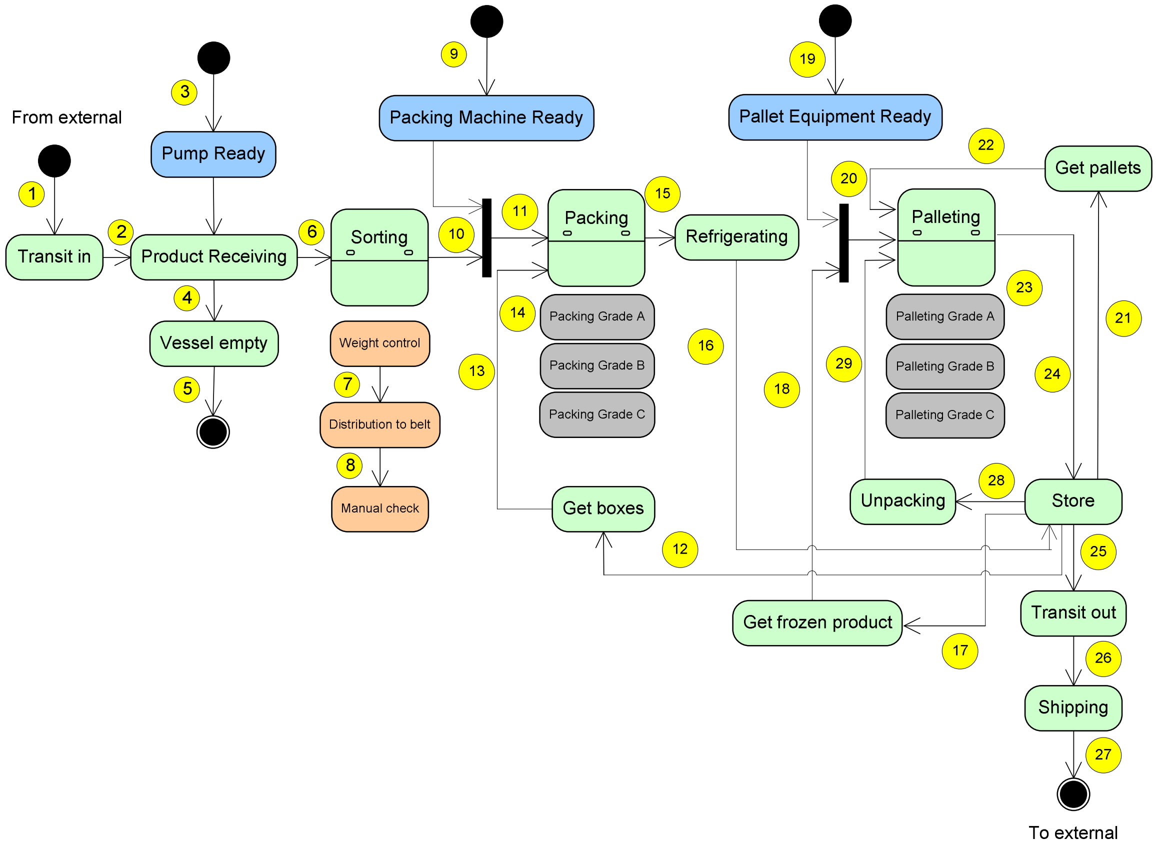


Figure 5. States and events for fishing vessel entity

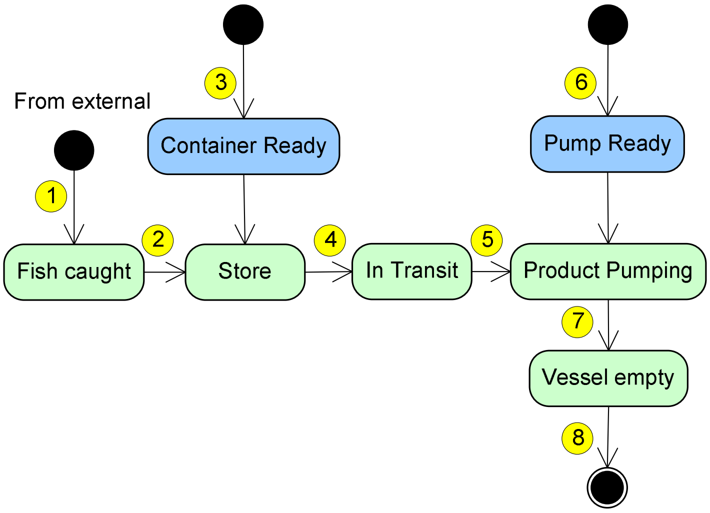


Figure 6. States and events for shipper entity

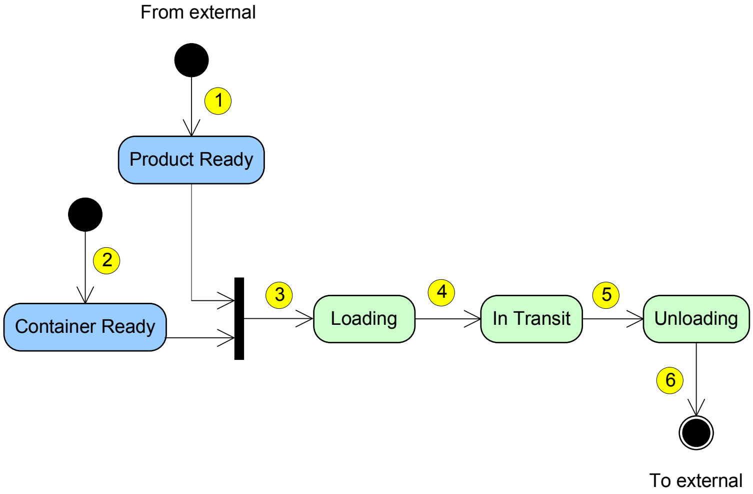


Figure 7. Flow of goods and information in the corn supply chain

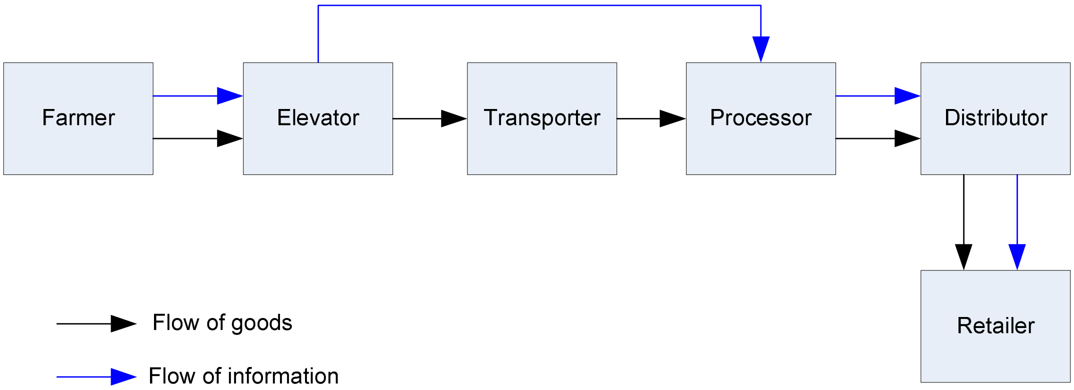


Figure 8. States and events in corn wet milling process

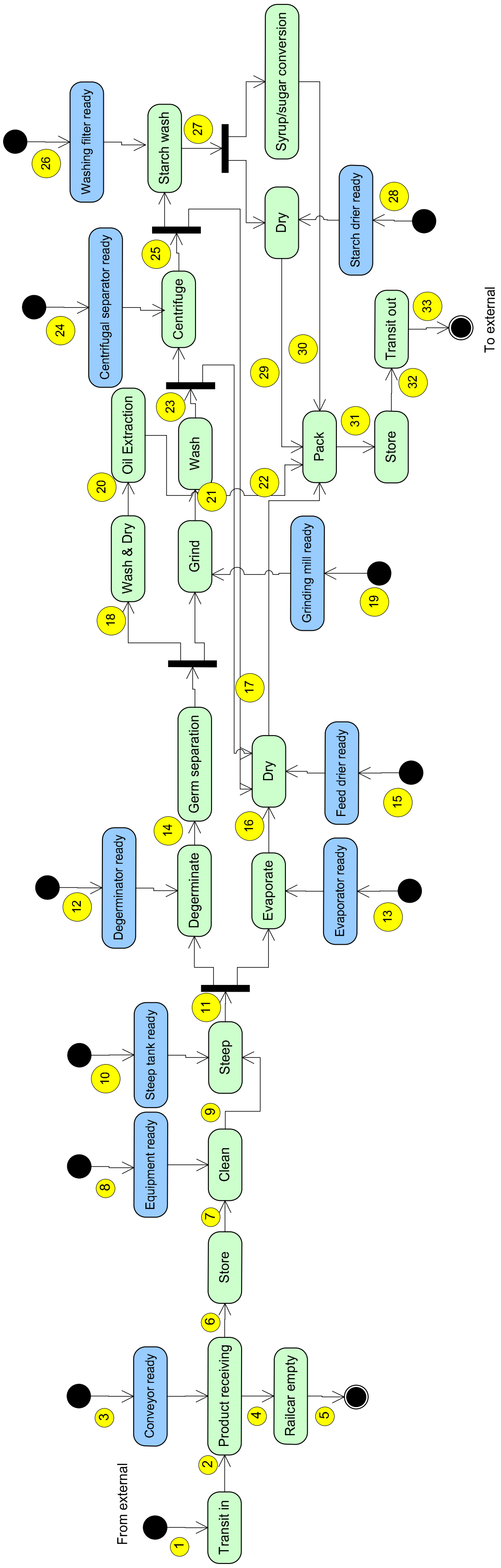


Figure 9. States and events in corn farming operation

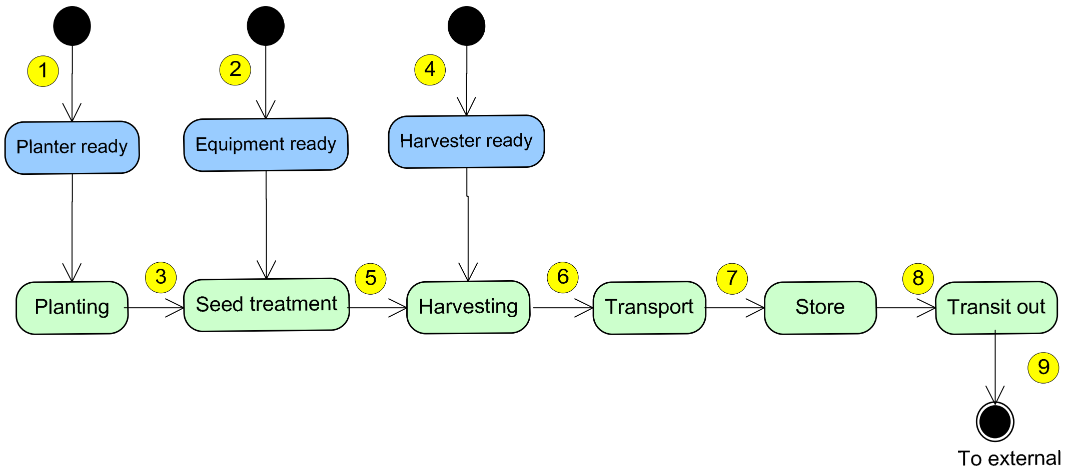


Figure 10. States and events in elevator operation

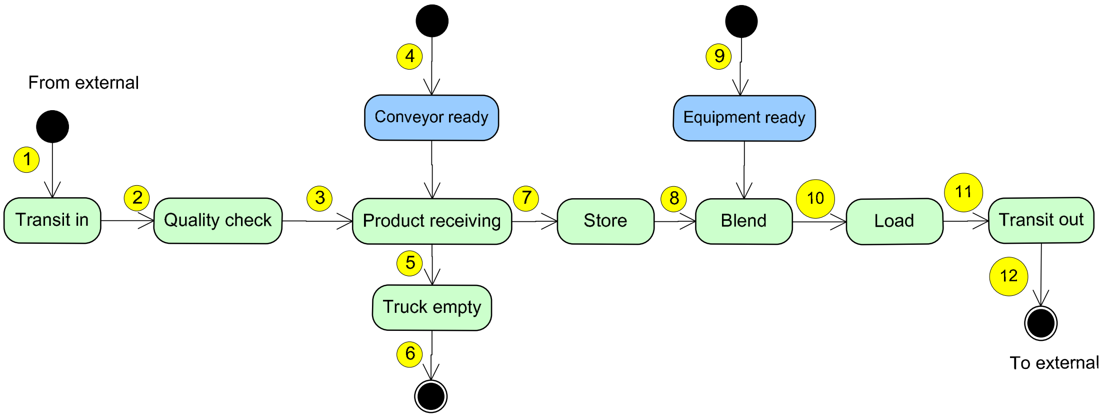


Table 1. Description of states in the frozen mackerel production

State	**Description**	**Start**	**End**	**Objects**	**Quality control**
Transit in	Denotes that fishing vessel is received at the production plant	Fishing vessel to be received	Fishing vessel received at production plant	Actor, Resource, Traceable Item	NA
Pump ready	Denotes that the pump is ready (clean) to be used for product receiving	Pump cleaned	Pump ready for use	Resource	Pump sterilized
Product receiving	Denotes that the fish is received by pumping into the production plant	Fish ready to be pumped into the production plant	Fish ready to be sorted	Resource, Traceable Item	Flow rate
Vessel empty	Denotes that the fishing vessel is emptied after pumping	Fish being pumped out	Fishing vessel empty	Actor, Resource, Traceable Item	NA
Sorting	This is a composite state comprised of three sub states: Weight control, Distribution to belt, and Manual check	Fish ready to be sorted after pumping	Fish sorted into different grades based on weight and ready to be packed	Resource, Traceable Item	Weight Visual inspection
Weight control	Denotes that fish is sorted using weight control technique	Fish ready to be sorted after pumping	Fish sorted based on weight	Resource, Traceable Item	NA
Distribution to belt	Denotes that fish is transferred to the conveyor belt after sorting	Fish ready to be distributed on conveyor belt after sorting	Fish distributed on conveyor belt	Resource, Traceable Item	Visual inspection
Manual check	Denotes that manual check is performed by taking random fish from the conveyor belt	Fish ready to be weighed manually	Fish checked manually and sorted into different grades based on weight	Resource, Traceable Item	Weight
Packing machine ready	Denotes that packing machine is ready to enter the packing state	Packing machine ordered	Packing machine ready for use	Resource	Packing machine sterilized
Store	Denotes the process of managing stock	Goods ready for storage	Goods stored	Resource, Traceable Item	Temperature (for fish storage)
Get boxes	Denoted the process of getting boxes from storage for packing	Boxes ready in storage	Boxes ready for use in packing	Resource, Traceable Item	NA
Packing	This is a composite state and denotes the packing process of fish using the packing material and graded fish. The state consists of 3 concurrent states:	Fish and packing material ready to be used	Fish with different packed into boxes	Resource, Traceable Item	NA
Packing Grade A	Denotes the process of packing of grade A fish	Grade A fish and packing material ready to be used	Grade A fish packed into boxes	Resource, Traceable Item	NA

Table 1. (continued)

State	**Description**	**Start**	**End**	**Objects**	**Quality control**
Packing Grade B	Denotes the process of packing of grade B fish	Grade B fish and packing material ready to be used	Grade B fish packed into boxes	Resource, Traceable Item	NA
Packing Grade C	Denotes the process of packing of grade C fish	Grade C fish and packing material ready to be used	Grade C fish packed into boxes	Resource, Traceable Item	NA
Refrigerating	Denotes that the packed boxes are refrigerated in tunnel freezers	Packed boxes ready to be refrigerated	Packed boxes refrigerated	Traceable Item	Temperature
Get frozen product	Denotes the process if getting the frozen product from cold storage	Frozen product ready in cold storage	Frozen product ready to be palleted	Traceable Item	NA
Pallet equipment ready	Denotes that pallet equipment is ready to enter the palleting state	Pallet equipment ordered	Pallet equipment ready for use	Resource	Pallet equipment clean
Get pallets	Denoted the process of getting pallets from storage for palleting	Pallets ready in storage	Pallets ready for use in palleting	Resource, Traceable Item	NA
Palleting	This is a composite state and denotes the palleting process of boxes containing frozen fish of different grades. The state consists of three concurrent states as follows:	Packed fish and palleting material ready to be used	Pallets of packed fish created	Resource, Traceable Item	NA
Palleting Grade A	Denotes the process of making pallets of boxes containing grade A fish	Grade A packed fish and palleting material ready to be used	Pallets of Grade A packed fish created	Resource, Traceable Item	NA
Palleting Grade B	Denotes the process of making pallets of boxes containing grade B fish	Grade B packed fish and palleting material ready to be used	Pallets of Grade B packed fish created	Resource, Traceable Item	NA
Palleting Grade C	Denotes the process of making pallets of boxes containing grade C fish	Grade C packed fish and palleting material ready to be used	Pallets of Grade C packed fish created	Resource, Traceable Item	NA
Unpacking	Denotes the process of splitting of pallets by unpacking and removing some boxes	Pallets in storage ready for unpacking	Pallets in storage unpacked	Resource, Traceable Item	NA
Transit out	Denotes the process of physical shipping of goods out from the production plant	Pallets ready for shipping	Pallets shipped	Resource, Traceable Item, Actor	NA
Shipping	Denotes the process of getting the product ready for shipment	Pallets picked from storage	Pallets ready for shipping	Resource, Traceable Item, Actor	NA

Table 2. Description of events in the frozen mackerel production

No.	Transition	From state	To state	Description
1	Fishing vessel to be received	Start state Another actor	Transit in	This transition denotes that the fishing vessel is in transit to the production plant
2	Fish to be pumped	Transit in	Product receiving	This transition denotes that the handover of fish from vessel to production plant
3	Pump made ready for use	Start state	Pump ready	This transition denotes that the pump is made ready for use in product receiving
4	Vessel to be emptied	Product receiving	Vessel empty	This transition denotes that the pumping of fish from vessel into the production plant
5	Vessel to exit	Vessel empty	End state	This transition denotes that the empty vessel left the production plant
6	Fish to be sorted	Product receiving	Weight control	This transition denotes the sorting of received fish based on weight control
7	Fish to be distributed on conveyor belt	Weight control	Distribution to belt	This transition denotes that the sorted fish is distributed to the conveyor belt
8	Fish to be checked manually	Distribution to belt	Manual check	This transition denotes that the fish on conveyor belt is checked (weighed) manually
9	Packing machine made ready for use	Start state	Packing machine ready	This transition denotes that the packing machine is made ready for use in production
10	Sorted fish to be packed	Manual check	Packing	This transition denotes that sorted fish is ready for packing
11	Packing machine used in packing process	Packing material ready Manual check	Packing	This transition denotes that the packing material is used to pack the sorted fish
12	Boxes to be taken from storage	Store	Get boxes	This transition denotes that the boxes are taken from storage to be used for packing
13	Boxes used in packing process	Get boxes	Packing	This transition denotes that the boxes are used to pack the sorted fish
14	Concurrent events for packing material used in packing of different grades of fish	Packing material ready Manual check	Palleting	This transition denotes that the packing material is used to pack the sorted fish based on grade
15	Packed fish ready to be refrigerated	Packing	Refrigerating	This transition denotes that the packed fish is refrigerated in tunnel freezers
16	Frozen fish ready to be stored in cold storage	Refrigerating	Store	This transition denotes that the frozen fish is stored in cold storage
17	Frozen fish to be taken from cold storage	Store	Get frozen product	This transition denotes that the boxes containing frozen product are taken from cold storage for palleting
18	Frozen product to be palleted	Get frozen product	Palleting	This transition denotes that the frozen product is ready to be palleted
19	Pallet equipment made ready for use	Start state	Pallet equipment ready	This transition denotes that the pallet equipment is made ready for use in production
20	Pallet equipment used in palleting process	Pallet equipment ready Get frozen product	Palleting	This transition denotes that the pallet equipment is used to make pallets of boxes containing frozen fish

Table 2. (continued)

No.	Transition	From state	To state	Description
21	Pallets to be taken from storage	Store	Palleting	This transition denotes that the pallets are taken from storage to be used for palleting
22	Pallets used in palleting process	Get pallets	Palleting	This transition denotes that the pallets are used for palleting the packed boxes
23	Concurrent events for pallet equipment used for palleting of packed graded fish	Pallet equipment ready Packing Unpacking	Store	This transition denotes that the pallet equipment is used to make pallets of packed fish based on grade
24	Pallets to be stored	Palleting	Store	This transition denotes that the pallets are ready to be stored
25	Pallets to be delivered	Store	Transit out	This transition denotes that the stored pallets are taken for storage for shipping
26	Pallets to be shipped	Transit out	Shipping	This transition denotes that pallets are ready to be shipped
27	Pallets shipped	Shipping	End state Another actor	This transition denotes that the pallets are shipped and outside the control of the production plant
28	Pallets to be unpacked	Store	Unpacking	This transition denotes that pallets in storage are unpacked
29	Boxes to be palleted	Unpacking	Palleting	This transition denotes that unpacked boxes are palleted

Table 3. Description of states for fishing vessel entity

State	Description	Start	End	Objects	Quality control
Fish caught	Denotes the process of catching fish	Fishing vessel ready	Fish caught	Resource, Traceable Item	NA
Store	Denotes the process of storing fish on the vessel	Fish ready for storage	Fish stored	Resource, Traceable Item	Temperature
Container ready	Denotes that the container is ready (clean) to be used for storage	Container cleaned	Container ready for use	Resource	Container sterilized
In transit	Denotes that fishing vessel is in transit to the production plant	Fishing vessel in transit	Fishing vessel received at production plant	Actor, Resource, Traceable Item	NA
Pump ready	Denotes that the pump is ready (clean) to be used for product receiving	Pump cleaned	Pump ready for use	Resource	Pump sterilized
Product pumping	Denotes that the fish is pumped into the production plant	Fish ready to be pumped into the production plant	Fish pumped into the production plant	Resource, Traceable Item	Flow rate
Vessel empty	Denotes that the fishing vessel is emptied after pumping	Fish being pumped out	Fishing vessel empty	Actor, Resource, Traceable Item	NA

Table 4. Description of events for fishing vessel entity

No.	Transition	From state	To state	Description
1	Fishing vessel to be caught	Start state	Fish caught	This transition denotes that the fishing vessel is ready to catch fish
2	Fish to be stored	Fish caught	Store	This transition denotes that the fish is ready to be stored on the vessel
3	Container made ready for use	Start state	Container ready	This transition denotes that the container is made ready to store fish
4	Vessel to start transit	Store	In transit	This transition denotes that the vessel starts the transit towards the production plant
5	Fish to be pumped into production plant	In transit	Product pumping	This transition denotes that the fish is ready to be pumped into the production plant
6	Pump made ready for use	Start state	Pump ready	This transition denotes that the pump is made ready for use in product pumping
7	Vessel to be emptied	Product pumping	Vessel empty	This transition denotes that the pumping of fish from vessel into the production plant
8	Vessel to exit	Vessel empty	End state	This transition denotes that the empty vessel left the production plant

Table 5. Description of states for shipper entity

State	Description	Start	End	Objects	Quality control
Product ready	Denotes that pallets of packed fish are ready to be shipped	Packed fish in storage	Packed fish ready	Resource, Traceable Item	NA
Container ready	Denotes that the container is ready (clean) to be used for shipping	Container cleaned	Container ready for use	Resource	Container sterilized
Loading	Denotes the process of loading the shipping contained with pallets of packed fish product	Packed fish and container ready	Packed fish loaded into container	Resource, Traceable Item	Weight
In transit	Denotes that container is in transit to the customer	Container in transit	Container received by the customer	Actor, Resource, Traceable Item	Temperature
Unloading	Denotes the process of unloading the product from shipping container	Container arrives at customer	Container unloaded	Actor, Resource, Traceable Item	NA

Table 6. Description of events for shipper entity

No.	Transition	From state	To state	Description
1	Product made ready	Start state	Product ready	This transition denotes that the packed fish is ready to be loaded for shipping
2	Container made ready	Start state	Container ready	This transition denotes that the container is ready to be loaded for shipping
3	Product ready for loading in container	Product ready Container ready	Loading	This transition denotes that the container is loaded with packed fish product
4	Shipping container to start transit	Loading	In transit	This transition denotes that the shipping container starts the transit towards the customer
5	Shipping container to be unloaded	In transit	Unloading	This transition denotes that the packed fish product is ready to be unloaded from the container
6	Shipping container unloaded	Unloading	End state	This transition denotes that the container is unloaded and product delivered to the customer

