Journal of Food Engineering 147 (2015) 24–38

Contents lists available at ScienceDirect

Journal of Food Engineering

journal homepage: www.elsevier.com/locate/jfoodeng

A national produce supply chain database for food safety risk analysis

Denyse I. LeBlanc [a], Sébastien Villeneuve [b,*], Leila Hashemi Beni [b], Ainsley Otten [c], Aamir Fazil [c], Robin McKellar [d], Pascal Delaquis [d]

[a] Agriculture and Agri-Food Canada, Atlantic Food and Horticulture Research Centre, c/o Food Research Centre, Pavillon Jacqueline-Bouchard, Université de Moncton, Moncton, New Brunswick E1A 3E9, Canada
[b] Agriculture and Agri-Food Canada, Food Research and Development Centre, 3600 Casavant Blvd West, Saint-Hyacinthe, Quebec J2S 8E3, Canada
[c] Public Health Agency of Canada, 160 Research Lane, Unit 206, Guelph, Ontario N1G 5B2, Canada
[d] Agriculture and Agri-Food Canada, Pacific Agri-Food Research Centre, Highway 97, Summerland, British Columbia V0H 1Z0, Canada

ARTICLE INFO

Article history:
Received 3 July 2014
Received in revised form 16 September 2014
Accepted 17 September 2014
Available online 28 September 2014

Keywords:
Relational database
Food supply chain
Logistics
Food safety risk assessment
Simulation
Escherichia coli O157:H7

ABSTRACT

During a foodborne crisis, risk assessors are often scrambling to assemble data needed to trace suspected foods along very complex supply chains. Although traceability systems ensure that stakeholders in the supply chain record lot-specific trace-back and trace-forward data, there are few databases available that describe in detail the flow of product in the complex web of supply chains. This paper presents the methodological approach used to design and assemble a relational database of nation-wide trade data for packaged ready-to-eat lettuce and leafy greens. The database was used in the development of an integrated simulation tool (Canadian GIS-based Risk Assessment, Simulation and Planning for food safety tool, i.e. CanGRASP) that can predict the spatial distribution and public health risk associated with contaminated food. The database includes the geographical coordinates of 5 domestic processors, 28 produce distribution centres and 2946 retail outlets from five of the top ten retail chains in Canada. It also includes other critical information to predict the fate of pathogens during distribution of contaminated product through the supply chain including: (a) product volumes handled by each stakeholder, (b) flow of product between stakeholders, (c) temperatures of product each season, and (d) times products spend in each step or during transit between steps, for each season. The database is used by both the simulation and mapping components of the integrated simulation tool during risk assessment exercises associated with emergency preparedness planning and training. Using the database, CanGRASP was able to assess the spread of the population at risk during a simulation of a hypothetical outbreak caused by fresh-cut leafy vegetables contaminated with *Escherichia coli* O157:H7 in the Canadian food distribution systems during both summer and winter seasons.

1. Introduction

A systems-based approach was recently developed for the identification of vulnerabilities and the assessment of risks associated with contamination of the Canadian food supply with microbiological threat agents along the farm-to-fork chain. This led to the design of an integrated simulation tool that can predict the spatial distribution and public health risk associated with contaminated food. The tool will enhance risk assessment capacity to better prepare for threats directed at foods disseminated through increasingly complex distribution systems.

The integrated simulation tool includes an innovative GIS-based interface that can dynamically map and predict the spread of specific microbiological threat agents along the farm-to-fork continuum (Hashemi Beni et al., 2011, 2012). Referred to as the Canadian GIS-based Risk Assessment, Simulation and Planning (CanGRASP) tool for food safety, it can be used to simulate a food contamination scenario, map the predicted spatial distribution of the contaminated product over time, and predict public health impact in specific regions of Canada. The tool can be adapted to the simulation of food safety risk behaviour in any food commodity provided the necessary relational databases and predictive risk equations are available. The ability to predict the rate of dissemination of food safety threats and their impact through a distribution system enhances response capacity during an actual contamination event. In addition, the ability to generate realistic contamination scenarios is invaluable for planning purposes, specifically for

* Corresponding author. Tel.: +1 450 768 3335; fax: +1 450 773 8461.
E-mail address: sebastien.villeneuve@agr.gc.ca (S. Villeneuve).

http://dx.doi.org/10.1016/j.jfoodeng.2014.09.026
0260-8774/

the identification of vulnerabilities in food systems, and for the development or optimization of risk-mitigation measures that can be deployed during a contamination event.

Development of the tool required the collection of nation-wide data on production, importation, processing, distribution and retailing systems, and the compilation of associated databases to enable interfacing with computerized vulnerability assessment software.