Table 7. Description of states in the corn wet milling process

State	Description	Start	End	Objects	Quality control
Transit in	Denotes that grain container is received at the corn wet milling plant	Grain container to be received	Grain container received at production plant	Actor, Resource, Traceable Item	NA
Conveyor ready	Denotes that the conveyor is ready (clean) to be used for product receiving	Conveyor cleaned	Conveyor ready for use	Resource	Conveyor cleaned
Product receiving	Denotes that the grain is received by conveying into the storage bins	Grain ready to be conveyed to the storage bins	Grain transferred to the storage bins	Resource, Traceable Item	Product quality
Railcar empty	Denotes that the railcar is emptied after receiving grain	Grain being transferred	Railcar empty	Actor, Resource, Traceable Item	NA
Store	Denotes that the grain is stored in the storage bins at the production plant	Grain ready to be stored after conveying	Grain stored until ready to be used in wet milling	Resource, Traceable Item	Product moisture and temperature
Equipment ready	Denotes that the equipment for cleaning grain (screens) is ready	Cleaning equipment available	Cleaning equipment ready for use	Resource	Equipment cleaned
Clean	Denotes that grain is cleaned	Grain ready to be cleaned	Grain cleaned	Resource, Traceable Item	Visual inspection
Steep tank ready	Denotes that the steep tank is ready to begin the steeping process	Steep tank available	Steep tank ready for use	Resource	Steep tank cleaned
Steep	Denotes that the cleaned grain is steeped in steep tanks	Clean grain ready for steeping	Corn ready for degermination and evaporation processes	Resource, Traceable Item	Water temperature, SO_2 concentration
Degerminator ready	Denotes that the degerminator is ready to begin the degermination of corn	Degerminator available	Degerminator ready for use	Resource	Degerminator cleaned

Table 7. (continued)

State	Description	Start	End	Objects	Quality control
Degerminate	Denotes the process of degermination where endosperm is separated from the corn kernels	Corn ready for degermination process after steeping	Corn ready for germ separation	Resource, Traceable Item	Mill clearance
Evaporator ready	Denotes that the evaporator is ready to concentrate the steeping water	Evaporator available	Evaporator ready for use	Resource, Traceable Item	Evaporator cleaned
Evaporate	Denotes the process of evaporating steep water	Steep water is ready for evaporation after steeping	Steep solids ready to be dried	Resource, Traceable Item	Moisture content
Germ separation	Denotes the process of separating germ from the corn kernels	Corn kernels are ready for germ separation after degermination	Separated germ is ready for washing and drying and slurry for grinding	Resource, Traceable Item	Flow rates
Wash and dry	Denotes the process of washing and drying of germ	Germ separated from corn kernels is ready for washing and drying	Dried germ is ready for oil extraction	Resource, Traceable Item	Moisture content
Oil extraction	Denotes the process of oil extraction from germ	Dried germ is ready for oil extraction	Extracted oil is ready to be packed	Resource, Traceable Item	Oil quality
Grinding mill ready	Denotes that the grinding mill is ready	Grinding mill available	Grinding mill ready for use	Resource	Grinding mill cleaned
Grind	Denotes the process of grinding the slurry from germ separation	Slurry from germ separation is ready to be ground	Ground slurry is ready to be washed	Resource, Traceable Item	Mill clearance
Wash	Denotes the process of washing the ground slurry	Ground slurry is ready to be washed	Hulls separated from wash ready to be dried and remaining mixture to be centrifuged	Resource, Traceable Item	Moisture content
Centrifugal separator ready	Denotes that the centrifugal separator is ready	Centrifugal separator available	Centrifugal separator ready for use	Resource	Centrifugal separator cleaned
Centrifuge	Denotes the process of centrifugal separation of gluten and starch	Remaining mixture after grinding ready for centrifuge separation	Gluten and starch separated using a centrifuge: gluten ready to be dried and starch to be washed	Resource, Traceable Item	Flow rates, specific gravity (Baume degrees)
Washing filter ready	Denotes that the washing filter is ready	Washing filter available	Washing filter ready for use	Resource, Traceable Item	Washing filter cleaned
Starch wash	Denotes the process of washing starch	Starch separated by centrifuge ready to be washed	Washed starch ready for drying and sugar conversion	Resource, Traceable Item	Moisture content, specific gravity

Table 7. (continued)

State	Description	Start	End	Objects	Quality control
Starch drier ready	Denotes that the starch drier is ready	Starch drier available	Starch drier ready for use	Resource	Starch drier cleaned
Feed drier ready	Denotes that the feed drier is ready	Feed drier available	Feed drier ready for use	Resource	Feed drier cleaned
Dry	Denotes the separate processes of drying starch, hulls and gluten	Products ready for drying	Dried products ready to be packed	Resource, Traceable Item, Actor	Moisture content
Syrup/sugar conversion	Denotes the process of converting starch into syrup/sugar	Washed starch ready for conversion to syrup/sugar	Syrup/sugar ready to be packed	Resource, Traceable Item	Sugar quality
Pack	Denotes the process of packing of various products	Products ready to be packed	Packed products ready to be stored	Resource, Traceable Item	NA
Store	Denotes the process of managing stock	Products ready for storage	Products stored	Traceable Item	Temperature
Transit out	Denotes the process of physical shipping of goods out from the production plant	Products ready for shipping	Products shipped	Resource, Traceable Item, Actor	NA

Table 8. Description of events in the corn wet milling process

No.	Transition	From state	To state	Description
1	Grain railcar to be received	Start state Another actor	Transit in	This transition denotes that the railcar containing grain is in transit to the corn wet milling plant
2	Grain to be received	Transit in	Product receiving	This transition denotes that the transfer of grain from railcar to production plant
3	Conveyor made ready for use	Start state	Conveyor ready	This transition denotes that the conveyor is made ready for use in product receiving
4	Railcar to be emptied	Product receiving	Railcar empty	This transition denotes that the transfer of grain from railcar into the production plant
5	Railcar to exit	Railcar empty	End state	This transition denotes that the empty railcar left the production plant
6	Grain to be stored	Product receiving	Store	This transition denotes the storing of received grain in storage bins
7	Grain to be cleaned	Store	Clean	This transition denotes that stored grain is cleaned before starting the wet milling process
8	Cleaning equipment made ready to use	Start state	Equipment ready	This transition denotes that the equipment is made ready for product cleaning
9	Clean grain (corn) to be steeped	Clean	Steep	This transition denotes that clean corn kernels are transferred to the steep tanks
10	Steep tank made ready for use	Start state	Steep tank ready	This transition denotes that the steep tank is made ready for the steeping process
11	Steeped kernels to be degerminated and steep water to be evaporated	Steep	Degerminate Evaporate	This transition denotes that the corn kernels after steeping enter degermination process while the steep water is evaporated to recover the solids
12	Degerminator made ready for use	Start state	Degerminator ready	This transition denotes that the degerminator is made ready for degermination of corn kernels

Table 8. (continued)

No.	Transition	From state	To state	Description
13	Evaporator made ready for use	Start state	Evaporator ready	This transition denotes that the evaporator is made ready for evaporation of steep water
14	Germ to be separated from degerminated corn kernels	Degerminate	Wash & Dry Grind	This transition denotes that the germ part is separated from the corn kernels after steeping
15	Feed drier made ready for use	Start state	Feed drier ready	This transition denotes that the feed drier is made ready for drying
16	Steep water solids to be dried	Evaporate	Dry	This transition denotes that the steep solids are dried using the feed drier
17	Dried products to be packed	Dry	Pack	This transition denotes that the dried products including hull and gluten are packed
18	Germ to be washed and dried	Germ separation	Oil extraction	This transition denotes that the germ separated from corn kernels is washed and dried
19	Grinding mill made ready for use	Start state	Grinding mill ready	This transition denotes that the grinding mill is made ready to grind the corn kernels
20	Dried germ to be used for oil extraction	Wash & dry	Oil extraction	This transition denotes that the washed and dried germ is used to extract corn oil
21	Ground corn kernels to be washed	Grind	Wash	This transition denotes that the ground corn kernels are washed
22	Corn oil to be packed	Oil extraction	Pack	This transition denotes that the corn oil is packed
23	Ground kernels ready to be separated into constituents	Wash	Centrifuge Dry	This transition denotes that the ground corn kernels are washed to separate hulls which are dried and rest is centrifuged to separate gluten and starch
24	Centrifugal separator made ready for use	Start state	Centrifugal separator ready	This transition denotes that the centrifugal separator is made ready to centrifuge the gluten-starch mix
25	Centrifuged parts to be dried or washed	Centrifuge	Starch wash Dry	This transition denotes that the separated gluten is dried and starch is washed
26	Washing filter made ready for use	Start state	Washing filter ready	This transition denotes that the washing filter is made ready to wash the separated starch
27	Starch to be dried or converted into sugar	Starch wash	Dry Syrup/sugar conversion	This transition denotes that the washed starch is dried into dry starch or converted into syrup/sugar
28	Starch drier made ready for use	Start state	Dry	This transition denotes that the starch drier is made ready to dry starch
29	Dried starch to be packed	Dry	Pack	This transition denotes that the dry starch is packed
30	Syrup/sugar to be packed	Syrup/sugar conversion	Pack	This transition denotes that the syrup/sugar is packed
31	Packed products to be stored	Pack	Store	This transition denotes that the packed products obtained from corn wet milling process are stored
32	Packed products to be delivered	Store	Transit out	This transition denotes that the stored products are taken from storage for shipping
33	Products shipped	Transit out	End state Another actor	This transition denotes that the products are shipped and outside the control of the production plant

Table 9. Description of states for farmer entity

State	Description	Start	End	Objects	Quality control
Planter ready	Denotes that the planter is ready to be used for planting seeds	Planter cleaned	Planter ready for use	Resource	Planter cleaned
Planting	Denotes the process of planting seeds	Seeds to be planted	Seeds planted in field	Resource, Traceable Item	NA
Equipment ready	Denotes that the equipment is ready for seed treatment	Equipment cleaned	Equipment ready for use	Resource	Equipment cleaned
Seed treatment	Denotes the process of treating seeds: applying pesticides, fungicides, etc.	Planted seeds to be treated	Planted seeds treated appropriately	Resource, Traceable Item	Application rates
Harvester ready	Denotes that the harvester is ready for harvesting the crop	Harvester cleaned	Harvester ready for use	Resource	Harvester cleaned
Harvesting	Denotes the process of harvesting the crop	Crop ready to be harvested	Crop harvested	Resource, Traceable Item	Yield
Transport	Denotes the process of transporting harvested crop to on-farm storage	Harvested crop to be transported	Crop transported to storage	Resource, Traceable Item	NA
Store	Denotes the process of storing the crop on on-farm storage	Crop ready to be stored	Crop stored in storage bins	Resource, Traceable Item	Grain quality (moisture)
Transit out	Denotes the process of transporting and selling the crop to an elevator	Crop ready to be transported	Crop transported and sold to an elevator	Actor, Resource, Traceable Item	NA

Table 10. Description of events for farmer entity

No.	Transition	From state	To state	Description
1	Planter made ready for use	Start state	Planter ready	This transition denotes that the planter is made ready to plant seeds
2	Equipment made ready for use	Start state	Equipment ready	This transition denotes that the equipment is made ready for seed treatment
3	Planted seeds to be treated	Planting	Seed treatment	This transition denotes that the planted seeds are treated
4	Harvester made ready for use	Start state	Harvester ready	This transition denotes that the harvester is made ready for harvesting the crop
5	Crop to be harvested	Seed treatment	Harvesting	This transition denotes that the crop is harvested using the harvester
6	Harvested crop to be transported to storage	Harvesting	Transport	This transition denotes that the harvested crop is transported to on-farm storage
7	Crop to be stored	Transport	Store	This transition denotes that the harvested crop is stored in storage bins on farm
8	Stored crop to be transported to elevator	Store	Transit out	This transition denotes that the crop is taken from storage to be transported to the next supply chain entity (an elevator)
9	Crop shipped	Transit out	End state Another actor	This transition denotes that the crop is sold to the elevator and outside the control of the farmer

Table 11. Description of states for elevator entity

State	Description	Start	End	Objects	Quality control
Transit in	Denotes that grain is received at elevator from farm	Grain to be received	Grain received at elevator	Actor, Resource, Traceable Item	NA
Quality check	Denotes the process of grading grain by checking quality	Grain ready to be graded	Grain graded	Resource, Traceable Item	Moisture, test weight, damaged matter and foreign matter
Conveyor ready	Denotes that the conveyor is ready (clean) to be used for transferring grain	Conveyor cleaned	Conveyor ready for use	Resource	Conveyor cleaned
Product receiving	Denotes that the grain is received by conveying into the storage bins	Grain ready to be conveyed to the storage bins	Grain transferred to the storage bins	Resource, Traceable Item	NA
Truck empty	Denotes that the truck is emptied after transferring grain into storage bins	Grain being transferred	Truck empty	Resource	NA
Store	Denotes that the grain is stored in the storage bins at the elevator	Grain ready to be stored after conveying	Grain stored until ready to be shipped	Resource, Traceable Item	Grain quality, temperature
Equipment ready	Denotes that the equipment is ready for blending grain	Blending equipment cleaned	Blending equipment ready for use	Actor, Resource, Traceable Item	Equipment cleaned
Blend	Denotes that the grain is blended before shipment to meet customer specifications	Grain ready to be blended	Grain blended according to specifications	Resource, Traceable Item	Quality specifications
Load	Denotes that the blended grain is ready to be loaded on railcars	Blended grain ready to be loaded	Grain loaded on railcars	Resource, Traceable Item	NA
Transit out	Denotes the process of transporting the grain to a processor	Grain ready to be transported	Grain transported to a corn wet miller	Actor, Resource, Traceable Item	NA

Table 12. Description of events for elevator entity

No.	Transition	From state	To state	Description
1	Grain truck to be received	Start state Another actor	Transit in	This transition denotes that the truck containing grain is in transit to the elevator
2	Received grain to be graded	Transit in	Quality check	This transition denotes that the received grain is graded by quality check at the elevator
3	Grain to be received	Quality check	Product receiving	This transition denotes that the grain is received at the elevator
4	Conveyor made ready for use	Start state	Conveyor ready	This transition denotes that the conveyor is made ready for transferring grain
5	Truck to be emptied	Product receiving	Truck empty	This transition denotes that the transfer of grain from truck to the elevator
6	Truck to exit	Truck empty	End state	This transition denotes that the empty truck left the elevator
7	Grain to be stored	Product receiving	Store	This transition denotes the storing of received grain in storage bins
8	Grain to be blended	Store	Blend	This transition denotes that the grain is blended to meet customer specifications
9	Equipment made ready for use	Start state	Equipment ready	This transition denotes that the blending equipment is made ready for use
10	Blended grain to be loaded on railcars	Blend	Load	This transition denotes that the blended grain is loaded on railcars
11	Grain to be transported to processor	Load	Transit out	This transition denotes that the railcars are prepared to be transported to the next supply chain entity (corn wet milling plant)
12	Grain shipped	Transit out	End state Another actor	This transition denotes that the grain is transported to the corn wet milling plant and outside the control of the elevator

CHAPTER 7. GENERAL CONCLUSIONS

1 Conclusions

In conclusion, this research has provided a holistic approach for minimizing food safety risk in bulk product supply chains. Several methods have been proposed for traceability and information exchange on various food supply chains, however, techniques for implementing internal traceability systems at food production facilities is lacking. This is particularly true for bulk food production industry. Bulk products replicate the fluid-like properties and normally undergo a continuous production process which makes it impossible to define a fixed lot-size of the traceable unit. To overcome this problem, this research focused on developing operational techniques for traceability in bulk product supply chains with special focus on commodity grain.

First, a framework for implementing traceability in the grain supply chain in United States was developed based on a systems approach. The usage requirements of this system were defined and information exchange protocols were discussed. Second, an internal traceability relational database model was developed for a grain elevator to record all product, quality and supplier/customer information. This database system can be queried to retrieve information related to incoming, internal and outgoing lots and to retrieve information that connects the individual incoming grain lots to an outgoing shipment.

In the third part of this research, an optimization technique was developed at an elevator level for minimizing the traceability effort in case of a food safety emergency. A mathematical multi-objective mixed integer programming (MIP) model was proposed with two objective functions; to calculate the minimum levels of lot aggregation and minimum total cost of blending grain in order to meet the customer contract specifications. Pareto optimal front was computed for simultaneous optimization of lot aggregation and cost of blending. Finally, a novel methodology for modeling the traceability information using the UML statecharts following an event management approach in bulk food production is introduced. In order for any traceability system to meet one of its most important requirements of ensuring food quality and safety, there is need to integrate all this information into one system so that a problem caused either due to processing or handling/logistics can be identified and traced back to the source.

Therefore, we integrate the product, process and quality information into the data that is recorded when transition takes place from one state to another. Food safety and quality issues generally occur due to incorrect processing and handling of food products. Bulk food production also has other challenges including product transformations such as blending or splitting of batches. Monitoring the flow of products, their quality and the process parameters throughout production and linking them to each transition in state of the products is an effective way of implementing and ensuring product safety and traceability.

2 Future Research

The focus of this study was to develop operational techniques for implementing traceability in bulk product supply chains to minimize the food safety risk. In future, the modeling techniques developed in this study need to be implemented by the food industry. In addition, there is a need to develop optimization strategies for initial handling of the bulk products, for instance, the initial bin assignments for the incoming grain lots at a grain elevator.

Sector-specific standards must be developed for information management in the food industry. Internal traceability data management systems must be implemented by all actors in a supply chain to effectively link raw materials with semi-processed and finished products. This would lead to faster response in identification of contaminated products during food processing as well in case of a recall.

APPENDIX A: Additional papers

This section includes additional papers published during the course of my doctoral program are relate to the areas of food safety and traceability.

Data Mining for Recognizing Patterns in Foodborne Disease Outbreaks

Published in *Journal of Food Engineering* (2010), 97(2):213-227

Maitri Thakur[1, 2, *], Sigurdur Olafsson[2], Jong-Seok Lee[2] and Charles R. Hurburgh[1, 3]

[1]Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011
[2]Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA 50011
[3]Department of Food Science and Human Nutrition, Iowa State University, Ames, IA 50011
* Primary author, Corresponding author

Abstract

This paper introduces a new methodology for discovering patterns in foodborne disease outbreaks using a data-driven approach. Specifically, our approach uses three data mining methods, namely attribute selection, decision tree learning, and association rule discovery, to extract previously unknown and meaningful patterns that connect specific types of foodborne diseases outbreaks with associated foods vehicles and consumption locations. We use this approach to study the four most common disease causing etiologies in the Center for Disease Control (CDC) database of foodborne disease outbreaks in the year 2006, namely *salmonella enteritidis*, *salmonella typhimurium*, *e. coli* and *norovirus*. The analysis reveals numerous patterns of how each of these outbreaks types relates to specific foods and locations. The discovery of such patterns in foodborne disease outbreak data can be very useful is determination and implementation of suitable intervention techniques. In particular, if the associations between different food types and consumption locations are known then custom intervention techniques including specific training methods can be designed to train individuals in hygienic food handling, preparation and consumption practices.

Keywords: foodborne disease outbreaks, surveillance databases, data mining, classification, association rule mining, attribute selection

1. Introduction

Food safety and food control continue to gain significant attention as our food supply chains and production practices become increasingly complex. Food safety is in fact a very important part of public health, and although several advanced surveillance and monitoring systems exist in developed countries, outbreaks of foodborne diseases continue to be commonplace. Such foodborne diseases are caused by consumption of contaminated foods or beverages. There are many different types of foodborne infections as many disease-causing microbes or pathogens can contaminate foods. In addition to these, several poisonous chemicals can also cause foodborne diseases if present in food (CDC, 2005). According to the Center for Disease Control and Prevention (CDC), an outbreak of foodborne illness occurs when a group of people consume the same contaminated food and two or more of them come down with the same illness. CDC (2005) estimates that foodborne diseases cause 76 million illnesses, 325,000 hospitalizations and 5,000 deaths in the United States every year.

1.1 Foodborne Disease Surveillance

Each state makes a decision regarding which diseases are to be under surveillance and the public health departments monitor these important diseases. In most states, the diagnosed cases of certain serious infections are reported to the health department, which in turn reports them to the CDC through the National Outbreak Reporting System (NORS). The reported data is investigated by the CDC to obtain information regarding the role of food in the outbreaks. The surveillance of foodborne disease outbreaks serves three main purposes (Olsen et al., 2000). The first purpose is to establish prevention and control measures in the food industry by identification of critical control points by the public health officials. Similar changes at all levels in the food production, handling and consumption contribute to a safer food supply chain. Secondly, the outbreak investigations provide critical means for identifying new and emerging pathogens, as well as maintain awareness about ongoing problems. Finally, analysis of several years of data provides epidemiologists ways to monitor trend over time in the prevalence of outbreaks caused by specific etiologies, foods and mistakes in food handling practices. This information provides the basis for regulatory changes and other advances to improve food safety.

Foodborne outbreak investigations, if carried out in a timely and systematic manner, aid in rapid identification of corresponding etiologies, which can then lead to appropriate prevention and control measures. The CDC surveillance system, and the Outbreak Surveillance Data made available through their website, has certain limitations in the way data is recorded. For instance, food vehicles of disease transmission can be classified in two different ways, both as individual food items (e.g. lettuce) and as food categories (e.g. salad, multiple vehicles). It can therefore be difficult to identify the item that contained the foodborne pathogens. There are also several cases where the etiologies are either unknown or unconfirmed. The CDC reports that in certain cases, the pathogens are not identified because of delayed or incomplete laboratory investigation, or inability to recognize a pathogen as a cause of foodborne disease.

This paper focuses on finding patterns involving specific food vehicles and locations, and connecting them to the type of outbreak. By food vehicle we mean the type of food that is believed to be the cause of the foodborne disease outbreak, and we will often refer to this simply as the vehicle. By location we mean the type of place where the outbreak occurred (e.g., home, office, hospital). There are many studies that look at the foodborne disease outbreaks caused by different foods or different locations. For example, Levin, et al. (1991) study foodborne disease outbreaks in nursing homes, and Cody, et al. (1999) study E. coli infections caused by unpasteurized commercial apple juice. According to Dewaal et al. (2006), it is important to know which foods are most frequently linked to outbreaks, because identifying specific food/hazards combinations allows for better targeting of food safety interventions. This study also emphasizes the evaluation of contamination locations to identify factors such as cross-contamination and inadequate personal hygiene.

Considerable work has thus been done analyzing specific food types and locations, and there is a general understanding of the importance of identifying links between foods and locations on the one hand and types of food outbreaks. However, no studies appear to have been conducted to extract possible hidden patterns in the disease outbreaks and relationships between different food types and outbreak locations based on automated data-driven learning. While not guaranteed to exist, such hidden patterns do exist in many databases. To address this gap, we suggest the use of data mining techniques to extract hidden patterns from the CDC Outbreak Surveillance Data of foodborne diseases.

1.2 Data Mining

Data mining is a semi-automated process of extracting meaningful, previously unknown patterns from large databases (Han and Kamber, 2001). In recent years, data mining techniques have been found to be useful in many application areas, including safety areas such as drug safety (Hochberg et al., 2007) and aviation safety (Nazeri et al., 2001), but as stated above its application in food safety appears to be largely unexplored. The increased popularity of data mining can be traced to the fact that data collection and storage has become easier, leading to massive databases that often contain a wealth of data that traditional methods of analysis fail to transform into relevant knowledge. Specifically, meaningful patterns are often hidden and unexpected, which implies that they may not be uncovered by hypothesis-driven methods. In such cases, inductive data mining methods, which learn directly from the data without an a priori hypothesis, can be used to uncover the hidden patterns that can then be transformed into actionable knowledge.

To illustrate the difference between data mining and traditional hypothesis-driven methods, consider how patterns may be found in a database such as the previously mentioned Outbreak Surveillance Data maintained by the CDC to track foodborne diseases. In a hypothesis-driven analysis, an analyst might query the database for all outbreaks that match a certain criteria, such as all *salmonella typhimurium* outbreaks involving potato salads at a wedding reception. But unless there is an expectation of a connection between *salmonella typhimurium*, potato salads and wedding receptions, that query is unlikely to be made. On the other hand, a data-mining approach can automatically extract from the database that when an incident description discusses potato salads and a wedding reception, then the outbreak is likely to involve *salmonella typhimurium*; thus generating a pattern of interest without any preexisting knowledge about this pattern. In other words, what defines data mining is that by employing data-driven methods, it can extract previously unknown and potentially useful knowledge from large databases.

The data mining process consists of numerous steps, which may include data integration, preprocessing of the data, and induction of a model with a learning algorithm. The model can then be used to identify and implement actions, such as interventions to reduce outbreaks of foodborne diseases. All data mining starts with a set of data called the training set, which consists of instances describing the observed values of certain variables. These instances are then used to learn a given target concept or pattern. One of the main approaches to learning a pattern is *classification* (Han and Kamber, 2001). In classification the training data is labeled, meaning that each instance is identified as belonging to one of two or more classes, and an inductive learning algorithm is used to create a model that discriminates between those class values. The label can for example be the specific etiology of a foodborne disease outbreak, such as *salmonella typhimurium*, and the model classifies each incident as either a *salmonella typhimurium* outbreak (positive) or not (negative). This model can then be used to classify any new instances according to this class variable, for example, to predict the etiology of an outbreak. The primary objective is usually for the classification to be as accurate as possible, but accuracy is not the only relevant measure of the quality of the model. The interpretability of the results of the model is also extremely important. For example, rather than predicting the etiology of an outbreak, it may be of more interest to understand *why* a specific type of etiology is predicted, which would provide insights into the circumstances of this when this type of outbreak occurs.

Data preprocessing is also an important part of data mining. The initial data preparation is very significant since to mine any useful knowledge from the raw data it must typically be transformed considerably. Specifically, it is often of great value to reduce the dataset to the most valuable data, and specifically to focus the analysis on the most important or most relevant variables. Variable or *attribute selection* has been relatively well studied for decades and some simple attribute selection is a standard part of most data mining projects (Liu and Motodo, 1999; Olafsson et al., 2008). Attribute selection involves a process for determining which variables or attributes are relevant in that they predict or explain the data, and conversely which attributes are redundant or provide little information. Such elimination of many or even most of the attributes makes it easier to train other learning models. The resulting model may also be simpler, which makes it easier for an analyst to interpret and thus more useful in identifying root causes and transform such insights into interventions. Identifying relevant attributes may also provide valuable information directly, such as showing which locations and/or foods are predictive of a specific etiology, and is therefore important in its own right. On the other hand, when attribute selection is used as preprocessing prior to classification, it is also possible that an attribute will be removed that would have been found valuable by the classification learning algorithm. Thus, in our analysis we perform the classification learning both with and without attribute selection.

Association rule discovery is another important type of learning method (Hipp et al., 2000), but unlike classification it is unsupervised and the data is unlabelled. This means that there is no specific class attribute, but rather the learning algorithm aims to discover interesting correlation between any attributes (Agrawal et al. 1993). Those correlations are represented as association rules $X \Rightarrow Y$, where both the antecedent X and the consequent Y are sets of attribute-value pairs, called item sets. An example of an association rule is a relationship such as 'location is wedding reception & vehicle is potato salad ⇒ etiology is *salmonella typhimurium*.'

An association rule has three measures that express the degree of uncertainty about the rule, and those numbers are used to select interesting rules from the set of all possible rules. The first measure as a probability is called the *support* for the rule that can be defined as below, and it is simply the portion of instances that contain all items in the antecedent and consequent parts of the rule.

$$Support(X \Rightarrow Y) = P(X \cap Y)$$

The *confidence* of the rule, which is the ratio of the number of instances that include all items in the consequent as well as the antecedent to the number of instances that include all items in the antecedent, can, by its definition, be interpreted as the probability of finding the consequent part of the rule in instances under the condition that these instances also include the antecedent part. Therefore, the confidence is given by

$$Confidence(X \Rightarrow Y) = P(Y \mid X) = \frac{P(X \cap Y)}{P(X)}.$$

The last measure, which is the *lift* of the rule, is the ratio of the confidence to the expected confidence (Berry and Linoff, 1997). The expected confidence means the confidence where the antecedent part does not enhance the probability of occurrence of the consequent part. It is the number of transactions that include the consequent part divided by the total number of transactions. Hence, the lift value gives us information about the increase in probability of the consequent part given the antecedent part. By such a definition of the lift, a meaningful rule should have the lift value that is greater than one. A lift value that is greater than one means that when the consequent part happens it is more likely that the antecedent happens (positive association), whereas a lift value of less than 1 means that if the consequent happens it is less likely that the antecedent happens (negative association). The lift is calculated as follows:

$$Lift(X \Rightarrow Y) = \frac{Confidence(X \Rightarrow Y)}{P(Y)} = \frac{P(X \cap Y)}{P(X) \cdot P(Y)}.$$

Association rules are required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time. To achieve this, association rule generation is a two-step process. First, minimum support is applied to find all frequent itemsets in a database. In a second step, these frequent itemsets and the minimum confidence constraint are used to form rules. While the second step is straight forward, the first step needs more attention. In order to implement this two-step process, *a-priori* algorithm is the most often used (Agrawal and Srikant, 1994).

2. Objectives

In this study we investigate methods to extract meaningful patterns from a surveillance database of foodborne disease outbreaks in order to improve our understanding of the outbreaks of a specific etiology. In particular, through data mining framework that we believe to be novel for the given application, we study the question of what vehicles and/or locations are associated with specific etiologies, and how outbreaks of those diseases occur. This is an important question as addressing it may help inform successful interventions related to food handling, preparation, and consumption practices.

The data mining framework employs classification, attribute selection, and association rule discovery as the primary learning methods. After developing the framework, we apply it to analyze the four most common outbreak etiologies in the 2006 CDC Outbreak Surveillance Data, namely *salmonella typhimurium*, *salmonella enteritidis*, *E. coli*, and *norovirus*. In addition to the value of the specific patterns obtained for those four etiologies, our framework provides a general approach for using data mining to identify patterns in food safety surveillance databases.

3. Discovering Outbreak Patterns in Surveillance Data

To achieve the objectives of this study, we have designed a framework for extracting meaningful patterns from foodborne illness outbreak surveillance data (see Figure 1). We first briefly describe the data and then explain each component of our new data mining framework.

3.1 Description of Outbreak Surveillance Data

The data for this study was obtained from the Outbreak Surveillance Data from the CDC for the year 2006. All the data was collected electronically through the Electronic Foodborne Outbreak Reporting System (EFORS) and all etiologies are as reported by the states. Table 1 shows the summary of foodborne illness outbreaks in the United States in 2006. A total number of 1247 outbreaks and 25,659 illnesses were reported in the year 2006. Out of 1247 outbreaks, 623 outbreaks had confirmed etiology while 275 etiologies were unconfirmed and 349 were unknown. The dataset from the CDC consists of eight attributes, described in Table 2.

3.2 Data Preparation for Classification

The dataset in its raw format described above is not directly appropriate for data mining. In this section we describe the process of converting such a raw surveillance database into a database that can be used for classification and other data mining. This means that a class attribute needs to be identified or constructed and each of the other attributes needs to be either numeric or nominal, that is, taking a given number of predefined values.

For the Outbreak Surveillance Data in particular, the following issues needed to be addressed:

1. The attribute *vehicle* that describes the types of food consumed and *location* that describes the location of food consumption was present in text format. In cases where multiple foods were consumed they were all grouped under this attribute. Such text data needs to be structured before data mining can be done.
2. It is characteristic of most surveillance databases that there are no negative instances present in the database. In other words, the outbreak information is reported to CDC only when an outbreak occurs by consumption of specific foods, so obviously there are no instances where an outbreak didn't occur on consumption of these foods. To apply a classification algorithm, the data must have two or more class types, for example positive and negative instances so that the algorithm can learn to discriminate between those, and in this case find a model that can predict any new instances of a foodborne outbreak. Thus, a class attribute(s) must be constructed.

3. For almost all etiologies there are relatively few examples of outbreaks. For example, although it is one of the most common types of outbreaks, there are only 28 instances of *salmonella enteritidis* outbreaks in the database. This causes what in data mining is called a class imbalance problem, that is, there are relatively few examples of one class value. The result is that any data mining algorithm tends to ignore the infrequent class unless some action is taken to balance the class value.

The raw Outbreak Surveillance Data was preprocessed to address these issues and thus to set up a classification problem where data mining algorithms can be applied. The first issue, namely that of dealing with text data, is well-known and we used a standard approach that converts a single text attribute into a (large) set of binary variables, each indicating if a word occurs in that text (Lewis, 1992). Specifically, rather than having a single string such as "beef, meatball; green salad; steak, unspecified" describing the vehicle of the outbreak, there are binary variables such as "beef," "black_grouper," "ceaser_salad," and "cheese," where the for the example string the "beef" binary attribute would be set to one and the other three to zero. Words that occurred only once were removed from the dataset since those are not useful for finding general patterns involving multiple outbreaks. This resulted in the two text attributes describing vehicle and location being replaced by 106 binary attributes, with each of those binary attributes describing a specific vehicle or location. Furthermore, since this study focuses on relating the vehicle and location of the outbreak to the etiology, all other attributes were deleted.

To address the second issue, we created the *negative* class type for instances attributed to all etiologies except the one being studied and repeated the process for each etiology. For example, when classifying all outbreaks caused by *e. coli*, all *e. coli* outbreaks were labeled *positive* instances and all others were labeled *negative* instances. We note that this implies that the classification problem does not discriminate between an outbreak of a specific etiology versus safe consumption, but rather between an outbreak of a specific etiology versus outbreaks of some other etiology. The output should hence be interpreted as identifying what is particularly characteristic of one etiology versus another. The same process of adding a class attribute taking two possible values was repeated for other three etiologies being studied, resulting in four classification problems. After adding a class attribute, the final datasets contained 107 attributes and 1206 instances.

For three of the four classification problems the classes are very imbalanced (Gu et al., 2008). For example, as noted above, there are 28 instances of *salmonella enteritidis* outbreaks in the database out of a total of 1167 instances. Thus, there are 28 instances with a positive class value and 1139 instances with a negative class value. The problem with this is that a model that predicts that *salmonella enteritidis* never occurs, simply ignores the minority class value, will be 97.6% accurate, and any learning algorithm will simply find this trivial, highly accurate, but useless classification model. To address this, we use a well-known method of non-uniform resampling to balance the class (Japkowicz, 2000). Specifically, we sample with replacement from the dataset 1167 times, each time giving much higher chance of being sampled to the positive instances, so that in expectation we end up with 583.5 positive and 583.5 negative instances. This means that in the final dataset, many of the original 1139 negative instances will not be present (some may also be present more than once), and each of the original 28 positive instances will be present multiple times. It is important to note that although this type of resampling, or a similar alternative, is inevitable for learning meaningful classification models, this process does introduce a bias, specifically by overemphasizing some of the positive instances that are sampled most frequently. The estimated prediction accuracy for any model learned on the resampled data is therefore not meaningful unless it is estimated independently of the resampling process. However, the objective of this project is not to accurately predict an etiology of an outbreak, but rather to identify patterns that provide insights into how and why outbreaks occur, a purpose for which this bias is not a significant concern. The resampling process does affect our analysis in that different repetitions of the sampling may lead to different patterns being discovered, some of which are likely to be more useful than others. Rather than simply resampling once, it may therefore be valuable to resample repeatedly.

It should be noted that an alternative to the binary classification problem suggested above would be a multiclass classification problem where each we would try to discriminate between all etiologies of interest simultaneously. This would automatically reduce the class imbalance problem, but our experimentation with the data indicated that the multiclass approach did not result in as interesting patterns. The binary class approach was therefore chosen and the multiclass results are not reported in the paper. However, we also caution that this conclusion can only be drawn for the particular classification method tried, and other classification methods might prove valuable for the multiclass problem.

3.3 Identifying Important Vehicle and Location Attributes

We described above how attribute selection is an important part of most data mining projects. Attribute selection may be done simply to improve subsequent data mining models (e.g., in order to obtain a more accurate classification model) or it may be done because identifying relevant attributes is important in its own right. For us both motivations hold. It is of intrinsic interest to identify the vehicle and location attributes that are relevant to being able to predict a specific etiology, as those provide insights into why and where certain outbreaks occur, and removing redundant and irrelevant attributes may also improve the subsequent classification models. Specifically, we will propose using decision trees as the classification model and as we will see in the results reported here, preceding the decision tree learning with attribute selection will result in smaller and easier to interpret trees.

Many methods have been proposed for attribute selection, and no single method can be identified as superior to all others. In our framework, we use either directly or indirectly three separate and complimentary measures of attribute worth. First, we use the Relief algorithm that identifies the attributes that best distinguish between classes if the classification is done based on nearest neighbors, also called instance-based learning (Kira and Rendel, 1992). Second, we use what is called a Wrapper method (Kohavi and John, 1997), which searches through the space of all possible subsets of attributes and evaluates the worth of the attribute subset based on how well it works for classification (that is, the accuracy of the classification model). Specifically, we use the accuracy of a Naïve Bayes classifier induced on the dataset using the particular attribute subset (Domingos and Pazzani, 1997). The Naïve Bayes classifier has been found to work well for text mining, namely datasets such as ours that has a large number of binary attributes. The basic idea of this classifier is to find the most likely class given the data. The third and final method for identifying important attributes is indirect and results from our choice of classification algorithm. As will be described in more details below, we choose a decision tree algorithm and the sequence in which attributes are used to construct the tree is an implicit attribute selection, with the attribute used for the top node judged the most important, and so forth. The measure used by the decision tree is information gain ratio (Quinlan, 1993), which is an information theory derived measure and may be thought of as complimentary to the instance-based and probabilistic measures used to evaluate attribute worth by the other two methods.

The output of the Relief algorithm is a ranked list of attributes, but it does not decide on a specific subset of most valuable attributes. We apply this algorithm before resampling to identify relevant attributes for all of the four etiologies individually. The primary purpose of this is to provide insight into which food vehicle and location factors are related to each foodborne illness outbreak category. The Naïve Bayes wrapper determines a subset of attributes to be used, but it cannot be applied before resampling because the Naïve Bayes algorithm simply identifies the trivial model that ignores all minority class values. The attribute subset found by the wrapper is therefore biased by the resampling, but as noted above this is not a significant concern since our objective is extraction of meaningful scenarios or patterns. This attribute subset is then used by the classification algorithm.

3.4 Classifying Etiology of an Outbreak using Decision Trees

As described above, for each of the four most common specific etiologies, we formulated a classification problem by creating an indicator for all of the incidents of that type. We used only the relevant attributes for each etiology type that were selected by using attribute evaluation techniques discussed in the previous section. Many methods exist for the actual classification, including support vector machine (Burges, 1998; Cortes and Vapnik, 1995), Bayesian methods (Heckerman, 1996), and decision tree induction (Quinlan, 1993).

While it does not usually provide the best prediction accuracy, in our approach we focus on decision tree induction because the resulting model (decision tree) is simple and interpretable, which allows us to achieve the primary objective of the study, namely to gain insights into the interaction between attributes. The process of decision tree induction is to construct a tree in a top-down manner by selecting variables one at a time and splitting the data according to the values of those variables. The most important variable is selected as the top split node, the next most important variable is considered at the next level, and so forth. For example, in the algorithm we employ, called the C4.5 algorithm (Quinlan, 1993), variables are chosen to maximize the information gain ratio in the split. This is an entropy measure designed to increase the average class purity of the resulting subsets as a result of the sequential splits.

We will use decision tree induction both using all of the attributes, and using the subset of attributes selected by the Naïve Bayes wrapper approach discussed above. The expectation is that that these trees will be mostly consistent, but the tree employing attribute selection will be simpler and easier to interpret.

However, some additional patterns regarding specific outbreaks could be extracted from the larger trees as well.

Given information about food consumption and location, the decision trees could be used as a predictive model to predict unknown etiologies and future foodborne disease outbreaks, although the applicability and accuracy of doing so is not evaluated here. Rather, we focus on insights that can be obtained from the decision trees by analyzing specific scenarios represented in the trees. Such insights can then be used to further enhance the decisions regarding intervention techniques and models that can reduce the occurrence of such outbreaks.

3.5 Discovering Associations between Vehicles, Locations and Etiologies

The final component to our data mining framework is to use association rule mining to discover relationships between the attributes in the database. As discussed above, interesting association rules are required to satisfy three user-specified measurements. Considering the sparseness of the dataset, we allowed enough tolerance for the support of a rule by setting the minimum support to three. Only rules having the lift value that is greater than one were under our consideration. Since our expectation is that the most useful rules are of the type ‘if X and Y then Z’, where X is a location information, Y is a food vehicle that caused the outbreak, and Z is a type of etiology, we chose three as the maximum number of items for generating frequent item sets. No lower limit of the confidence was decided to prevent losing some interesting rules due to the sparseness of dataset.

Recall that association rule mining is an unsupervised learning method, that is, it will find relationships called association rules between any attributes. Most of those relationships will therefore not describe the etiologies of interest, and after generating all association rules, we prune them to only include those rules that include one of the target etiologies in the consequent (e.g., *salmonella enteritidis*, *salmonella typhimurium*, *e. coli*, and *norovirus* in the results reported below). Hence, we expect these patterns to provide insights into what types of outbreaks (etiology) are caused by specific types of food items and/or locations.

Note that while being unsupervised is a drawback to using association rule mining to study specific etiologies, as most of the patterns obtained will be discarded, unlike the decision tree learning association rule mining does not require resampling of the database. The estimated lift and confidence of each association rule will therefore be unbiased.

4. Results

In this section we use the data mining framework described above to analyze outbreaks of the four most common etiologies of foodborne illness outbreaks.

4.1 Analysis of *Salmonella Enteritidis* Outbreaks

Salmonella enteritidis is a bacterium found inside eggs and can cause illness, called salmonellosis if contaminated eggs are consumed raw or undercooked. The current salmonella outbreaks are caused by intact and disinfected eggs. Government agencies and egg industry has taken several steps to reduce *salmonella enteritidis* outbreaks which includes identifying and removing infected flocks from the egg supply and increasing quality assurance and sanitation measures. According to CDC, every year, approximately 40,000 cases of salmonellosis are reported in the United States. Because many milder cases are not diagnosed or reported, the actual number of infections may be thirty or more times greater. It is estimated that approximately 400 persons die each year with acute salmonellosis.

When applying our data mining framework to *salmonella enteritidis* outbreaks, the first type of pattern obtained is a list of attributes found to be the most relevant in classifying this etiology versus another etiology. The attribute selection outputs a ranked list, and the order of each attribute is given in brackets. We list the ten most relevant attributes, and those are shown in Table 3. Note that the table organizes the most relevant attributes according to their type (location versus vehicle) and whether they are an indicator of the target etiology (*salmonella enteritidis*) or if they indicate that the etiology of the outbreak is something else.

In attribute selection an attribute can be found important either because it is strongly indicative of positive classification (that is, *salmonella enteritidis* outbreak), or a negative classification (that is, any other outbreak). In Table 3 eight out of the ten attributes indicate negative classification (not *salmonella enteritidis*). This is not an unexpected outcome since the class values are highly unbalanced (28 positive versus 1139 negative instances). From Table 3 we can observe that if the location is either a private home or a banquet facility then the outbreak is relatively more likely to be *salmonella enteritidis* than another type of outbreak, and we have a list of five locations and three vehicles where *salmonella enteritidis* is unlikely to be the cause of an outbreak.

The second type of pattern obtained is a decision tree classifying outbreaks as either positive or negative for *salmonella enteritidis* etiology. This decision tree learned without attribute selection is shown in Figure 2 and from it we can observe relationships between the target etiology of the outbreak and different foods and consumption locations. For example, the decision tree shows that in an outbreak where *beef* was consumed at *private home,* the disease can be attributed to *salmonella enteritidis* etiology.

Figure 3 shows the decision tree for *salmonella enteritidis* without attribute selection. As expected, the tree is somewhat larger than before (11 leave nodes versus 9 before). First note that the new decision tree has the same root node, *prison/jail* which shows that this attribute provides the maximum information gain. New locations, *restaurant/deli* and *school* are simultaneously linked to salmonella enteritidis outbreaks. *Beef* consumed at *private/home* and *restaurant/deli* is also simultaneously linked to these outbreaks. *Ground beef* consumed at *workplace - not cafeteria* also caused some outbreaks while l*ettuce* and *turkey* are linked to outbreaks other than *salmonella enteritidis*. The results found by this technique are consistent with the attributes selected and the association rules found that are shown in Figure 4.

The third and final type of pattern is a set of association rules linking the target etiology in the consequent with location and vehicle attributes in the antecedent. The rules with the highest lift and confidence are shown in the bar chart in Figure 4. Note that the relevant attributes to each type of etiology, i.e. the antecedent part of the rule, are shown in the horizontal axis with their lift values and confidence values. From the figure we observed that the lift value of prison/jail in which *salmonella enteritidis* was involved is approximately 6.5. This means that the probability that prison/jail will be involved in *salmonella enteritidis* is 6.5 times higher than the general probability of prison/jail in the dataset. Similar interpretations can be made on the rules involving the other attributes: private home, banquet facility, ground beef, and beef.

Given the complementary nature of the three methods of extracting patterns, it is worth noting when the same pattern is found by two or more methods. For predicting *salmonella enteritidis* outbreaks, two locations, namely private home and banquet facility, are found to be indicative of this type of outbreak by all three methods. Furthermore, the location of prison or jail is found to be the most important indicator of *salmonella enteritidis* outbreaks by both the decision tree and the association rule mining, and both of those methods also identify the food vehicle beef as the second most important indicator of an outbreak. While outside the scope of this paper, these results call for further analysis of what causes such outbreaks to be particularly prevalent in these three locations, as well as why this infection that is transmitted through eggs appears to have a strong the connection with beef, especially beef in a private home as indicated by the decision tree.

4.2 Analysis of *Salmonella typhimurium* Outbreaks

Salmonella typhimurium is among the most common Salmonella bacterium causing salmonellosis in the United States. *Salmonella typhimurium* multiplies in the gastrointestinal tract of many animal species where it usually causes no disease, but in humans its growth causes gastroenteritis. Isolations of Salmonella causing gastroenteritis in humans have increased in recent years in developed countries, primarily because modern methods of animal husbandry, food preparation, and distribution encourage the spread of Salmonella (Resource Center for Biodefense Proteomics Research, 2009). Contaminated foods are often beef, poultry, milk and eggs, but according to CDC, any foods, including vegetables, can become contaminated if they come into contact with feces from an infected animal.

To extract interesting patterns related to *salmonella typhimurium* outbreaks, we repeat the data mining analysis as in the previous section. The most relevant attributes are reported in Table 4. From the table we note that two locations (restaurant or deli and private home) are strongly linked to *salmonella typhimurium* outbreaks, whereas several others (especially banquet facility, which was the second highest ranked attribute overall) indicate that the etiology of the outbreak is something else. One food vehicle, namely chicken, is indicated as a relatively common cause of outbreaks (versus a cause for some other outbreak), whereas lettuce is more likely to be a vehicle for an outbreak with a different etiology. All three of the positive indicators (restaurant or deli, private home, and chicken) ranked as one of the four top attributes, indicating a fairly strong relationship.

Figure 5 represents the decision tree obtained for *salmonella typhimurium* outbreaks by learning the decision tree with attribute selection. There was a known *salmonella typhimurium* outbreak caused by tomatoes in 2006 (FDA, 2006). The positive instances are classified by *tomatoes* consumption is 100%. But other outbreaks that could not be attributed to tomatoes can be analyzed using this decision tree. In this case, it is very interesting to see that the two different locations: *fair/festival/temporary mobile device* and *private home* are simultaneously related to *salmonella typhimurium* outbreak.

Figure 6 represents the decision tree obtained without attribute selection. This decision tree is now much more complex than the one obtained with attribute selection, but reveals some additional patterns. *Chicken teriyaki* was not present in the decision tree with attribute selection but is the root node in this case. This decision tree provides additional information about the food and location combinations. For example, it can be noted that *turkey* consumed at *prison/jail* caused *salmonella typhimurium* outbreaks. Similarly, *tomatoes* consumed at locations other than *hospital* caused these outbreaks. Some of the attributes linked to *salmonella typhimurium* outbreaks were same as those chosen by the Relief algorithm (Table 4). For instance, the decision tree shows that *chicken* consumed at *private home* is linked to these outbreaks. The results obtained were consistent with both attribute selection and association rule mining.

As before, we also obtain association rules with *salmonella typhimurium* in the consequent, and Figure 7 reports the rules with the highest lift and confidence. For this type of outbreak four rules are obtained, and of those one involves a combination of location and food vehicle and one involves two locations:

- restaurant or deli & chicken ⇒ *salmonella typhimurium*
- restaurant or deli & private home ⇒ *salmonella typhimurium*

It is again interesting to note the patterns that are found by two or more of our methods. Here, two locations, namely private home and restaurant or deli, and one food vehicle, namely chicken, are found as indicators of *salmonella typhimurium* outbreaks by both the attribute selection and the association rule mining. Neither method finds any other positive relationships so there is a perfect match between those two methods. The association rule mining further identifies interesting combinations of those attributes as noted before. The decision tree also finds that the location of private home indicates this type of outbreak, but does not include the other two attributes. (Note, however, that the decision tree does indicate the known tomato related outbreak of *salmonella typhimurium* in 2006, whereas the other two methods do not.)

4.3 Analysis of *E. coli* Outbreaks

E. coli are a bacterium that live in the guts of ruminant animals, including cattle, goats, sheep, deer, and elk. The major source for *E. coli* outbreak is cattle (Foodborne illness, 2005). CDC estimates that *E. coli* causes about 70,000 infections in United States each year. Exposures that result in illness include consumption of contaminated food, consumption of unpasteurized milk, consumption of water that has not been disinfected, contact with cattle, or contact with the feces of infected people. Some foods are considered to carry such a high risk of infection with *E. coli* and include unpasteurized milk, unpasteurized apple cider, and soft cheeses made from raw milk.

We next conduct our data mining analysis with *E. coli* as the target etiology of the outbreaks. The most relevant attributes are reported in Table 5. This table indicates two vehicles that are strong indicators of *E. coli* outbreaks versus other types of outbreaks, namely lettuce and milk. It also shows two locations where if an outbreak occurs this etiology is indicated, namely restaurant or deli or private home; and several locations that indicate another etiology. Finally, if the vehicle is chicken then an etiology other and *E. coli* is indicated.

Figure 8 shows the decision tree for *E. coli* related outbreaks. This decision tree, which is learned following attribute selection, is quite simple compared to those for *salmonella enteritidis* and *salmonella typhimurium*. It contains just one consumption location and other nodes represent different foods that were related to the *E. coli* outbreaks. *Steak* is chosen as the root node of this tree which suggests that the highest information gain is provided by this attribute. We note that the two vehicles indicated as being linked with *E. coli* outbreaks by the selection of relevant attributes (milk and lettuce) are also present in the decision tree.

Figure 9 shows the decision tree without attribute selection. Again, this tree is much more complex and difficult to interpret than the tree utilizing attribute selection (Figure 8), but some additional interesting patterns are discovered from this decision tree. The decision tree for *E. coli* related outbreaks with attribute selection was very simple. It did not provide information about food-location combinations that were linked with these outbreaks. The decision tree using all attributes provides this information. For example, *milk* consumed at *private home* is found to be linked to several *E. coli* outbreaks. Similarly, *ground beef* consumed at locations other than *workplace-not cafeteria* and *banquet facility* are linked with these outbreaks. *Lettuce* consumed at *restaurant/deli* is also linked with *E. coli* outbreaks. These findings are consistent with the attributes selected by Relief algorithm and association rules found for these outbreaks.

The association rules obtained that include *E. coli* in the consequent are reported in Figure 10, and we note that these rules are significantly stronger than those reported for the other etiologies. For example, the lift value of spinach is almost thirty and the confidence of spinach is greater than 60%. It means that the rule, 'spinach ⇒ *e. coli*', is highly promising. The other selected attributes overall have very high lift values with good confidence numbers. Five rules involving a combination of location and food vehicle are obtained, namely:

- Restaurant or deli & private home ⇒ *E. coli*
- Milk & private home ⇒ *E. coli*
- Ground beef & private home ⇒ *E. coli*
- Restaurant or deli & lettuce ⇒ *E. coli*
- Restaurant or deli & ground beef ⇒ *E. coli*

Two food vehicles are identified by all three methods, namely milk and lettuce, as being indicators of *E. coli* outbreaks. Spinach is also identified by both the decision tree and the association rule mining as being an important vehicle for this disease. Furthermore, restaurant or deli and private home are identified by both the attribute selection and the association rule mining as locations where such outbreaks occur relatively frequently. Further analysis of those three food types and two locations is therefore indicated by the data mining results.

4.4 Analysis of *Norovirus* Outbreaks

Noroviruses are a group of related, single-stranded RNA viruses that cause acute gastroenteritis in humans. *Noroviruses* are transmitted primarily through the fecal-oral route, either by consumption of fecally contaminated food or water or by direct person-to-person spread (CDC, 2006). CDC estimates that 23 million cases of acute gastroenteritis are due to *norovirus* infection, and that at least 50% of all foodborne outbreaks of gastroenteritis can be attributed to *norovirus*.

As the final illustration of our data mining framework, we analyze outbreaks with *norovirus* as the target etiology. The most relevant attributes are reported in Table 6. We observe that there is one location that indicates outbreaks where the etiology is *norovirus,* namely nursing home; whereas if the location is either a hospital or a picnic, other etiology is indicated. There is no vehicle identified that specifically indicates *norovirus*, but numerous vehicles, such as chicken, tuna and milk, indicate that the etiology of the outbreak is not *norovirus*.

Figure 11 shows the decision tree with attribute selection for *norovirus* related outbreaks. This tree is very complicated and involves several nodes. The leaf nodes with low support are not very attractive for our objective but they cannot be removed because they are the parent nodes for other leaf nodes. It can be noted that the *norovirus* outbreaks are caused by many different combinations of foods and consumption locations. *Chicken salad* is chosen as the root node of this tree which suggests that the highest information gain is provided by this attribute. But unlike all other decision trees where the root rode classifies the positive instances, root node for this decision tree classifies the negative *norovirus* instances. The first three nodes (*Chicken salad*, *pork,* and *picnic*) in fact eliminate the negative instances, which is an interesting finding. In other words, if a person consumed *chicken salad*, *pork* or the consumption location was *picnic*, the outbreak is very unlikely to be caused by *norovirus*. *Turkey sandwich* consumed at *workplace, not cafeteria* caused a very significant number of *norovirus* outbreaks. As was the case for the other three analyses, the decision tree obtained for *norovirus* outbreaks without attribute selection was even more complex than the tree reported in Figure 11, and in this case we were not able to extract any additional information from that tree. It is therefore not included in the paper.

Finally, Figure 12 shows the association rules obtained to indicate *norovirus*. We note that these results indicate a long list of locations (banquet facility, office setting, school, nursing home, wedding reception, church or temple, workplace not cafeteria, and camp) that indicate that the etiology is *norovirus*. Also, there is a similar list of food vehicles (lettuce, salad, green salad, turkey sandwich, ice, submarine sandwich, potato salad, and mixed fruit). This compliments the results of the attribute selection, which consist primarily of vehicles that indicate an etiology other than *norovirus*. Furthermore, two association rules are obtained involving both a location and a food vehicle, namely

- Restaurant or deli & lettuce ⇒ *Norovirus*
- Restaurant or deli & salad ⇒ *Norovirus.*

One location, namely nursing home, is identified by all three methods as being somehow associated with frequent *norovirus* outbreaks. Furthermore, four other locations (banquet facility, wedding reception, workplace (not cafeteria) and camp) are identified by both the decision tree and the association rule mining. Seven food types (lettuce, salad, turkey sandwich, ice, submarine sandwich, potato salad and mixed fruit) are also identified by those two methods as indicating a *norovirus* outbreak. As before, further analysis may thus be warranted for investigating the link between those locations and foods and *norovirus* outbreaks.

4.5 Discussion of Results

The results reported above for four common types of foodborne disease outbreaks illustrates how data mining can find interesting patterns in food safety surveillance databases. The results will, however, always be limited by the quality and availability of data. The CDC database analyzed here has two short text fields, one describing the food vehicle responsible, and the other describing the location where the outbreak occurred. Our approach is thus limited to finding patterns of relatively simple relationship between various vehicles and locations. If a more detailed description of each outbreak was to be made available in the database then we conjecture that the same methodology could find more nuanced patterns involving other characteristics of an outbreak. Since our data mining framework involves text mining of free-form text, this additional data could be a completely open ended description of the outbreak.

It should also be noted that the analysis of each type of outbreak should be interpreted separately as we do in each subsection above, and there is no reason to believe that a pattern obtained for one etiology must be unique for that etiology. This is in fact revealed by our results above. For example, *prison/jail* is classified as having a positive relationship for both *salmonella enteritidis* and *salmonella typhimurium* (see Figure 2 and Figure 4, respectively). Intuitively this situation is not surprising because one than one type of disease outbreak can occur at any given location. From the data mining perspective such scenarios are also not surprising as the negative examples (that is, the set of instances representing 'not *salmonella enteritidis*' or 'not *salmonella typhimurium*') have a great deal of overlap. All that can be inferred is that if the location is prison or jail then and both *salmonella enteritidis* and *salmonella typhimurium* are more likely causes of outbreaks than the average cause, which should indeed be inferred by independently analyzing each of the two etiologies.

When comparing the value of the proposed approach to analyzing each of the four etiologies above, it is noteworthy that for some types of outbreaks very simple trees are obtained. For example, the decision tree in Figure 2 describes only three scenarios of *salmonella enteritidis* outbreaks, whereas the decision tree in Figure 11 describes twenty one scenarios for how *Norovirus* outbreaks occur. This difference in complexity can be explained by the number of ways in which outbreaks occurred in the database, specifically with respect to food vehicle and the outbreak location. Relatively few vehicles and locations point to *salmonella enteritidis* as the cause of the outbreak, whereas many vehicles and locations point to *Norovirus* as the likely cause. Such differences in complexity of the patterns are to be expected, which also implies that the data mining approach may not be equally useful for analyzing all etiologies.

The main objective of this paper was to demonstrate how data mining can be used to extract hidden patterns from the surveillance database of foodborne disease outbreaks. However, observations such as those obtained here for four common types of outbreaks of foodborne illnesses can be very helpful in devising intervention techniques, including safe handling, processing procedures for different foods as well as safe hygiene practices that can be individually formulated for different types of locations where the food is consumed. With the knowledge of the type of outbreak that is most likely to occur, say, at home, the related agencies can plan training techniques targeted to individuals. Similarly, if a certain type of outbreak occurs at hospitals more often and is related to specific foods (or combination of foods, more realistic situation); the hospital staff can be better trained. Same will be true for different food production industries.

5. Conclusions

In this paper, we have introduced a framework for using data mining techniques to discover hidden patterns in the foodborne disease outbreak data from the Center of Disease Control. We demonstrate how data can be preprocessed appropriately to apply data mining techniques and the use of attribute selection, decision trees, and association rule mining to discover patterns in the data. This technique can be used to gain insight into the types of foods, food combinations and consumption locations that are more frequently linked to certain types of foodborne disease outbreaks. The knowledge gained can be used to create modified intervention techniques for different types of foods and disease causing etiologies. This knowledge can be very useful for designing customized food safety training methods for all food safety stakeholders. Such knowledge of interrelationships can also indicate whether specific foods are more prone to contamination at different locations, for example at home, in restaurants, etc.

Also, cross-contamination of food can occur during consumption. Our data mining techniques can be used to discover frequently occurring patterns where multiple foods caused a foodborne disease outbreak. This knowledge can be used to design food safety procedures for consumers for safe food handling practices.

Further work is required to develop robust prediction models that can be used for rapid classification of unknown or unconfirmed foodborne disease outbreak etiologies. The outbreak reporting practices vary for different states in the US as the criteria of each State Health Department for reporting outbreaks to CDC is different. Discovering hidden patterns and comparing outbreaks from different states also needs further investigation to determine the type foods that cause certain outbreaks more frequently in a given state. This can be done by using the same approach as developed in this paper but also including the *State* information. State Health Departments can benefit considerably from this type of information as they can develop strategies for ensuring food safety in their regions.

The CDC database provides critical information about various foodborne disease outbreaks to consumers. Although, the results from applying data mining techniques cannot be better than the data that is available. Further steps can be taken by the CDC to improve the database by recording all parameters for each type of etiology in a consistent manner. But, in this paper we show how data mining techniques can be used to prepare this database for discovering previously unknown patterns and to study interrelationships between different types of foods and other parameters that affect food safety. Knowledge discovered from this approach can be used by various food safety stakeholders such as producers, processors, consumers, policy-makers and regulatory officials for developing food safety measures as they relate to them.

References

Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of the 1994 International Conference on Very Large Data Bases* (VLDB'94), 487-499.

Berry, M.J.A. & Linoff, G.S. (1997). Data Mining Techniques for Marketing, Sales and Customer Support. John Wiley & Sons, Inc., Hoboken, New Jersey.

Burges, C.J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Knowledge Discovery and Data Mining*, 2(2).

CDC (2006). Norovirus Technical Factsheet, http://www.cdc.gov/ncidod/dvrd/revb/gastro/ norovirus-factsheet.htm

Cody, S.H., Glynn, M.K., Farrar, J.A., Cairns, K.L., Griffin, P.M., Kobayashi, J., Fyfe, M., Hoffman, R., King, A.S., Lewis, J.H., Swaminathan, B., Bryant, R.G. & Vugia, D.J. (1999). An outbreak of Escherichia coli O157:H7 infection from unpasteurized commercial apple juice. *Annals of Internal Medicine,* 130(3), 202-209.

Cortes, C. & Vapnik, V. (1995). Support vector networks. *Machine Learning* 20, 273-297.

Dewal, C.S., Hicks, G., Barlow, K., Alderton, L., & Vegosen, L. (2006). Foods Associated with Food-borne Illness Outbreaks from 1990 through 2003. *Food Protection Trends*, 26(7), 466-473.

Domingos, P. & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29, 103–137.

FDA News (2006). FDA Notifies Consumers that Tomatoes in Restaurants Linked to Salmonella Typhimurium Outbreak. http://www.fda.gov/bbs/topics/NEWS/2006/NEW01504.html

Foodborne Illness (2005). Centers for Disease Control and Prevention http://www.cdc.gov/ncidod/dbmd/diseaseinfo/foodborneinfections_g.htm

Gu, Q., Z. Cai, L. Zhu, & Huang, B. (2008). Data Mining on Imbalanced Data Sets. In *2008 International Conference on Advanced Computer Theory and Engineering*, 1020-1024.

Hall, M.A., & Holmes, G. (2003). Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. *IEEE Transactions on Knowledge and Data Engineering*, 15(6), 1437-1447.

Han, J., & Kamber, M. (2006). Data mining: concepts and techniques, 2nd Edition. Morgan Kaufmann Publishers, San Francisco.

Heckerman, D. (1996). Bayesian networks for knowledge discovery. In *Advances in Knowledge Discovery and Data Mining*, 273-305, MIT Press, Cambridge, MA.

Hipp, J., Güntzer, U., & Nakhaeizadeh, G. (2000). Algorithms for association rule mining - a general survey and comparison. *SIGKDD Explorations* 2, 58-64.

Hochberg, A.M., Reisinger, S.J., Pearson, R.K., O'Hara, D.J., & Hall, K. (2007). Using data mining to predict safety actions from FDA Adverse Event Reporting System data. *Drug Information Journal,* 41(5), 633-642.

Japkowicz, N. (2000). Learning from imbalanced data sets: a comparison of various strategies. In *Papers from the AAAI Workshop on learning from imbalanced data sets.*

Kira, K., & Rendel, L. (1992). The Feature Selection Problem: Traditional Methods and a new algorithm. *Proc. Tenth National Conference on Artificial Intelligence*, MIT Press, 129-134.

Kohavi, R., & John, G.H. (1997). Wrappers for feature selection. *Artificial Intelligence* 97(1-2), 273-324.

Levine, W.C., Smart, J.F., Archer, D.L., Bean N.H., & Tauxe, R.V. (1991). Foodborne Disease Outbreaks in Nursing Homes, 1975 through 1987. *The Journal of the American Medical Association,* 266(15), 2105-2109.

Lewis, D. (1992). An evaluation of phrasal and clustered representation on a text categorization task.In *Proceedings of the 15th annual international ACM SIGIR conference on research and development in information retrieval*, 37-50.

Li, X., & Olafsson, S. (2005). Discovering Dispatching Rules using Data Mining. *Journal of Scheduling*, 8(6), 515-527.

Liu, H., & Motoda, H. (1998). Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic, Boston.

Nazeri, Z., Bloedorn, E., & Ostwald, P. (2001). Experiences in mining aviation safety data. In *Proceedings of the 2001 ACM SIGMOD international Conference on Management of Data* (Santa Barbara, California, United States, May 21 - 24, 2001). T. Sellis, Ed. SIGMOD '01. ACM, New York, NY, 562-566.

Olafsson, S., X. Li, & Wu, S. (2008). Operations Research and Data Mining. *European Journal on Operational Research*, 187(3), 1429-1448.

Olsen, S.J., MacKinon, L.C., Goulding, J.S., Bean, N.H., & Slutsker, L. (2000). Surveillance for Foodborne Disease Outbreaks – United States, 1993-1997. *Morbidity and Mortality Weekly Report*, 49(SS01), 1-51.

Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning*. Morgan-Kaufmann, San Mateo, CA.

Resource Center for Biodefense Proteomics Research (2009). Salmonella Typhimurium. http://www.proteomicsresource.org/Resources/viewOrganismSt.aspx

Simoudis E. (1996). Reality Check for Data Mining. *IEEE Expert*, 11(5), 26-33.

Witten I.H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. 2nd Edition. Morgan Kaufmann Publishers, San Francisco.

Figure 1. A framework for discovering patterns in a foodborne illness surveillance database

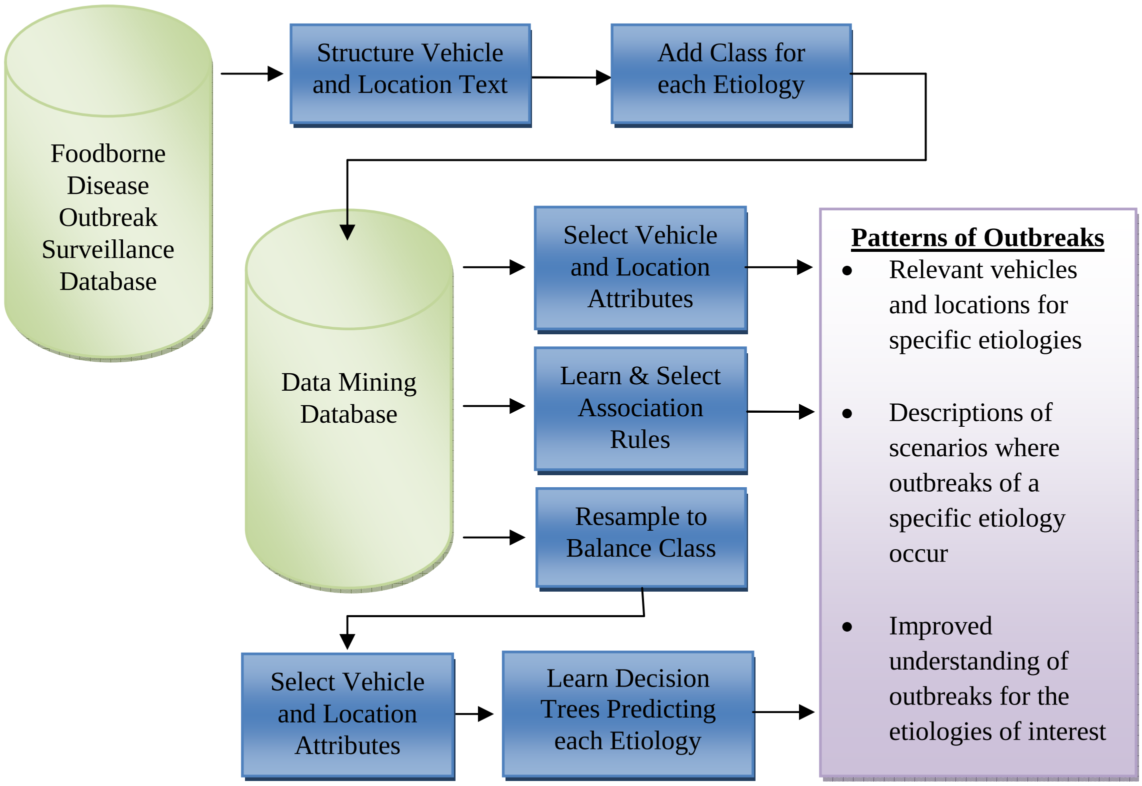
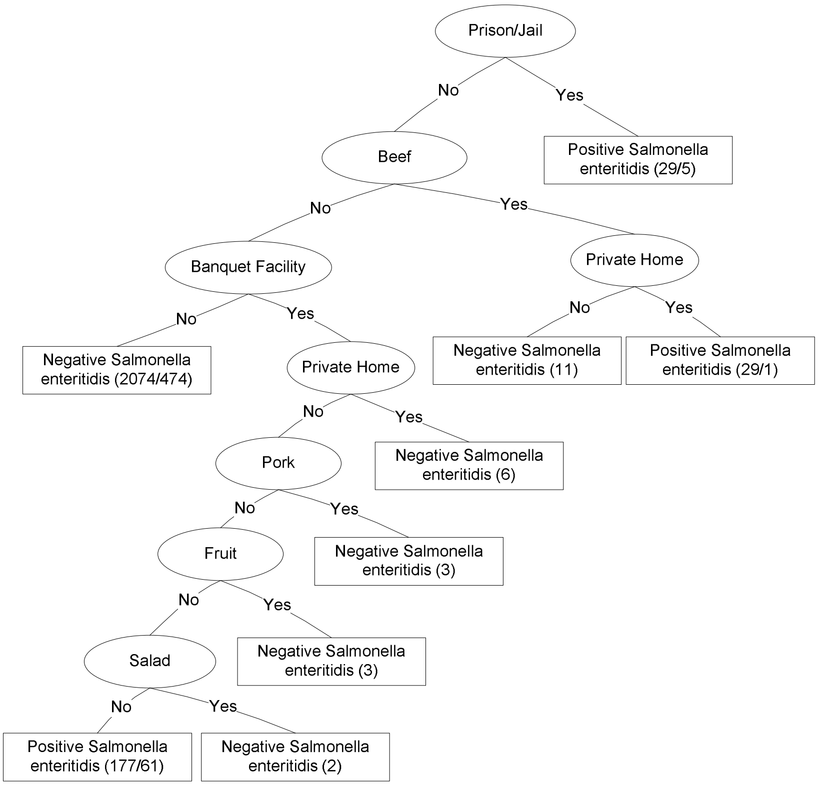
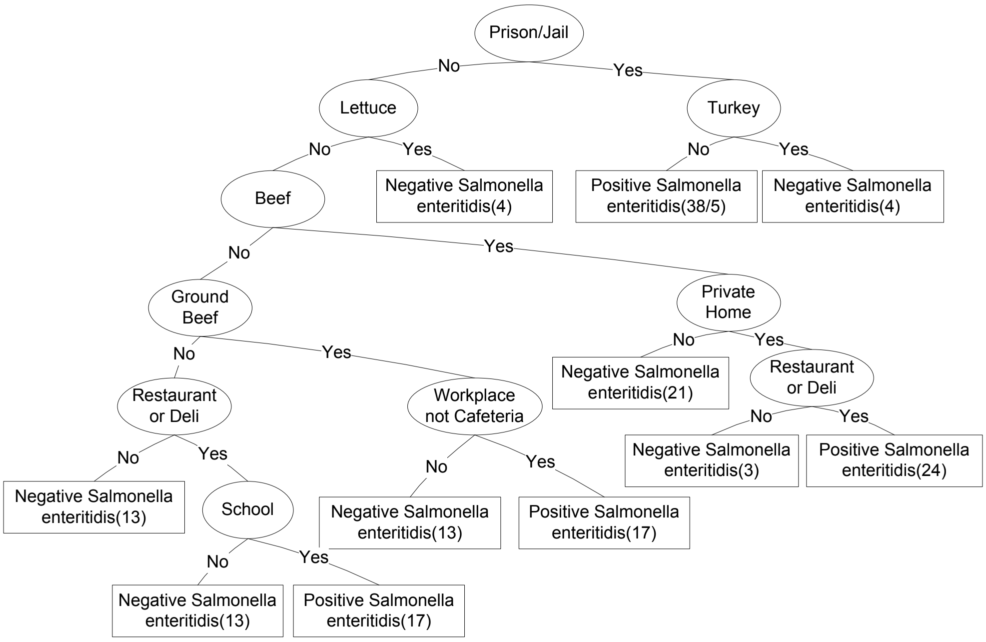


Figure 2. Decision tree for *Salmonella enteritidis* related outbreaks[a]



[a] The numbers in the parenthesis of each leaf represent the associated error. The first number represents the total number of instances classified by that leaf and the second number represents the incorrectly classified instances. For example, 2074/474 represents 1600 (=2074-474) correctly classified instances and 474 incorrectly classified instances.

Figure 3. Decision tree for *Salmonella enteritidis* related outbreaks using all attributes[a]



[a] The numbers in the parenthesis of each leaf represent the associated error. The first number represents the total number of instances classified by that leaf and the second number represents the incorrectly classified instances. For example, 38/5 represents 33 correctly classified instances and 5 incorrectly classified instances.

Figure 4. Associations found for *Salmonella enteritidis* outbreaks

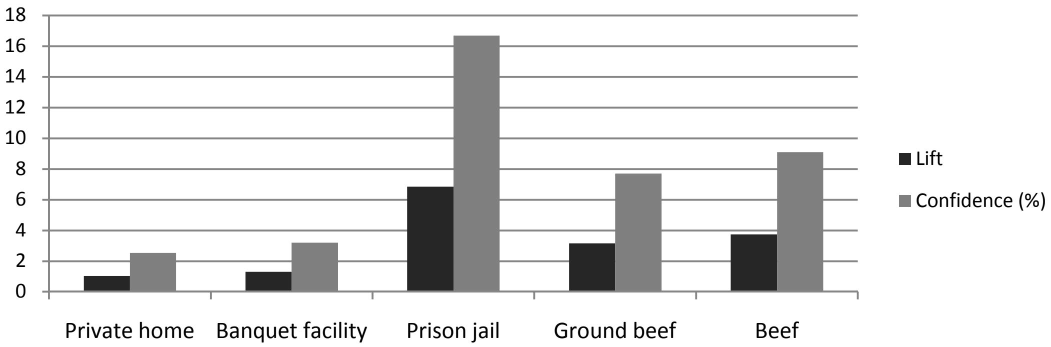
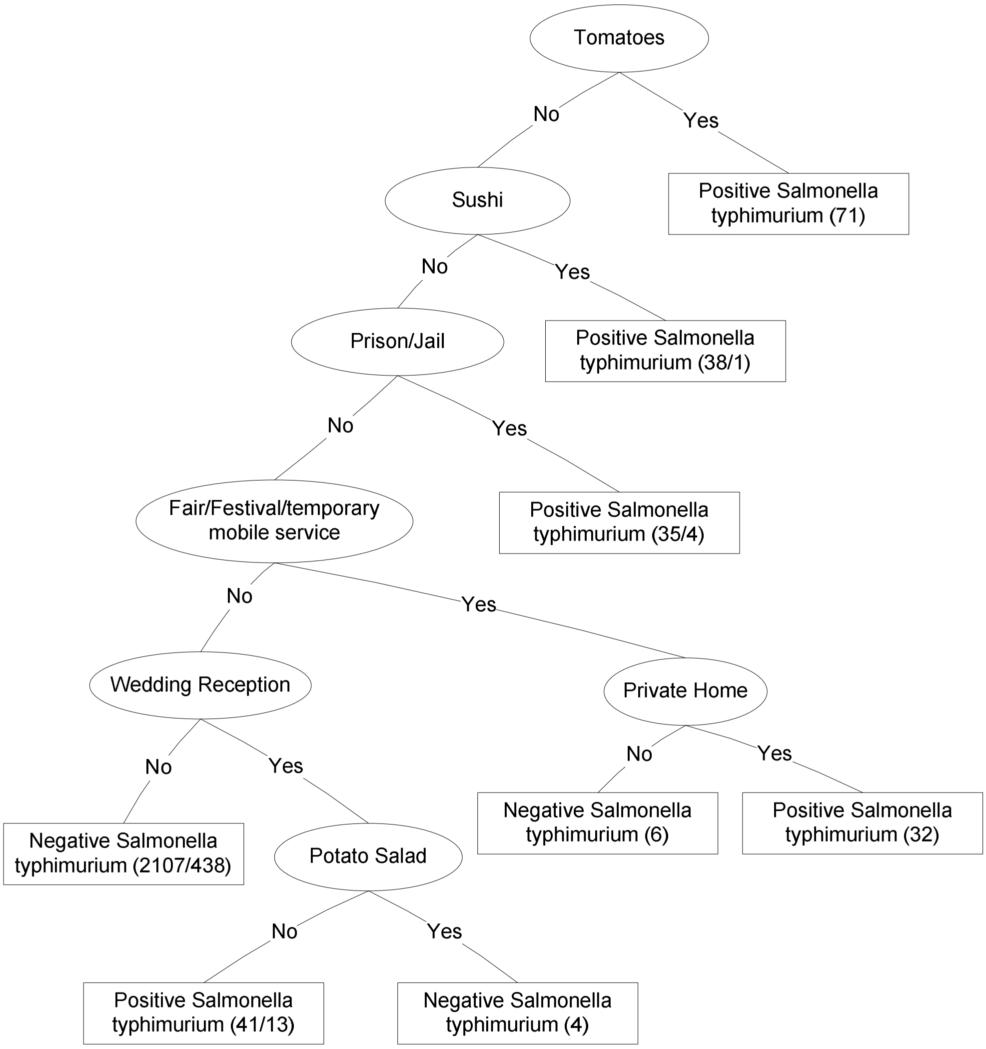


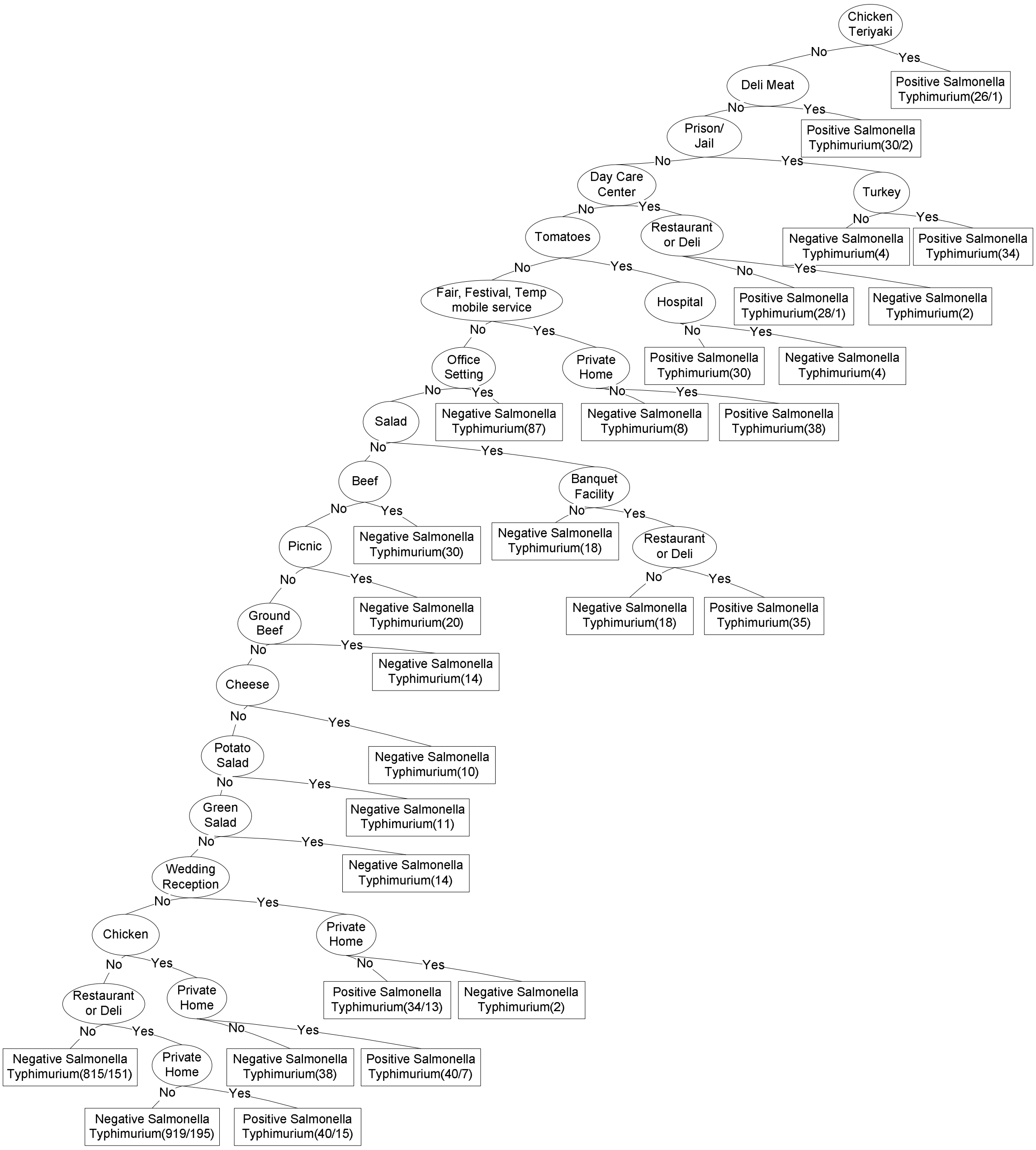
Figure 5. Decision tree for *Salmonella typhimurium* related outbreaks[a]



---

[a] The numbers in the parenthesis of each leaf represent the associated error. The first number represents the total number of instances classified by that leaf and the second number represents the incorrectly classified instances. For example, 2107/438 represents 1669 correctly classified instances and 438 incorrectly classified instances.

Figure 6. Decision tree for *Salmonella typhimurium* related outbreaks using all attributes[a]



[a] The numbers in the parenthesis of each leaf represent the associated error. The first number represents the total number of instances classified by that leaf and the second number represents the incorrectly classified instances. For example, 919/195 represents 724 correctly classified instances and 195 incorrectly classified instances.

Figure 7. Associations found for *Salmonella typhimurium* outbreaks

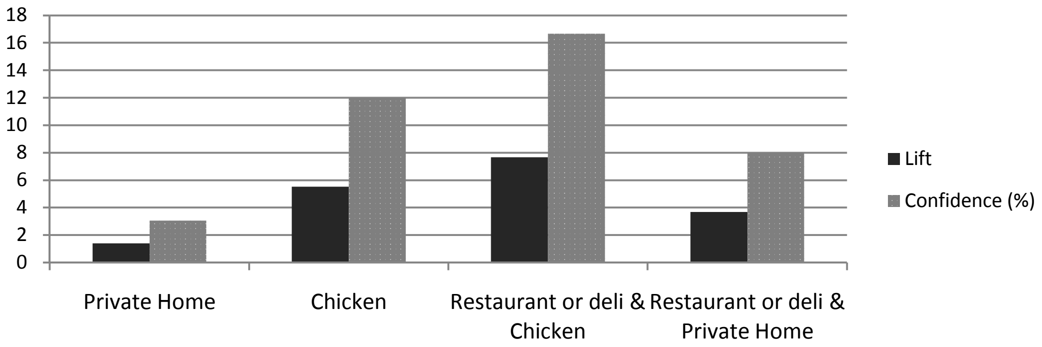
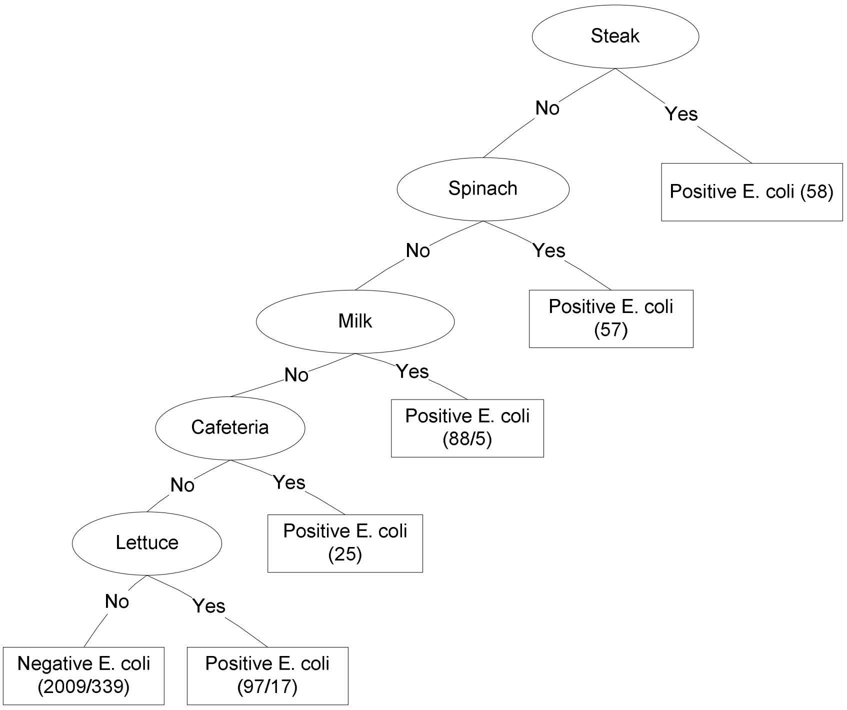
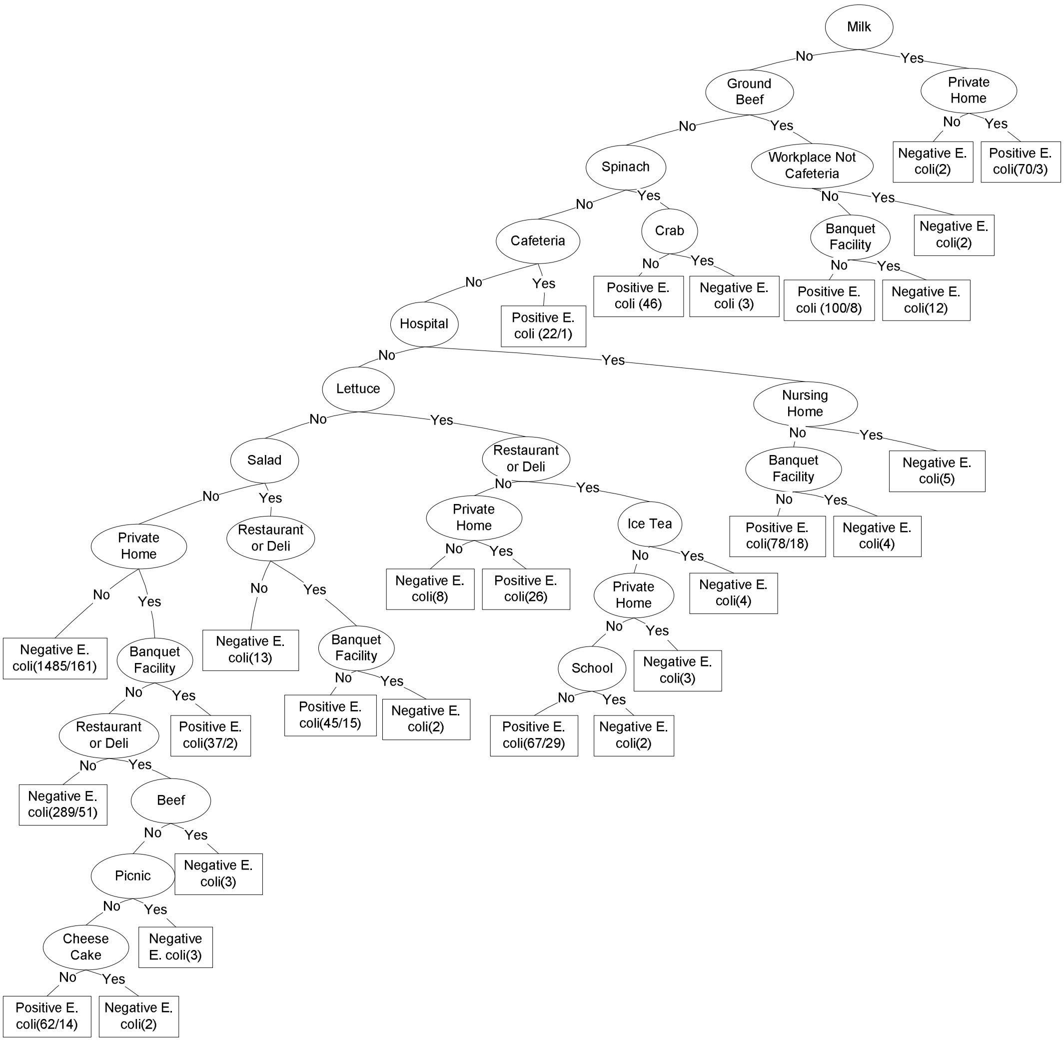


Figure 8. Decision tree for *E. coli* related outbreaks[a]



[a] The numbers in the parenthesis of each leaf represent the associated error. The first number represents the total number of instances classified by that leaf and the second number represents the incorrectly classified instances. For example, 2009/339 represents 1670 correctly classified instances and 339 incorrectly classified instances.

Figure 9. Decision tree for *E. coli* related outbreaks using all attributes[a]



[a] The numbers in the parenthesis of each leaf represent the associated error. The first number represents the total number of instances classified by that leaf and the second number represents the incorrectly classified instances. For example, 62/14 represents 48 correctly classified instances and 14 incorrectly classified instances.

Figure 10. Associations for *E. coli* outbreaks

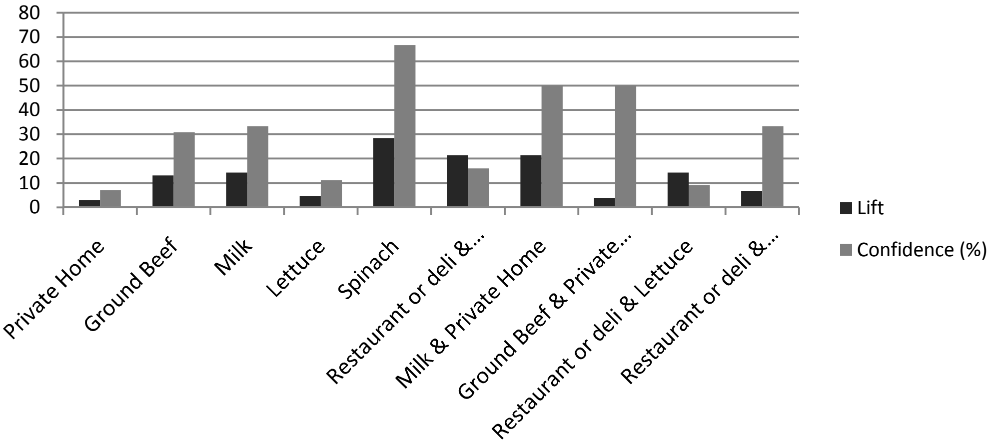
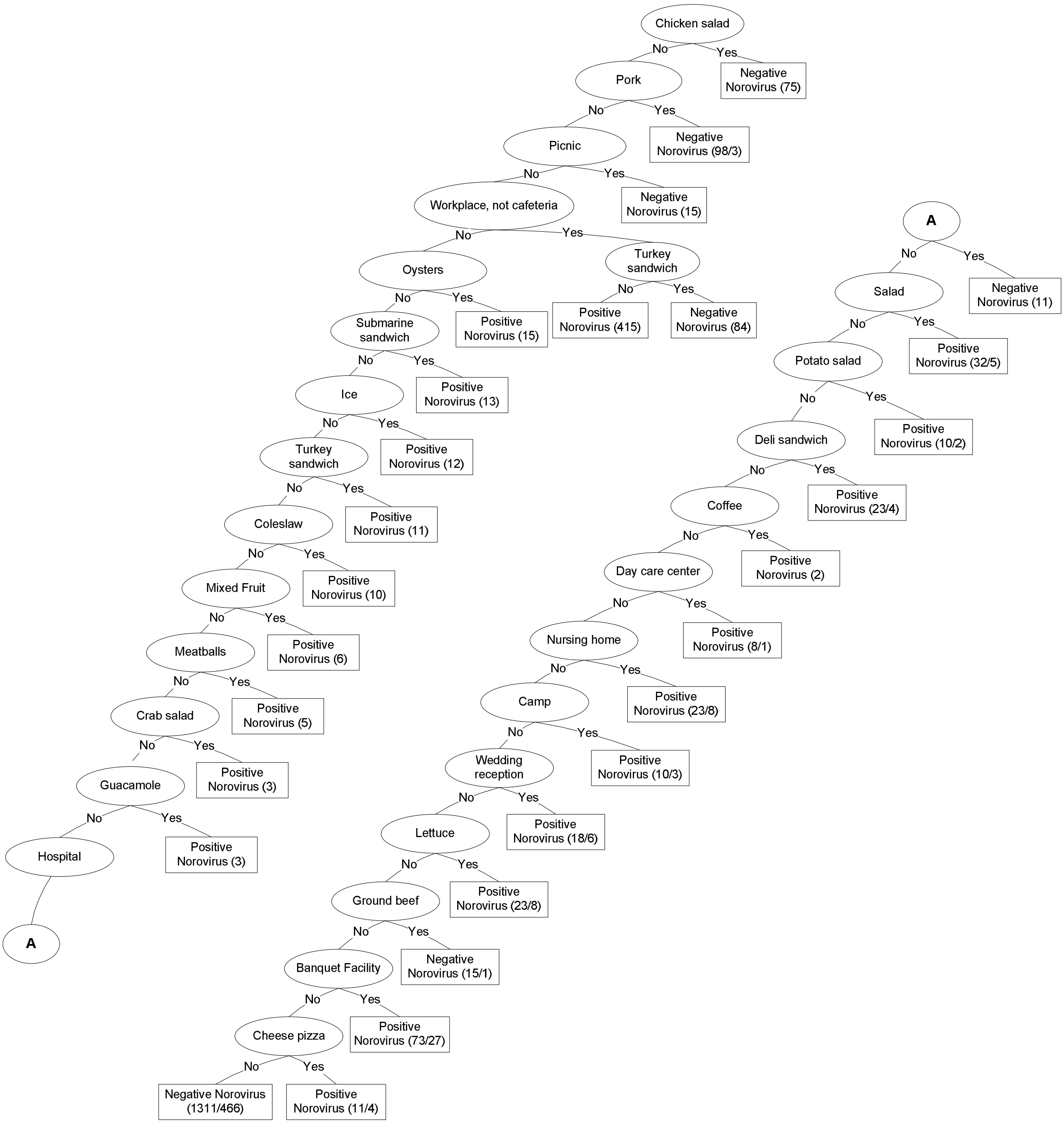


Figure 11. Decision tree for *Norovirus* related outbreaks[a]



[a] The numbers in the parenthesis of each leaf represent the associated error. The first number represents the total number of instances classified by that leaf and the second number represents the incorrectly classified instances. For example, 1311/466 represents 845 correctly classified instances and 466 incorrectly classified instances.

Figure 12. Associations found for *Norovirus* outbreaks

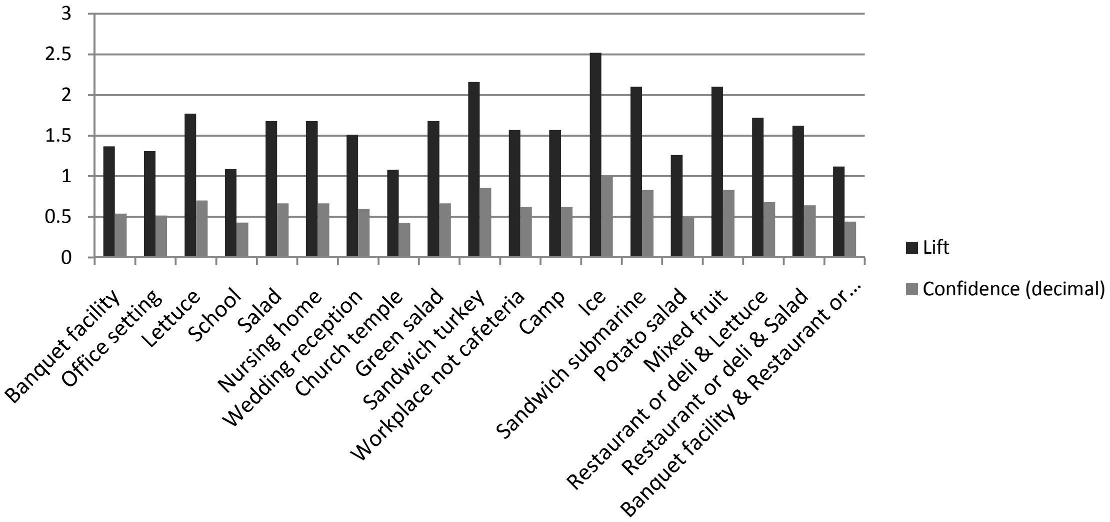


Table 1. Summary of Foodborne Illness Outbreaks, 2006

Confirmed Etiology	**No. Outbreaks**	**No. Cases**
Bacterial	223	5,336
Chemical	53	221
Parasitic	9	129
Viral	337	11,122
Suspect Etiology	**No. Outbreaks**	**No. Cases**
Bacterial	75	1,440
Chemical	11	39
Parasitic	3	18
Viral	165	2,841
Multiple Etiology	**No. Outbreaks**	**No. Cases**
Confirmed	1	96
Suspect	20	254
Confirmed and Suspected	1	32

Table 2. Attribute summary of the original dataset

Attribute	Type	Description
Confirmed Etiology	Nominal	Cause of outbreak, e.g. - E. Coli
State	Nominal	State where the outbreak occurred
Month	Nominal	Month when the outbreak occurred
Illnesses	Numeric	Number of illnesses reported
Hospitalizations	Numeric	Number of hospitalizations reported
Deaths	Numeric	Number of deaths reported
Vehicle	Text	Food item/s that caused the outbreak
Location	Text	Location where food was consumed, e.g. - restaurant

Table 3. Most predictive attributes for classifying *Salmonella enteritidis* outbreaks

	Indicates *salmonella enteritidis*	**Indicates another etiology**
Location	Private home (1) Banquet facility (3)	Office (2) Workplace, not cafeteria (4) School (5) Church or temple (8) Restaurant or deli (10)
Vehicle		Lettuce (6) Chicken (7) Salad (9)

Table 4. Most predictive attributes for classifying *Salmonella typhimurium* outbreaks

	Indicates *Salmonella typhimurium*	**Indicates another etiology**
Location	Restaurant or deli (1) Private home (4)	Banquet facility (2) Office setting (5) School (6) Workplace, not cafeteria (7) Church or temple (9) Nursing home (10)
Vehicle	Chicken (3)	Lettuce (8)

Table 5. Most predictive attributes for classifying *E. coli* outbreaks

	Indicates *e. coli*	**Indicates another etiology**
Location	Restaurant or deli (1) Private home (2)	Banquet facility (4) Office setting (5) School (7) Workplace, not cafeteria (8) Church or temple (10)
Vehicle	Lettuce (3) Milk (6)	Chicken (9)

Table 6. Most predictive attributes for classifying *Norovirus* outbreaks

	Indicates *norovirus*	**Indicates another etiology**
Location	Nursing home (6)	Hospital (7) Picnic (8)
Vehicle		Chicken (1) Tuna (2) Milk (3) Fish, escolar (4) Pork (5) Fish, mahi mahi (7) Turkey (10)

Modeling traceability information in soybean value chains

Published in *Journal of Food Engineering (2010),* 99(1):98-105.

Maitri Thakur [1, 2, *] and Kathryn A-M Donnelly[3]

[1]Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011
[2]Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA 50011
[3]Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima), Muninbakken 9-13, Breivika, 9291 Tromsø, Norway
* Corresponding author

Abstract

Identification of the information to be recorded is the most important requirement for developing an effective traceability system. In this paper, we present a soybean value chain and model the information capture by three links in the chain including the farming, bulk handling and processing sectors. Internal information capture points were identified for each sector and the corresponding traceability information to be recorded was determined. In-depth analyses were conducted for a soybean elevator and an oil and meal processor to determine the importance of traceability information from their perspective. A lot of information is available at different links in the soybean value chain. The method presented here can be used to create a standardized list of data elements that need to be recorded internally or exchanged with other links in the chain. A UML class diagram is developed to represent a method for modeling the product, process, quality and transformation information at any link in the chain. Finally, some suitable technologies for electronic information exchange within the food supply chains are presented.

Keywords: soybean value chain; traceability; information modeling; information exchange; soybean oil; elevator; processor

Introduction

Consumers all over the world have faced various food safety and health issues in the recent years. This has led to a growing interest in developing systems for food supply chain traceability (Carriquiry and Babcock, 2007; Folinas et al., 2006; Jansen-Vullers et al., 2003; Madec et al., 2001; McKean, 2001). Various food safety and traceability guidelines and regulations exist in several countries. Under the European Union Law, ‘‘traceability” is defined as the ability to track any food, feed, food-producing animal or substance that will be used for consumption, through all the stages of production, processing and distribution (Official Journal of the European Communities, 2002). It is a risk-management tool that allows food business operators or authorities to withdraw or recall products which have been identified as unsafe. In the United States, the Bioterrorism Act of 2002 requires that all companies involved in the food and feed industry to self-register with the Food and Drug Administration and maintain records and information for food traceability purposes (Food and Drug Administration, 2002). In Canada, Can-Trace was launched in July 2003 which is a collaborative and open initiative committed to the development of traceability standards for all food products sold in Canada (Can-Trace, 2003).

The General Food Law (Official Journal of the European Communities, 2002) requires traceability throughout the food supply chain. In order to be able to track and trace products throughout the supply chain, food business operators must maintain relevant information from the suppliers and keep track of all products and their transformation through all stages of production and then pass this information to the next link in the supply chain (Donnelly et al., 2009; Schwägele, 2005; Thakur and Hurburgh, 2009). Senneset et al. (2007) states that in order to achieve chain traceability, the identities of traceable units must be recorded at reception and shipping. The paper also states that internal traceability requires recording of all transformations during the production process.

One of the biggest challenges with supply chain traceability is the exchange of information in a standardized format between various links in the chain. Globalization combined with the ever-increasing complexity of food supply chain networks has led to an increase in the significance of efficient systems for information exchange between food businesses. This information needs to be exchanged in a precise, effective and electronic manner (FSA, 2002; Moe, 1998). To facilitate electronic interchange of such product information, international, non-proprietary standards are required such as the ones highlighted by

Jansen-Vullers et al. (2003). Standards must describe how information can be constructed, sent and received and also how the data elements in the information should be identified, measured, interpreted and stored (Folinas et al., 2006). Previous studies have shown that there is currently no standardized way of formatting information for exchange in traceability systems. Research suggested that structured data lists, vocabularies ad ontology will be appropriate tools in achieving effective universal data exchange (Donnelly et al. 2009, Dreyer et al., 2004; TRACE 2, 2008). Individual companies have made great progress in proprietary technologies for automated data capture and electronic data coding. However the benefit of these is lost when the data element transmission is required for use outside the originating company as it is only effective when there is an identical software system at the receiving end (Donnelly, 2008).

The TraceFood Framework developed under the European Commission sponsored TRACE project provides a toolbox with principles and guidelines for how to implement electronic chain traceability. The framework consists of the following components (TraceFood Wiki, 2009):

(g) Principle of unique identifications
(h) Documentation for joining and splitting (transformations) of units
(i) Generic language for electronic exchange of information
(j) Sector-specific language for electronic information exchange
(k) Generic guidelines for implementation of traceability
(l) Sector-specific guidelines for implementation of traceability

Based on this framework, the implementation of chain traceability requires industry analysis to understand the material flow, information flow and information handling practices. Using this method, based on the industry analysis, recommendations can be provided for new sector-specific data terminology and what information needs to be recorded by each link and communicated to other links in the chain.

To enable effective, electronic information exchange, work needs to be carried out on a sector-specific level. Analysis of what product information the particular food sector already records should be carried out and a method and format for identifying this product information should be developed in a standard form (Donnelly, 2009). The need for such systems has already been identified throughout the food industry, but particularly in areas where the authenticity of a product is in question. The viability of such non-proprietary standards were shown in the TraceFish project (CEN 14659, 2003; CEN 14660, 2003; Denton, 2003) where both sector-specific standards (for captured fish and farmed fish) and generic standards (for electronic coding and request-response scheme) were developed. The TraceFish work established sector-specific data models that not only contain information about data elements (including the relationship between them) relevant for product information in one link of the supply chain, but also information for each link. Standardized lists for data elements which can be included in data models have been acknowledged as a key technology for resolving semantic heterogeneity and are important in knowledge management in large organizations (FAO AGROVOC, 2006; Haverkort, 2007; Haverkort, 2006; Stuckenschmidt, 2003).

In this paper, we present a bulk product supply chain. A soybean value chain is presented and the work has been inspired by the TraceFood Framework, TraceFish project as well as the study carried out in the chicken processing sector by Donnelly et al. (2009). Bulk products supply chains present additional complexities in terms of defining the traceable units. In addition, blending and splitting of individual batches complicates how information is tied to a specific entity (traceable unit) (Thakur and Hurburgh, 2009). We present a soybean value chain with soybean oil used for cooking as the end product. The objective of this paper is to present a model for information capture at various stages in the soybean chain. We specify the information that must be recorded by three links in the soybean value chain; by the farmer, by an elevator handling bulk soybeans and by the soybean oil and meal processor. In-depth analysis of a soybean elevator based in US and a soybean processor in Europe was also conducted. The results related to the importance of different product, process, and quality information from their perspective are also presented.

Soybean chain stakeholders

Soybeans are native to East Asia and today are cultivated around most of the Americas and East Asia. A small amount of cultivation takes place Eastern Europe. Europe however is a consumer of soybeans imported from the Americas for the production of both animal feed and products for human consumption. Soybeans in the USA are primarily grown in the northern Midwestern states from Ohio to Kansas and South Dakota, in the states along Mississippi river, and in the southeastern states. After harvest, the farmers sell their crop to the grain elevators that handle and sell soybeans marketed against generic grade standards. Soybeans are transported by truck, rail, barge or ship to the processors. Beans are loaded, unloaded, conveyed, and blended several times while on the way from the field to processors. Bulk handling is most common in the soybean value chain. Soybeans on average contain 11% moisture, 37.9%

protein, 17.8% fat, 4.7% fiber, and 4.5% ash. The most common end use of soybeans include soybean oil that is used for cooking and soybean meal used as animal feed.

Three soybean chain stakeholders are presented in this paper; farmer, elevator and processor. Figure 1 shows a simple flowchart of the soybean value chain and the main inputs and outputs at each stage.

Farmer

The farmer is the first link in the soybean value chain. Farmers purchase seeds from a seed company and sell their crop to an elevator after harvesting. Several chemical compounds including fungicides and herbicides are used for soybean seed treatment to inhibit damage to the crop. Combines are commonly used for harvesting the soybean crop. After harvest, soybeans can be stored on farm before selling to an elevator.

The data available at the farming stage includes the information related to the seed supplier, seed variety, geographical location, farming practices, pesticides/ herbicides applications, harvest time, on-farm storage duration, and selling date.

Elevator

An elevator is a very important link between the farmer and the processor. Elevators buy soybeans from the farmers, keep it in storage, and blend it before selling to the processors. Soybeans received at the elevator are sampled and graded based on moisture content, test weight, foreign material and damaged material. The farmers are paid according to the quality grade. The beans are then conveyed to the storage silos before shipping to the customers. One storage silo can contain soybeans from several farmers. The incoming lots from the farmers are blended before shipment in order to meet the buyer's quality specifications. Thus, a specific lot shipped to the processor can contain soybeans from all different sources that may end up in the finished product.

The data available at this stage includes the information related to incoming product deliveries from the farmers (quality and quantity), farmer identification, time of delivery, product transformations (mixing and splitting of lots) within the elevator, product blending for shipments, and shipment date.

Processor

The processor link presented in this paper corresponds to a soybean oil and meal processor. Soybean oil and meal are the products of soybean processing using solvent extraction method. The soybean oil is used for human consumption while meal is used for animal consumption in the form of animal feed. Soybeans generally arrive at the processing plant by railcars from the elevators. The soybeans received by the processor are sampled and analyzed for moisture, test weight, foreign material and damaged material as done at the elevator and they are stored in silos until the facility is ready to process. Before processing, the soybeans are cleaned to remove any foreign materials and loose hulls. Figure 2 shows a flowchart of soybean oil and meal processing using solvent extraction.

The data available at this stage includes information related to incoming deliveries from the elevator, production information, batch transformations, quality data at different production stages, information related to the solvent used, and final product information.

Consumers

Soybean oil is used for human consumption while soybean meal is used to manufacture animal feed. Refined soybean oil products include cooking oils, margarine, mayonnaise, salad dressings, spreads, vegetable shortenings, etc. Soybean meal is used as animal feed for poultry feeds, swine feeds, fish feeds, pet foods, etc.

Methodology

A basic requirement for designing an effective traceability system is to determine the information that needs to be traced (Regattieri, et al., 2007). Conceptual process flow diagrams were created for farming, handling and processing sectors in the soybean value chain. Information capture points were identified for each sector and the corresponding product, process, and quality information to be captured was determined. In-depth analyses were conducted for a soybean elevator and a processor to determine the importance of traceability information from their perspective. The method used to investigate the importance of traceability information was devised during the creation of the TraceFish standards (CEN 14659, 2003; CEN 14660, 2003; Denton, 2003) and the mineral water initial standard (Karlsen, et al., 2008). A questionnaire was developed in order to gather information about what data elements are important where a list of possible data elements for each link was created using published sources. The data elements on the questionnaire corresponded to the product, process, transformations and quality information. The information was collected from an elevator in US and an oil and meal processor in Europe. Table 1 shows a list of questions that were asked on the questionnaire.

Results

Information modeling

There are three categories of information that needs to be captured by each entity; the product information, process information, and quality information. The information capture methods can be different for each entity in the chain. Figure 3 shows a detailed process flow model for all three sectors. The inputs and outputs of each process are also shown. The information capture points at each stage in the supply chain are numbered. These numbers represent the points where specific information must be captured. Table 2 and 3 show the product, process and quality related information that must be captured at these points. The location of information capture points were identified based on the responses from the soybean chain stakeholders.

Linking traceability information to Traceable Units

The concept of a traceable resource unit (TRU) was first introduced by Kim et al. (1999) where a TRU was defined as a batch of any resource. A Traceable Unit (TU) can be defined as any item upon which there is a need to retrieve predefined information and that may be priced, or ordered, or invoiced at any point in a supply chain. In practice, it refers to the smallest unit that is exchanged between two parties in the supply chain (TraceFood Wiki, 2009). Each traceable unit must be uniquely identified as described previously by the TraceFood framework. In order to capture and retrieve traceability information when required, this information must be associated with a uniquely identified TU. The definition of a TU would be different for each link in the soybean value chain. For example, at an elevator a truckload of soybean delivery is a TU while for a processor, a production batch is a TU. Table 4 lists the TUs as identified at each stage of in the chain. The logistic unit referred to in Tables 2, 3 and 4 is defined as an item that is established for transportation and/or storage which needs to be managed through the supply chain (for example, a 100 lb bas of soybean seeds).

One of the challenges related to bulk product traceability is the concept of transformations. Since, different lots are mixed and split at different stages of production, it is necessary to keep track of all these transformations as well as linking them to the new traceable units created (Thakur and Hurburgh, 2009). These traceable units must also be uniquely identified and linked to the original TUs that created them.

Figure 4 shows a UML class diagram for internal information capture by any link in the value chain. UML class diagrams are used for object-oriented analysis and design. They represent the classes of the system, their interrelationships and the operations and attributes of the classes (Ambler, 2008). The class diagram consists of the following main components: (1) Classes, that represents any person, place, thing, concept or an event, (2) Associations, that represents how objects are associated with (or related to) other objects. Classes are modeled as rectangles with three sections. The top section is for the name of the class, the middle section for the attributes of the class and the bottom section for the methods of the class. Attributes are the information stored about an object while the methods are the things an object or class does. The association between objects is depicted by a line connecting two classes which also identifies the multiplicity of an association.

For the sake of simplicity, only the basic structure of the UML class diagram for internal information capture is shown in Figure 4. The diagram shows the classes, their attributes and associations with other classes. All product, process, and quality properties must be linked to a uniquely identified TU. Each traceable unit can have several properties (product parameters, quality information, etc.) associated with it. On the other hand, each TU can have several transformations. One TU (for example, a truckload of soybeans from the farmer) can be split into different parts and transferred to different storage silos at the elevator where this one TU is mixed with other units already present in that silo. Therefore, each TU can have several transformations, each of which must be uniquely identified and linked back to the original TU. Finally, each transformation would generate new TU(s) which must be assigned unique identification by the system. This simple model represents how to model product, process, quality as well as transformation information internally.

Case Studies

Detailed analyses of an elevator and a processor were conducted to determine the importance of product information for these two stakeholders. The following section presents the findings of this analysis. An important observation was made from this analysis. All the information that is being recorded internally by each link corresponds to the information that is communicated to another link. Also, some data elements are reported to be somewhat important but no information is captured because it is not communicated to the next link in the chain. The soybean processor reported that some of the important parameters related to the solvent used for extraction of oil are not recorded by them but communicated by the suppliers are they rely on the information provided to them. These include the normal hexane level, sulphur content and benzene content of the solvent used. This information is provided by the solvent supplier.

Elevator

Figure 5 presents the level of importance of soybean product properties for the elevator. As described in the questionnaire in the methodology section, the level of importance is based on a scale of 1-

5, 1 being unimportant and 5 being very important. The graph shows that information including moisture, foreign material, damaged material, heat damaged and total damaged material is the most important for the elevator. This finding was expected as soybeans are traded based on generic grade standards and grade is determined on the basis of this quality information. However, the test weight is not considered important by the elevator which was unexpected as it is one of the factors that are used to determine the soybean grade standard. Information related to all data elements except mycotoxins is recorded by the elevator. Mycotoxins are toxic chemical products produced by fungal infection of crops. Soybeans in general are low in mycotoxins. However, contamination of mycotoxins in soybean meal is highly dependent upon the level of soy hulls, because hulls are more concentrated with mycotoxins (Agriculture Business Week, 2009). Soybean meal accounts for a large proportion of animal feed. Thus, level of mycotoxins in soybeans would be very important in case of contamination. This information must be recorded by the elevator so that it is available in case of a food-related emergency.

Processor

Figure 6 presents the importance of crude oil properties for the processor. According to the American Oil Chemists Society, the factors that affect crude soybean oil quality are: total gums/phosphatides, free fatty acids, iron/metal content, nonhydratable phosphatides, oxidation products, and pigments (Debruyne, 2004). Our findings do not match the AOCS criterion. The processor reported that the information related to total gums/phosphatides and free fatty acids is very important and is captured internally. Also, the information related to nonhydratable phosphatides and pigments is important and is captured internally by the processor. It is interesting to note that while the processor indicated the information related to iron/metal content and oxidation products is somewhat important, yet this information is not recorded by them. The importance of other crude oil properties is also summarized in Figure 5.

Figure 7 presents the importance of soybean meal properties for the processor. The processor reported that information related to moisture content, protein, oil, protein digestibility index and urease activity is very important. This was expected as soybean meal is used to manufacture animal feed and all these factors determine the quality of the feed (Thakur and Hurburgh, 2007). However, it was interesting to note that trypsin inhibitor activity and ash content were reported as somewhat important but this information is not captured by the processor.

Technologies for information exchange

Electronic Data Interchange (EDI) is commonly used in the B2B (Business-to-Business) environment as a reliable mode for electronic data exchange between business and trading partners. EDI is a set of standards for structuring information that is to be electronically exchanged between and within business organizations and other groups. EDI implies a sequence of messages between two parties, either of whom may serve as originator or recipient. The effectiveness of using EDI has been widely investigated and it is evident that the standard can be used efficiently by organizations with mature IT capabilities. This is generally not the case for all actors in the supply chain (Bechini et al., 2008). On the other hand, the increasing popularity of XML (Extensible Markup Language) for information interchange has made it easy for businesses of any size to use this technology. The main purpose of XML is to facilitate the sharing of structured data across different information systems, particularly via the internet. Both EDI and XML formats are structured to describe the data they contain. The main difference is that the EDI structure has a record-field-like layout of data segments and elements; which makes the EDI file shorter, but not easily understandable. The XML format has tags, which are more easily understood, but make the file bigger and verbose (Electronic Data Interchange Development, 2008).

A lot of information is available at different links in the soybean supply chain. All the information recorded by a given link corresponds to the information that is communicated to the next link in the supply chain. The method used in this paper can be used to create a standardized list of data elements that need to be recorded internally or exchanged with other links in the soybean value chain. Figure 4 presented a UML class diagram for capturing internal traceability information linked a unique traceable unit. The traceable unit is represented as a class in the UML class diagram and all attributes related to the traceable unit are the data elements that need to be recoded internally. Each link in the soybean value chain must develop such models for capturing internal information before it can be exchanged with other links. All data elements must be recorded in a standardized format by all chain links. The information gathered could form the basis for standardized electronic interchange in the supply chain, for instance as an extension of the Universal Business Language (UBL). UBL is a library of standard electronic XML business documents such as purchase orders and invoices developed and supported by Organization for the Advancement of Structured Information Standards (OASIS) and already supported by many national governments, in particular by Denmark and Iceland. TraceCore eXtensible Markup Language (TCX) developed under the TraceFood project is a standard way of exchanging traceability information electronically in the food industry. TCX makes it possible to exchange the information that is common for all food products, like the identifying number, the origin, how and when it was processed, transported and received, the joining and splitting of units, etc. (TraceFood, 2007). The TraceCore XML standards can be adapted to soybean value chain where all actors can exchange information using this standard.

Conclusions

Development of data management systems to facilitate product traceability in food supply chains has gained significant importance in the past years. The ability to track and trace individual product units depends on an efficient supply chain traceability system which in turn depends on both internal data management systems and information exchange between supply chain actors. To enable effective, electronic information exchange, work needs to be carried out on a sector-specific level. Standardized lists for data elements which can be included in data models have been acknowledged as a key technology for resolving semantic heterogeneity and are important in knowledge management in large organizations.

We present a soybean value chain with soybean oil used for cooking as the end product and a model for information capture at various stages in the soybean chain including three links: the farmer, the elevator and the soybean oil and meal processor. Detailed analysis of a soybean elevator based in US and a soybean processor in Europe highlighting the importance of various quality parameters of soybean product from their perspective are also presented. Internal data capture points were identified for each of these links. Traceable Units were defined for each stage in the supply chain and the traceability data that needs to be captured at each point linked to a TU was identified. The traceability data consists of product, process and quality data that must be recorded by each link in the chain.

One of the most interesting findings was that only the information that is communicated to the next link in the chain is recorded internally by both the elevator and the processor. Another interesting finding was that some data elements were reported as being “somewhat important” by both but no information related to these was recorded. On further investigation, it was found that the soybean processor relies on the information provided by the supplier and this information is not recorded again during processing. In this scheme each actor is responsible for maintaining and communicating their own product, process and transformation information. Soybean meal accounts for a large proportion of animal feed and the level of mycotoxins in soybeans is very important in case of contamination. This information, however, is not recorded by the elevator. In addition, the level of mycotoxins was reported as being “unimportant” by the elevator; which was an unexpected finding. This information must be recorded by the elevator so it is available in case of a food safety emergency and the source of the problem can be tracked.

A lot of information is available at different links in the soybean value chain. The method used in this paper can be used to create a standardized list of data elements that need to be recorded internally or exchanged with other links in the soybean value chain. A UML class diagram was developed to represent a method for modeling the product, process, quality as well as transformation information by any link in the value chain. All the traceability data captured must be linked to a uniquely identified TU.

Acknowledgements

The authors would like to thank the industry participants for their contributions to this work. The authors offer a special thanks to Dr. Charles R. Hurburgh at Iowa State University for his valuable insight into the soybean industry. Petter Olsen at the Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima) should also be thanked for his support. The authors would like to thank both the EU project TRACE and Nofima for making this work possible.

References

Agriculture Business Week, 2009. How real is the threat of mycotoxins for feed and animal producers in Asia? <http://www.agribusinessweek.com/how-real-is-the-threat-of-mycotoxins-for-feed-and-animal-producers-in-asia-part-1/>

Ambler, S.W., 2008. The Object Primer 3rd edition: Agile Model-Driven Development with UML 2.0, Cambridge University Press, New York pp 351-380.

Bechini, A., Cimino, M.G.C.A., Marcelloni, F., Tomasi, A., 2008. Patterns and technologies for enabling supply chain traceability through collaborative information e-business. Information and Software Technology 50, 342–359.

Can-Trace, 2003. Agriculture and Agri-Food Canada. <http://www.can-trace.org>

Carriquiry, M., Babcock, B.A., 2007. Reputations, market structure and the choice of quality assurance systems in the food industry. American Journal of Agricultural Economics 89, 12–23.

CEN14659, 2003. CEN Workshop Agreement. Traceability of Fishery products. Specification of the information to be recorded in caught fish distribution chains. European Committee for Standardization.

CEN14660, 2003. CEN Workshop Agreement. Traceability of Fishery products. Specification of the information to be recorded in farmed fish distribution chains. European Committee for standardization.

Debruyne, 2004. Soybean Oil Processing; Quality Criteria and Flavor Reversion, IUPAC-AOCS Workshop on Fats, Oils, and Oilseeds Analysis and Production, Tunis, Tunisia.

Denton, W., 2003. TraceFish: The development of a traceability scheme for the fish industry. In: Luten, J.O., Olafsdottir, G. (eds.) Quality of fish from catch to consumer, pp. 75–91. Wageningen Academic Publishers, Wageningen.

Donnelly, K.A.-M., Karlsen, K.M., Olsen, P., van der Roest, J., 2008. Creating Standardized Data Lists for Traceability – A Study of Honey Processing. International Journal of Metadata, Semantics and Ontologies, 3(4), 283-291.

Donnelly, K.A.-M., van der Roest, J., Höskuldsson, S.T., Olsen, P., Karlsen, K.M., 2009. Improving Information Exchange in the Chicken Processing Sector using Standardised Data Lists. Communications in Computer and Information Science, 46, 312-321.

Donnelly, K.A.-M., Karlsen, K.M., Olsen, P., 2009. The importance of transformations for traceability - A case study of lamb and lamb products. Meat Science, 83, 68-73.

Dreyer, C., Wahl, R., Storøy, J. and Forås, O.P., 2004. 'Traceability standards and supply chain relationships', in Aronsson, H. (Ed.): *Proceedings of the 16th Annual Conference for Nordic Researchers in Logistics, NOFOMA 2004*, Challenging Boundaries with Logistics, Linköping, Sweden, pp.155–170.

EDI vs. XML, 2008. Electronic Data Interchange Development. <http://www.edidev.com/XMLvsEDI.html>

FAO AGROVOC, 2006. < www.fao.org/agrovoc>

Folinas, D., Manikas, I., Manos, B., 2006. Traceability data management for food chains. British Food Journal 108 (8), 622–633.

FSA, 2002. Traceability in the Food Chain- A Preliminary Study, FSA, Editor.

Haverkort, A., 2007. The Canon of Potato Science: 36. Potato Ontology. Potato Research 50(3), 357–361.

Haverkort, A., Top, J., Verdenius, F., 2006. Organizing Data in Arable Farming: Towards an Ontology of Processing Potato. Potato Research 49(3), 177–201.

Jansen-Vullers, M.H., van Dorp, C.A., Buelens, A.J.M., 2003. Managing traceability information in manufacture. International Journal of Information Management 23, 395–413.

Karlsen, K.M., Van der Roest, J., Olsen, P., 2008. Traceability of Mineral Water – Specification of the Information to be Recorded in Mineral Water Distribution Chains, in Nofima Reports Nofima, Editor.

Kim, H.M., Fox, M.S., Grüninger, M., 1999. An ontology for quality management – enabling quality problem identification and tracing. BT Technology Journal 17(4), 131-140.

Madec, F., Geers, R., Vesseur, P., Kjeldsen, N., Blaha T., 2001. Traceability in the pig production chain. Revue Scientifique Et Technique (International Office of Epizootics) 20 (2), 523–537.

McKean, J.D., 2001. The importance of traceability for public health and consumer protection. Revue Scientifique Et Technique (International Office of Epizootics) 20 (2), 363–371.

Moe, T., 1998. Perspectives on traceability in food manufacture. Trends in Food Science & Technology 9(5), 211–214.

National Soybean Research Laboratory, 2009. About Soy. <http://www.nsrl.uiuc.edu/aboutsoy/images/soychart.gif>

Official Journal of the European Communities, 2002. Regulation (EC) No. 178/2002 of the European Parliament and the Council of 28 January 2002.

Regattieri, A., Gamberi, M., Manzini, R., 2007. Traceability of food products: general framework and experimental evidence. Journal of Food Engineering 81, 347–356.

Schwägele, F., 2005. Traceability from a European perspective. Meat science 71(1), 164-173.

Senneset, G., Forås, E., Fremme, K.M., 2007. Challenges regarding implementation of electronic chain traceability. British Food Journal, 109(10), 805-818.

Stuckenschmidt, H., 2003. Ontology based information in dynamic environments. In: Twelfth IEEE Internation Worskshops on Enabling Technologies: Infrastructures for Collaborative Enterprises(WETICE 2003).

Thakur, M., Hurburgh, C.R. ,2009. Framework for implementing traceability system in the bulk grain supply chain. Journal of Food Engineering, 95(4), 617-626.

Thakur, M., Hurburgh, C.R., 2007. Quality of US soybean meal compared to the quality of soybean meal from other origins. Journal of American Oil Chemists' Society, 84(9), 835-843.

TRACE 2 (2008) *Annex I – TRACE – Tracing Food Commodities in Europe 'Description of Work'*, FP6-2003-FOOD-2-A Proposal No 006942, SIXTH FRAMEWORK PROGRAMME.
TraceFood, 2007. TraceCore – XML Standard Guidelines, TraceFood. <http://193.156.107.66/ff/po/TraceFood/TraceCore%20XML.htm>
TraceFood Wiki, 2009. <http://www.tracefood.org>
US Food and Drug Administration, 2002. The Bioterrorism Act of 2002.

Figure 1. Inputs and outputs at each stage in the soybean value chain

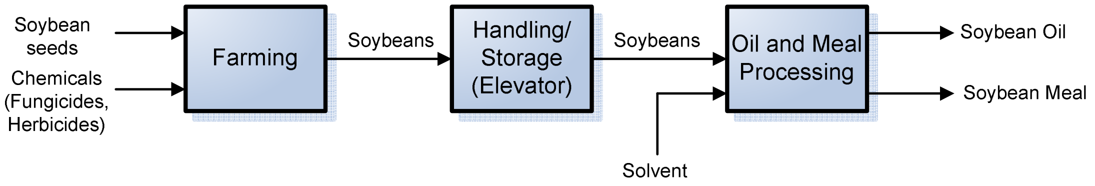


Figure 2. Flowchart of soybean processing (National Soybean Research Laboratory, 2009)

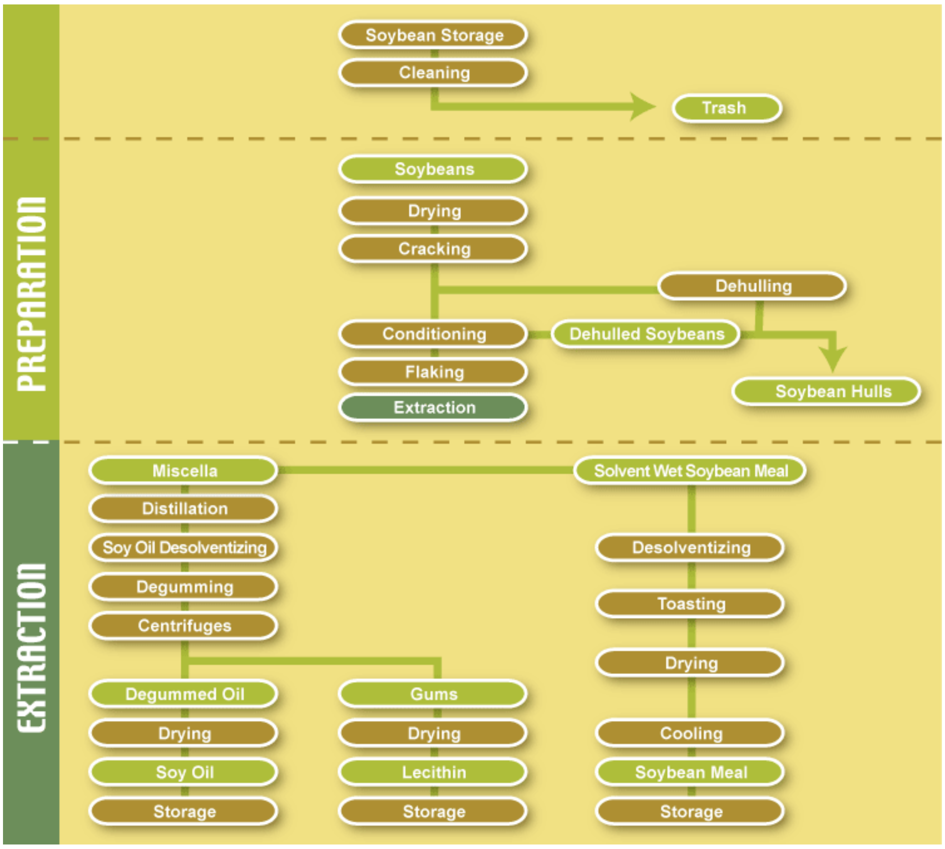


Figure 3. Process flow models for the soybean value chain

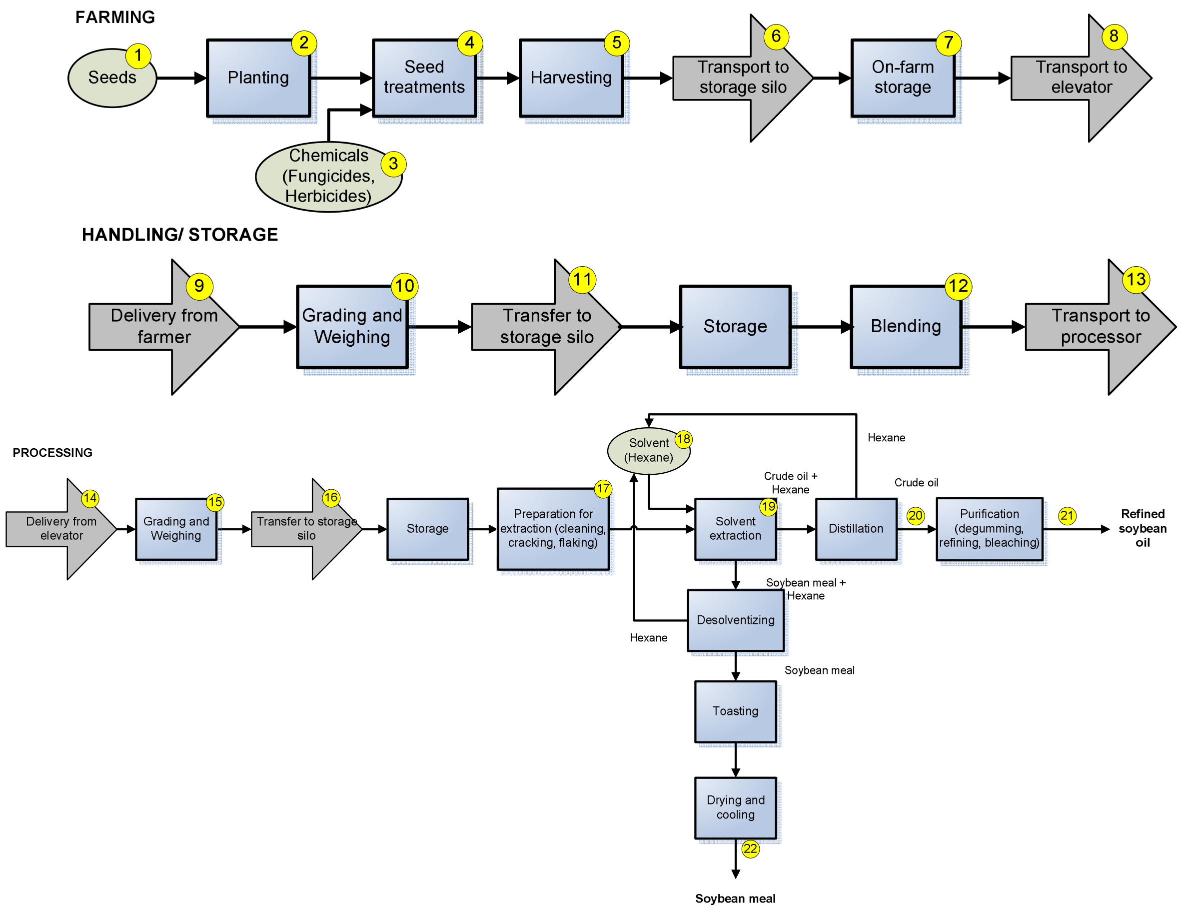


Figure 4. UML class diagram for internal information capture

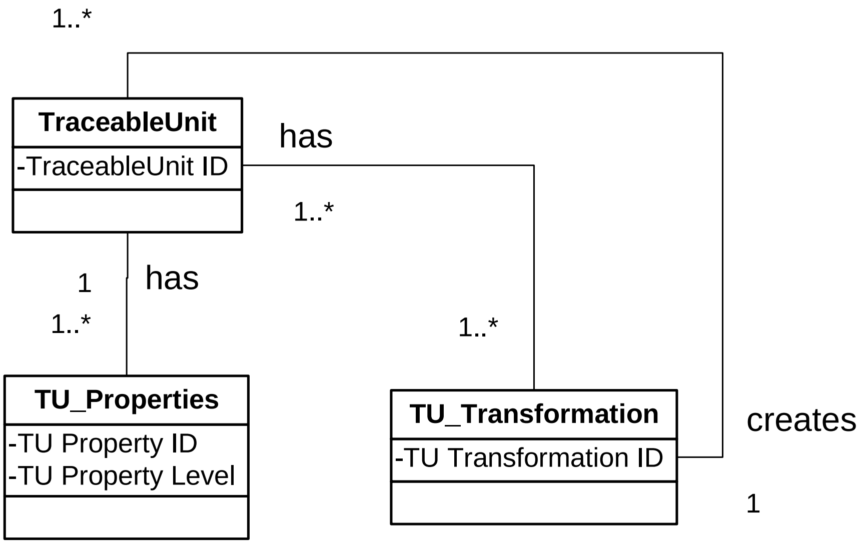
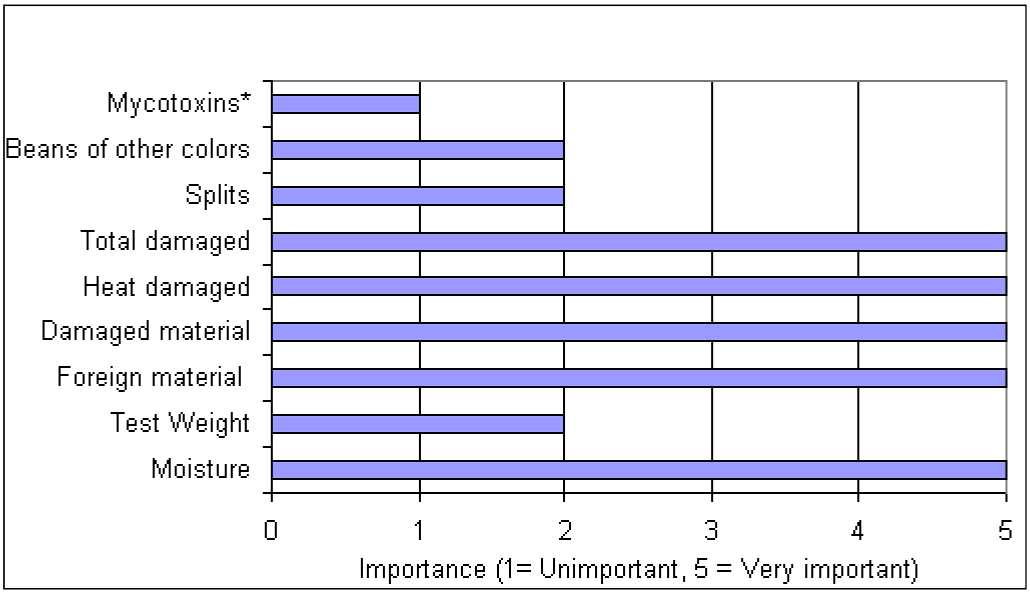
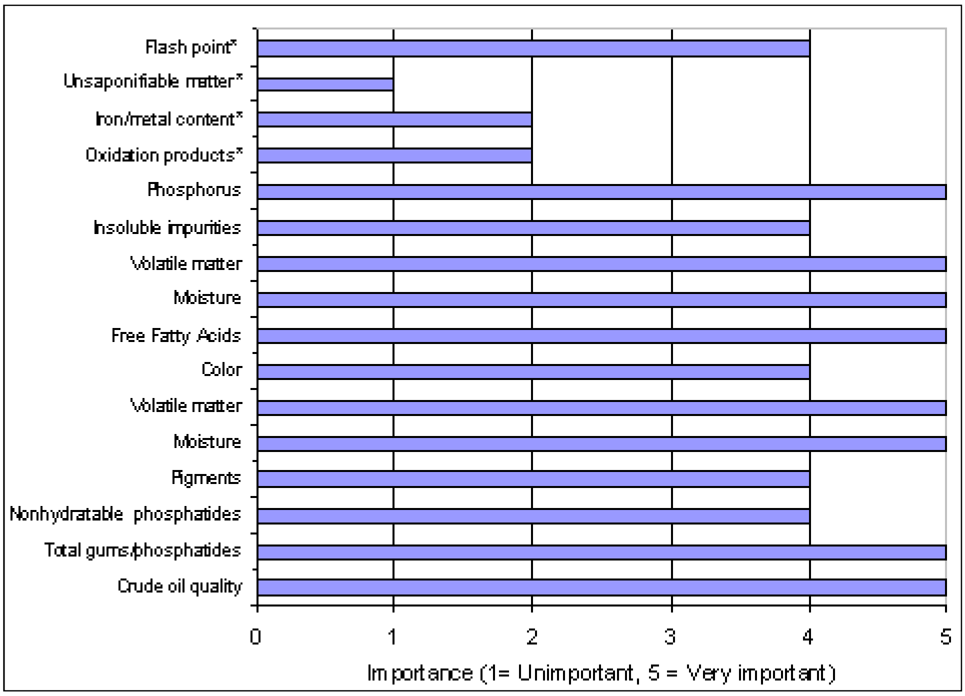


Figure 5. Importance of soybean product properties for the elevator



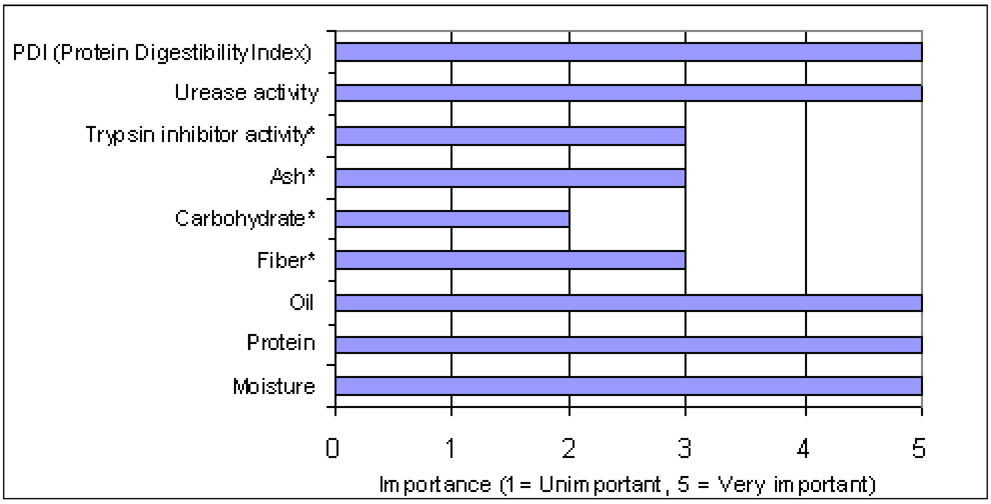
*** The importance of this information is indicated by the processor but the data is not recorded**

Figure 6. Importance of crude oil properties for the processor



*** The importance of this information is indicated by the processor but the data is not recorded**

Figure 7. Importance of soybean meal properties for the processor



*** The importance of this information is indicated by the processor but the data is not recorded**

Table 1. Questions asked on the survey

Question	Possible responses
1. Do you record this information?	Yes or No
2. How important is this information?	Scale 1-5 1 = Unimportant, 5 = Very important
3. Do you communicate this information to anyone outside of your company?	Yes or No
4. How important is this information to your customers?	Scale 1-5 1 = Unimportant, 5 = Very important
5. How important is this information to the end consumers (refined soybean oil used as cooking oil)?	Scale 1-5 1 = Unimportant, 5 = Very important

Table 2. Information to be captured in the soybean farming and handling sectors

Information capture point	Product information	Process information	Quality information
1	Seed variety Seed supplier Logistic unit ID		
2	Seed variety Seed supplier Logistic unit ID	Time of planting Field lots planted Machinery ID	
3	Chemical name Chemical supplier Logistic unit ID		
4	Chemical name Chemical supplier Logistic unit ID	Time of application Quantity applied Field lots treated	
5	Field lot ID	Time of harvesting Field lots harvested Machinery ID Quantity (bushels)	
6	Field lot ID Quantity (bushels)	Time of transport Vehicle ID Destination silo (Storage ID)	
7	Storage ID		Moisture
8	Storage ID Quantity (bushels)	Time of transport Vehicle ID Elevator ID	
9	Farmer ID Logistic unit ID	Time of delivery	
10	Logistic unit ID	Time of grading Quantity (bushels) Grade	Moisture Test weight Foreign material Damaged material
11	Logistic unit ID	Assigned storage ID	
12	Customer order ID	Time of blending Storage ID Quantity used from each storage bin	Moisture Test weight Foreign material Damaged material
13	Customer order ID	Time of transport Transportation ID Processor ID	
14	Elevator ID Logistic unit ID	Time of delivery	

Table 3. Information to be captured in the soybean oil and meal processing sector

Information capture point	Product information	Process information	Quality information
15	Logistic unit ID	Time of grading Quantity (bushels) Grade	Moisture Test weight Foreign material Damaged material
16	Logistic unit ID	Assigned storage ID	
17	Storage ID	Time of preparation Quantity (bushels)	
18	Solvent name Solvent supplier Logistic unit ID		Normal hexane Sulphur content Benzene content
19	Storage ID	Time of extraction Quantity (bushels)	
20	Batch ID	Time of process	Crude oil quality Total gums/phosphatides Nonhydratable phosphatides Pigments Moisture Volatile matter Color Free Fatty Acids Insoluble impurities Phosphorus Triglycerides Trace metals (Iron, Copper)
21	Batch ID	Time of process	Free Fatty Acids Peroxide value Phosphorus Color Moisture Triglycerides Trace metals (Iron, Copper)
22	Batch ID	Time of process	Moisture Protein Oil Urease activity Protein digestibility index (PDI)

Table 4. Identification of Traceable Units at different stages in the supply chain

Information capture point	Traceable Unit Identification (Example)	Information capture point	Traceable Unit Identification (Example)
1	Logistic unit ID (Bag of seeds)	**12**	Customer order ID + Storage bin ID
2	Logistic unit ID (Bag of seeds)	**13**	Customer order ID + Shipment ID
3	Logistic unit ID (Box of chemicals)	**14**	Elevator ID + Customer order ID
4	Logistic unit ID (Box of chemicals)	**15**	Customer order ID
5	Field lot ID (GPS coordinates)	**16**	Customer order ID + Storage bin ID
6	Field lot ID (GPS coordinates)	**17**	Storage bin ID + Process batch ID
7	On-farm storage silo number (Silo 2)	**18**	Logistic Unit ID (Tank of solvent)
8	On-farm storage silo number (Silo 2)	**19**	Storage bin ID + Process batch ID
9	Farmer ID + Transportation ID (Scale ticket number)	**20**	Process batch ID
10	Scale ticket number	**21**	Process batch ID
11	Scale ticket number + Storage ID	**22**	Process batch ID

APPENDIX B: Elevator database code

Data Definition Language

The following section illustrates the use of Data Definition Language to create the tables.

Table Constructions

/* Table Bin*/

```
CREATE TABLE Bin (
Bin_No          VARCHAR(5) PRIMARY KEY,
Depth           NUMBER(6,2) NOT NULL,
Capacity        NUMBER(8,2) NOT NULL);
```

/* Table Farmer*/

```
CREATE TABLE Farmer (
Farmer_ID               CHAR(5) PRIMARY KEY,
Farmer_Name             VARCHAR(30) NOT NULL,
Farmer_Address          VARCHAR(30) NOT NULL,
Farmer_City             VARCHAR(20) NOT NULL,
Farmer_Phone_Num        CHAR(10) NOT NULL);
```

/* Table Elevator_Customer*/

```
CREATE TABLE Elevator_Customer (
Customer_ID             CHAR(5) PRIMARY KEY,
Cus_Name                VARCHAR(30) NOT NULL,
Cus_Address             VARCHAR(30) NOT NULL,
Cus_City                VARCHAR(20) NOT NULL,
Cus_Phone_Num           CHAR(10) NOT NULL);
```

/* Table Purchase*/

/ A check is performed on the grain_type attribute to ensure that a valid grain type is entered in the table. */*

```
CREATE TABLE Purchase (
Scale_Ticket            VARCHAR(12) PRIMARY KEY,
Farmer_ID               CHAR(5) NOT NULL,
Purchase_Date           DATE,
Grain_Type              VARCHAR(20) NOT NULL
                        CHECK (Grain_Type IN ('Corn', 'Soybeans', 'Screenings')),
Bushels                 NUMBER(8,2) NOT NULL,
Moisture                NUMBER(5,2) NOT NULL,
Test_Weight             NUMBER(5,2) NOT NULL,
Damaged_Mt              NUMBER(5,2) NOT NULL,
Foreign_Mt              NUMBER(5,2) NOT NULL,
CONSTRAINT Farmer_ID_FK FOREIGN KEY (Farmer_ID) REFERENCES Farmer(Farmer_ID));
```

/* Table Bin_Activity*/

/ Checks are performed on the grain_type and movement_type attributes to ensure that valid values are entered in the table. * /*

```
CREATE TABLE Bin_Activity (
Activity_Date           TIMESTAMP NOT NULL,
Bin_No                  VARCHAR(5) NOT NULL,
Grain_Type              VARCHAR(20) NOT NULL
                        CHECK (Grain_Type IN ('Corn', 'Soybeans', 'Screenings')),
Moisture                NUMBER(5,2) NOT NULL,
Test_Weight             NUMBER(5,2) NOT NULL,
Damaged_Mt              NUMBER(5,2) NOT NULL,
Foreign_Mt              NUMBER(5,2) NOT NULL,
Movement_Type           VARCHAR(3)
```

```
                        CHECK (Movement_Type IN ('Int', 'In', 'Out')),
Bushels                 NUMBER(8,2) NOT NULL,
CONSTRAINT Bin_Activity_PK PRIMARY KEY (Activity_Date,  Bin_No),
CONSTRAINT Bin_Activity_FK FOREIGN KEY (Bin_No) REFERENCES Bin(Bin_No)
ON DELETE CASCADE);
```

```
/* Table Internal*/
CREATE TABLE Internal (
Activity_Date           TIMESTAMP NOT NULL,
Bin_No                  VARCHAR(5) NOT NULL,
Origin_Bin_No           VARCHAR(5),
Dest_Bin_No             VARCHAR(5),
Emp_Responsible         VARCHAR(30),
CONSTRAINT Internal_PK PRIMARY KEY (Activity_Date, Bin_No),
CONSTRAINT Internal_FK1 FOREIGN KEY (Activity_Date, Bin_No) REFERENCES
Bin_Activity(Activity_Date, Bin_No));
```

```
/* Table Incoming*/
CREATE TABLE Incoming (
Activity_Date           TIMESTAMP NOT NULL,
Bin_No                  VARCHAR(5) NOT NULL,
Scale_Ticket            VARCHAR(12),
CONSTRAINT Incoming_PK PRIMARY KEY (Activity_Date,  Bin_No),
CONSTRAINT Incoming_FK1 FOREIGN KEY (Activity_Date, Bin_No) REFERENCES
Bin_Activity(Activity_Date, Bin_No),
CONSTRAINT Incoming_FK2 FOREIGN KEY (Scale_Ticket) REFERENCES Purchase(Scale_Ticket));
```

```
/* Table Contract*/
/* A check is performed on the grain_type attribute to ensure that a valid grain type is entered in the table.
*/
CREATE TABLE Contract (
Contract_Num            VARCHAR(10) PRIMARY KEY,
Customer_ID             CHAR(5) NOT NULL,
Contract_Date           DATE,
Grain_Type              VARCHAR(20) NOT NULL
                        CHECK (Grain_Type IN ('Corn', 'Soybeans')),
Bushels                 NUMBER(8,2) NOT NULL,
Moisture                NUMBER(5,2) NOT NULL,
Test_Weight             NUMBER(5,2) NOT NULL,
Damaged_Mt              NUMBER(5,2) NOT NULL,
Foreign_Mt              NUMBER(5,2) NOT NULL,
CONSTRAINT Customer_ID_FK FOREIGN KEY (Customer_ID) REFERENCES
Elevator_Customer(Customer_ID));
```

```
/* Table Shipment_Info*/
/* A check is performed on the ship_mode attribute to ensure that a valid shipment mode is entered in the
table. */
CREATE TABLE            Shipment_Info (
Shipment_ID             VARCHAR(12) PRIMARY KEY,
Contract_Num            VARCHAR(10) NOT NULL,
Ship_Mode               CHAR(1)
                        CHECK (Ship_Mode IN ('T', 'R')),
CONSTRAINT Contract_Num_FK FOREIGN KEY (Contract_Num) REFERENCES
Contract(Contract_Num));
```

```
/* Table Outgoing*/
CREATE TABLE Outgoing (
Activity_Date           TIMESTAMP NOT NULL,
Bin_No                  VARCHAR(5) NOT NULL,
```

```
Shipment_ID                VARCHAR(12),
CONSTRAINT Outgoing_PK PRIMARY KEY (Activity_Date,  Bin_No),
CONSTRAINT Outgoing_FK1 FOREIGN KEY (Activity_Date, Bin_No) REFERENCES
Bin_Activity(Activity_Date, Bin_No),
CONSTRAINT Outgoing_FK2 FOREIGN KEY (Shipment_ID) REFERENCES
Shipment_Info(Shipment_ID));

/* Table Truck*/
CREATE TABLE Truck (
Shipment_ID                VARCHAR(12),
Truck_ID                   VARCHAR(5),
CONSTRAINT Truck_PK PRIMARY KEY (Shipment_ID),
CONSTRAINT Truck_FK1 FOREIGN KEY (Shipment_ID) REFERENCES
Shipment_Info(Shipment_ID),
CONSTRAINT Truck_UI1 UNIQUE(Shipment_ID, Truck_ID));

/* Table Rail*/
CREATE TABLE Rail (
Shipment_ID                VARCHAR(12),
Rail_ID                    VARCHAR(5),
Railcar_ID                 VARCHAR(5),
CONSTRAINT Rail_PK PRIMARY KEY (Shipment_ID),
CONSTRAINT Rail_FK1 FOREIGN KEY (Shipment_ID) REFERENCES Shipment_Info(Shipment_ID),
CONSTRAINT Rail_UI1 UNIQUE(Shipment_ID, Rail_ID, Railcar_ID));
```

Data Manipulation Language

The following section illustrates the use of Data Manipulation Language to insert records in all tables.

Insert Statements

```
/* Insert rows in BIN table */
INSERT INTO bin VALUES ('2', 42.1, 4218);
INSERT INTO bin VALUES ('3', 42.1, 1987);
INSERT INTO bin VALUES ('8', 94, 43268);
INSERT INTO bin VALUES ('9', 94, 43268);
INSERT INTO bin VALUES ('11', 84.3, 299375);
INSERT INTO bin VALUES ('12', 84.3, 299375);
INSERT INTO bin VALUES ('13', 84.3, 299375);
INSERT INTO bin VALUES ('14', 84.3, 299375);
INSERT INTO bin VALUES ('19', 48.3, 70257);
INSERT INTO bin VALUES ('20', 94, 109767);
INSERT INTO bin VALUES ('21', 136, 397038);
INSERT INTO bin VALUES ('22', 136, 397038);

/* Insert rows in FARMER table */
INSERT INTO farmer VALUES ('F0001', 'John Smith', '701 4th Ave W.', 'Spencer', '7122626650');
INSERT INTO farmer VALUES ('F0002', 'Ron Penning', '222 West Broadway', 'Leland', '6415673321');
INSERT INTO farmer VALUES ('F0003', 'Pat Torreson', '102 1st Street North', 'Altoona', '5159674215');
INSERT INTO farmer VALUES ('F0004', 'Karl Haglund', '105 4th Avenue SW', 'Dayton', '5155472813');
INSERT INTO farmer VALUES ('F0005', 'Paul Olson', '1800 130th Street', 'Perry', '5154653516');
INSERT INTO farmer VALUES ('F0006', 'Robert Jensen', '2200 RR Street', 'Yale', '6414392243');

/* Insert rows in ELEVATOR_CUSTOMER table */
INSERT INTO elevator_customer VALUES ('C0001', 'Cargill, Inc.', '15615 McGinty Road West',
'Minneapolis, MN', '8002274455');
INSERT INTO elevator_customer VALUES ('C0002', 'Archer Daniels Midland Company', '4666 Faries
Parkway', 'Decatur, IL', '8006375843');
INSERT INTO elevator_customer VALUES ('C0003', 'Grain Processing Corporation', '1600 Oregon Street',
'Muscatine, IA', '5632644211');
```

```sql
INSERT INTO elevator_customer VALUES ('C0004', 'Conagra Grain Processing Co.', '11 ConAgra Drive',
'Omaha, NE', '4025954567');
INSERT INTO elevator_customer VALUES ('C0005', '21st Century Grain Processing', '4800 Main Street',
'Kansas City, MO', '8169947600');

/* Insert rows in PURCHASE table */
INSERT INTO purchase VALUES ('1010', 'F0002', '15-Mar-08', 'Soybeans', 2200, 14.2, 55, 3, 2);
INSERT INTO purchase VALUES ('1011', 'F0002', '15-Mar-08', 'Soybeans', 1564, 14.4, 54.7, 3.2, 2.4);
INSERT INTO purchase VALUES ('1012', 'F0002', '15-Mar-08', 'Soybeans', 3150, 15.1, 54, 3.3, 2.1);
INSERT INTO purchase VALUES ('1027', 'F0005', '15-Mar-08', 'Soybeans', 1000, 15.2, 54, 3.2, 1.2);
INSERT INTO purchase VALUES ('1028', 'F0005', '15-Mar-08', 'Soybeans', 1125, 15.4, 53.5, 3.6, 2.2);
INSERT INTO purchase VALUES ('1029', 'F0005', '15-Mar-08', 'Soybeans', 1054, 15.5, 53.4, 4.0, 3.1);
INSERT INTO purchase VALUES ('1030', 'F0005', '15-Mar-08', 'Soybeans', 1031, 15.3, 54.1, 3.4, 2.9);
INSERT INTO purchase VALUES ('1018', 'F0001', '16-Mar-08', 'Corn', 3200, 15.4, 54.0, 4.4, 2.9);
INSERT INTO purchase VALUES ('1019', 'F0001', '16-Mar-08', 'Corn', 1508, 15.0, 54.5, 3.3, 3.0);
INSERT INTO purchase VALUES ('1020', 'F0001', '16-Mar-08', 'Corn', 2124, 15.2, 54.2, 3.4, 3.1);
INSERT INTO purchase VALUES ('1045', 'F0003', '16-Mar-08', 'Corn', 4850, 15.6, 55.0, 3.4, 2.2);
INSERT INTO purchase VALUES ('1046', 'F0003', '16-Mar-08', 'Corn', 3025, 15.0, 55.0, 3.0, 2.2);
INSERT INTO purchase VALUES ('1047', 'F0003', '16-Mar-08', 'Corn', 4205, 15.2, 54.8, 3.4, 2.5);
INSERT INTO purchase VALUES ('1048', 'F0004', '17-Mar-08', 'Soybeans', 3548, 15.0, 54.2, 3.1, 2.2);
INSERT INTO purchase VALUES ('1049', 'F0004', '17-Mar-08', 'Soybeans', 2045, 15.4, 54.0, 3.4, 2.8);
INSERT INTO purchase VALUES ('1050', 'F0004', '17-Mar-08', 'Soybeans', 4530, 15.5, 54.2, 3.6, 3.0);
INSERT INTO purchase VALUES ('1051', 'F0002', '20-Mar-08', 'Soybeans', 1550, 15.2, 54.0, 3.6, 3.2);
INSERT INTO purchase VALUES ('1052', 'F0004', '21Mar-08', 'Soybeans', 1120, 16.0, 54.0, 3.8, 3.3);

/* Insert rows in BIN_ACTIVITY table */
/* The trigger TRG_ACTIVITY_TYPE inserts the primary key values in corresponding sub-type tables.
Other attributes are added using Update statements. So, each row is inserted partially by the INSERT
statement and a corresponding UPDATE statement.*/

INSERT into bin_activity VALUES (to_timestamp('15-Mar-08 9:00:15', 'DD-MON-YY HH24:MI:SS') ,'2',
'Soybeans', 14.2, 55, 3, 2,'In', 2200);
Update incoming
Set scale_ticket = '1010'
WHERE Activity_date = '15-Mar-08 9:00:15';

INSERT into bin_activity VALUES (to_timestamp('15-Mar-08 10:21:19', 'DD-MON-YY HH24:MI:SS')
,'2', 'Soybeans', 14.4, 54.7, 3.2, 2.4, 'In', 1564);
Update incoming
Set scale_ticket = '1011'
WHERE Activity_date = '15-Mar-08 10:21:19';

INSERT into bin_activity VALUES (to_timestamp('15-Mar-08 11:30:00', 'DD-MON-YY HH24:MI:SS')
,'8', 'Soybeans', 15.1, 54, 3.3, 2.1, 'In', 3150);
Update incoming
Set scale_ticket = '1012'
WHERE Activity_date = '15-Mar-08 11:30:00';

INSERT into bin_activity VALUES (to_timestamp('15-Mar-08 11:55:10', 'DD-MON-YY HH24:MI:SS')
,'8', 'Soybeans', 15.2, 54, 3.2, 1.2, 'In', 1000);
Update incoming
Set scale_ticket = '1027'
WHERE Activity_date = '15-Mar-08 11:55:10';

INSERT into bin_activity VALUES (to_timestamp('15-Mar-08 12:25:15', 'DD-MON-YY HH24:MI:SS')
,'8', 'Soybeans', 15.4, 53.5, 3.6, 2.2, 'In', 1125);
Update incoming
Set scale_ticket = '1028'
WHERE Activity_date = '15-Mar-08 12:25:15';
```

```
INSERT into bin_activity VALUES (to_timestamp('15-Mar-08 13:44:00', 'DD-MON-YY HH24:MI:SS')
,'8', 'Soybeans', 15.5, 53.4, 4.0, 3.1, 'In', 1054);
Update incoming
Set scale_ticket = '1029'
WHERE Activity_date = '15-Mar-08 1:44:00 PM';

INSERT into bin_activity VALUES (to_timestamp('15-Mar-08 14:50:00', 'DD-MON-YY HH24:MI:SS')
,'8', 'Soybeans', 15.3, 54.1, 3.4, 2.9, 'In', 1031);
Update incoming
Set scale_ticket = '1030'
WHERE Activity_date = '15-Mar-08 2:50:00 PM';

INSERT into bin_activity VALUES (to_timestamp('16-Mar-08 08:10:29', 'DD-MON-YY HH24:MI:SS')
,'9', 'Corn', 15.4, 54.0, 4.4, 2.9, 'In', 3200);
Update incoming
Set scale_ticket = '1018'
WHERE Activity_date = '16-Mar-08 08:10:29';

INSERT into bin_activity VALUES (to_timestamp('16-Mar-08 09:21:00', 'DD-MON-YY HH24:MI:SS')
,'9', 'Corn', 15.0, 54.5, 3.3, 3.0, 'In', 1508);
Update incoming
Set scale_ticket = '1019'
WHERE Activity_date = '16-Mar-08 09:21:00';

INSERT into bin_activity VALUES (to_timestamp('16-Mar-08 09:56:00', 'DD-MON-YY HH24:MI:SS')
,'9', 'Corn', 15.2, 54.2, 3.4, 3.1, 'In', 2124);
Update incoming
Set scale_ticket = '1020'
WHERE Activity_date = '16-Mar-08 09:56:00';

INSERT into bin_activity VALUES (to_timestamp('16-Mar-08 11:05:00', 'DD-MON-YY HH24:MI:SS')
,'9', 'Corn', 15.6, 55.0, 3.4, 2.2, 'In', 4850);
Update incoming
Set scale_ticket = '1045'
WHERE Activity_date = '16-Mar-08 11:05:00';

INSERT into bin_activity VALUES (to_timestamp('16-Mar-08 13:10:00', 'DD-MON-YY HH24:MI:SS')
,'9', 'Corn', 15.0, 55.0, 3.0, 2.2, 'In', 3025);
Update incoming
Set scale_ticket = '1046'
WHERE Activity_date = '16-Mar-08 1:10:00 PM';

INSERT into bin_activity VALUES (to_timestamp('16-Mar-08 15:22:00', 'DD-MON-YY HH24:MI:SS')
,'9', 'Corn', 15.2, 54.8, 3.4, 2.5, 'In', 4205);
Update incoming
Set scale_ticket = '1047'
WHERE Activity_date = '16-Mar-08 3:22:00 PM';

INSERT into bin_activity VALUES (to_timestamp('17-Mar-08 10:25:00', 'DD-MON-YY HH24:MI:SS')
,'11', 'Soybeans', 15.0, 54.2, 3.1, 2.2, 'In', 3548);
Update incoming
Set scale_ticket = '1048'
WHERE Activity_date = '17-Mar-08 10:25:00';

INSERT into bin_activity VALUES (to_timestamp('17-Mar-08 11:44:00', 'DD-MON-YY HH24:MI:SS')
,'11', 'Soybeans', 15.4, 54.0, 3.4, 2.8, 'In', 2045);
Update incoming
Set scale_ticket = '1049'
```

```sql
WHERE Activity_date = '17-Mar-08 11:44:00';

INSERT into bin_activity VALUES (to_timestamp('17-Mar-08 14:15:00', 'DD-MON-YY HH24:MI:SS')
,'11', 'Soybeans', 15.5, 54.2, 3.6, 3.0, 'In', 4530);
Update incoming
Set scale_ticket = '1050'
WHERE Activity_date = '17-Mar-08 2:15:00 PM';

INSERT into bin_activity VALUES (to_timestamp('18-Mar-08 14:15:00', 'DD-MON-YY HH24:MI:SS')
,'11', 'Soybeans', 15.2, 54, 3.8, 3.2, 'Int', -1000);
Update internal
Set dest_bin_no  = '12', emp_responsible = 'Jacob Smith'
WHERE Activity_date = '18-Mar-08 2:15:00 PM';

INSERT into bin_activity VALUES (to_timestamp('19-Mar-08 10:25:00', 'DD-MON-YY HH24:MI:SS')
,'9', 'Corn', 15.2, 55, 3.6, 3.0, 'Int', -500);
Update internal
Set dest_bin_no  = '3', emp_responsible = 'John Bolson'
WHERE Activity_date = '19-Mar-08 10:25:00';

INSERT into bin_activity VALUES (to_timestamp('19-Mar-08 11:39:00', 'DD-MON-YY HH24:MI:SS')
,'2', 'Soybeans', 15.4, 55.2, 3.6, 3.0, 'Int', 400);
Update internal
Set origin_bin_no  = '8', emp_responsible = 'John Bolson'
WHERE Activity_date = '19-Mar-08 11:39:00';

/* Insert rows in CONTRACT table */
INSERT into contract VALUES ('C032208', 'C0001', '22-Mar-08', 'Soybeans', 7000, 15.5, 53, 3.6, 2.6);
INSERT into contract VALUES ('A042508', 'C0002', '25-Apr-08', 'Soybeans', 6000, 15.6, 53.2, 3.8, 2.8);
INSERT into contract VALUES ('G042808', 'C0003', '22-Mar-08', 'Corn', 5000, 15.4, 53.8, 3.6, 2.9);
INSERT into contract VALUES ('CA031708', 'C0004', '17-Mar-08', 'Soybeans', 3000, 15.5, 53, 3.6, 2.6);
INSERT into contract VALUES ('CG040608', 'C0005', '06-Apr-08', 'Corn', 4000, 15.6, 52.8, 3.8, 2.7);
INSERT into contract VALUES ('CG040908', 'C0005', '06-Apr-08', 'Corn', 4000, 15.6, 52.8, 3.8, 2.7);

/* Insert rows in SHIPMENT_INFO table */
/* The trigger TRG_SHIP_MODE inserts the primary key values in corresponding sub-type tables. The
other attributes are added using Update statements. So, each row is inserted partially by the INSERT
statement and a corresponding UPDATE statement.*/

INSERT into shipment_info VALUES ('S10001', 'C032208', 'R');
INSERT into shipment_info VALUES ('S10002', 'A042508', 'R');
INSERT into shipment_info VALUES ('S10003', 'G042808', 'R');
INSERT into shipment_info VALUES ('S10004', 'CA031708', 'T');
INSERT into shipment_info VALUES ('S10005', 'CG040608', 'T');
INSERT into shipment_info VALUES ('S10006', 'CG040608', 'R');

Update rail
Set rail_ID  = '10001', railcar_ID = '01'
WHERE shipment_id = 'S10001';

Update rail
Set rail_ID  = '10001', railcar_ID = '11'
WHERE shipment_id = 'S10002';

Update rail
Set rail_ID  = '10002', railcar_ID = '02'
WHERE shipment_id = 'S10003';
```

```
Update truck
Set truck_ID = '20001'
WHERE shipment_id = 'S10004';

Update truck
Set truck_ID = '20002'
WHERE shipment_id = 'S10005';

Update rail
Set rail_ID  = '10003', railcar_ID = '12'
WHERE shipment_id = 'S10006';
```

/* Insert rows in BIN_ACTIVITY table */
/ The trigger TRG_ACTIVITY_TYPE inserts the primary key values in corresponding sub-type tables. Other attributes are added using Update statements. So, each row is inserted partially by the INSERT statement and a corresponding UPDATE statement.*/*

```
INSERT into bin_activity VALUES (to_timestamp('25-Mar-08 10:25:00', 'DD-MON-YY HH24:MI:SS')
,'2', 'Soybeans', 14.7, 54.9, 3.27, 2.47, 'Out', -2000);

INSERT into bin_activity VALUES (to_timestamp('25-Mar-08 10:25:00', 'DD-MON-YY HH24:MI:SS')
,'8', 'Soybeans', 15.3, 53.8, 3.5, 2.3, 'Out', -5000);

Update outgoing
Set shipment_ID = 'S10001'
WHERE Activity_date = '25-Mar-08 10:25:00';

INSERT into bin_activity VALUES (to_timestamp('28-Apr-08 11:30:00', 'DD-MON-YY HH24:MI:SS')
,'11', 'Soybeans', 15.28, 54.1, 3.48, 2.8, 'Out', -6000);

Update outgoing
Set shipment_ID = 'S10002'
WHERE Activity_date = '28-Apr-08 11:30:00';

INSERT into bin_activity VALUES (to_timestamp('29-Apr-08 09:25:00', 'DD-MON-YY HH24:MI:SS')
,'9', 'Corn', 15.23, 54.64, 3.5, 2.7, 'Out', -5000);

Update outgoing
Set shipment_ID = 'S10003'
WHERE Activity_date = '29-Apr-08 09:25:00';

INSERT into bin_activity VALUES (to_timestamp('02-May-08 14:25:00', 'DD-MON-YY HH24:MI:SS')
,'9', 'Corn', 15.23, 54.64, 3.5, 2.7, 'Out', -4000);

Update outgoing
Set shipment_ID = 'S10005'
WHERE Activity_date = '02-May-08 2:25:00 PM';
```

/* An outgoing shipment can contain grain from more than one bin as demonstrated by the following INSERT statements. */

```
INSERT into bin_activity VALUES (to_timestamp('02-May-08 10:21:00', 'DD-MON-YY HH24:MI:SS')
,'2', 'Soybeans', 14.7, 54.9, 3.27, 2.47, 'Out', -1500);

INSERT into bin_activity VALUES (to_timestamp('02-May-08 10:21:00', 'DD-MON-YY HH24:MI:SS')
,'11', 'Soybeans', 15.28, 54.1, 3.48, 2.8, 'Out', -1500);
```

```
Update outgoing
Set shipment_ID = 'S10004'
WHERE Activity_date = '02-May-08 10:21:00';

INSERT into bin_activity VALUES (to_timestamp('28-Mar-08 10:25:00', 'DD-MON-YY HH24:MI:SS')
,'2', 'Soybeans', 14.7, 54.9, 3.27, 2.47, 'Out', -664);

INSERT into bin_activity VALUES (to_timestamp('28-Mar-08 10:30:00', 'DD-MON-YY HH24:MI:SS')
,'2', 'Soybeans', 14.7, 54.9, 3.27, 2.47, 'In', 2000);

INSERT into bin_activity VALUES (to_timestamp('28-Mar-08 10:33:00', 'DD-MON-YY HH24:MI:SS')
,'2', 'Soybeans', 14.7, 54.9, 3.27, 2.47, 'In', 1500);

Update outgoing
Set shipment_ID = 'S10006'
WHERE Activity_date = '28-Mar-08 10:25:00';
```

Database Triggers

Two database triggers were created to populate the sub-type tables. The trigger trg_activity_type populates the Internal, Incoming and Outgoing tables depending on the movement_type attribute entered in each row of the Bin_activity table. The trigger trg_ship_mode populates the Truck and Rail tables depending on the ship_mode attribute entered in each row of the Shipment_info table.

```
/* Create Trigger TRG_ACTIVITY_TYPE */
CREATE OR REPLACE TRIGGER trg_activity_type
AFTER INSERT ON bin_activity
FOR EACH ROW
BEGIN
        IF :new.movement_type = 'Int' THEN
          INSERT into Internal(activity_date, bin_no) VALUES (:new.activity_date, :new.bin_no);
        ELSIF :new.movement_type = 'In' THEN
           INSERT into Incoming(activity_date, bin_no) VALUES (:new.activity_date, :new.bin_no);
        ELSE
          INSERT into Outgoing(activity_date, bin_no) VALUES (:new.activity_date, :new.bin_no);
        END IF;
END;
/

/* Create Trigger TRG_SHIP_MODE */
CREATE OR REPLACE TRIGGER trg_ship_mode
AFTER INSERT ON shipment_info
FOR EACH ROW
BEGIN
        IF :new.ship_mode = 'R' THEN
          INSERT into Rail(shipment_ID) VALUES (:new.shipment_ID);
       ELSE
         INSERT into Truck(shipment_ID) VALUES (:new.shipment_ID);
  END IF;
END;
/
```