This paper describes the method used to design and assemble a relational database of nation-wide trade data for a single food category, packaged ready-to-eat (RTE) lettuce and leafy greens, a product amenable to the widespread dissemination of infectious microorganisms. The relational database includes data on Canadian processing/packing facilities and retail distribution systems for analysis using computerized mapping tools developed to predict the dissemination of contaminated food along temporal and geographic planes. To our knowledge, this represents the first attempt to assemble a detailed database describing a retail supply network on a national scale. The relational database was used to assess the new CanGRASP tool for food safety by comparing the spread of the population at risk during the simulation of an outbreak caused by fresh-cut leafy vegetables contaminated with *Escherichia coli* O157:H7 in the Canadian food distribution systems, during both summer and winter seasons.

2. Literature review

Several modelling approaches were developed in the past few years to analyse the risk in food supply chains. Vialette et al. (2005) used meta-analysis of food safety information based on a combination of relational databases associated with quantitative microbiology models. Laguerre et al. (2013) identified the need to improve the design and management of the cold chain and proposed combined deterministic and stochastic approaches for modelling the evolution of food products along the cold chain (Flick et al., 2012; Hoang et al., 2012). Sym'Previus, a project initiated several years ago, involved the development of two software tools to help manage food safety (Couvert et al., 2005, 2006; Haemmerlé et al., 2007; Hignette et al., 2008; Buche et al., 2011; Coroller et al., 2012; Destercke et al., 2013). These include (1) a database containing information on the behaviour of microorganisms and natural contaminants in foods, that can be queried with a specifically developed system (MIEL) allowing formulation of interrogations on specific foods and microorganisms; (2) a user-friendly software tool that simulates the growth of microorganisms in a food matrix. In the Sym'Previus system, a semi-automatic acquisition tool (i.e. @WEB) retrieves scientific documents from the Web. Moreover, a flexible querying system using the domain ontology to scan simultaneously local and Web data was developed in order to feed the predictive modelling tools available on the Sym'Previus platform. Otherwise, more specific models and approaches that assess microbiological risks associated with the growth or decay of specific pathogens in fresh-cut lettuce and leafy greens in distribution chains were also recently published (Rijgersberg et al., 2010; Tromp et al., 2010, 2012; Danyluk and Schaffner, 2011; McKellar and Delaquis, 2011; McKellar et al., 2012; Zeng et al., 2014; McKellar et al., 2014).

However, there are few examples of the use of Geographical Information Systems (GIS) to model the dispersal of specific commodities in food distribution systems on a national scale and to map the consequential risks for the surrounding populations. Quantification of the public health impact related to a spatio-temporal distribution of contaminated product in a food supply system, to multiple retail outlets for example, presents serious challenges. Under such circumstances, calculating consumer accessibility to the product is as important as estimating the level of contamination and number of contaminated items in each retail outlet in a food supply system. GIS tools are well adapted for calculating consumer accessibility and have often been used to map accessibility to retail outlets as part of food desert assessments (Apparicio et al., 2007; Larsen and Gilliland, 2008; Sharkey, 2009; Smith et al., 2010; Michimi and Wimberly, 2010; Eckert and Shetty, 2011; Van Meter et al., 2011; Gordon et al., 2011).

During the 2011 *E. coli* O104:H4 outbreak in Germany, the lack of availability of detailed databases describing the spatial and temporal distributions of foods was noted. Since then, risk assessors have been exploring new data sources and data analysis technologies which would support the risk assessment process during a food chain crisis investigation. For example, Filter et al. (2012) used commercial data from retail tracking services (e.g. commercial B2B data) to generate hypotheses on possible contamination sources during a foodborne crisis. They assembled retail sales data from German food retailing companies (e.g. weight of a specific product sold per week) for each of the 88 German postal code 2 regions. In another instance, Pinior et al. (2012) used partial trade data and milk consumption data to model trade flow and trade links of the German milk supply chain. However, they were forced to simplify the German dairy trade to a network with three categories of nodes (e.g. milk producers, dairies and consumers), where all milk producers in a county were aggregated into a single milk producer node and all consumers in a municipality were aggregated into one consumer node. Since the only trade data and trade connection information available was between milk producer nodes and dairy nodes, trade data between dairy nodes and consumer nodes were estimated based on average per capita milk consumption while trade connection information was predicted using a standard randomized gravity model. Unfortunately, the product flow information in the databases assembled by Filter et al. (2012) and Pinior et al. (2012) was not associated with individual retail outlets but rather with a group of outlets in a discrete region. During the 2011 German *E. coli* O104:H4 outbreak, Weiser et al. (2013) assembled a more detailed relational database based on lot-specific trace-back and trace-forward data supplied by food business operators. Risk assessors were able to use this database to summarize the large amount of complex trade data and identify a common link of all outbreak clusters for which data were available. However, two of the limitations identified by the risk assessors during their investigation were the lack of information in the database on the classification of food business operators (e.g. producer, processor, wholesaler, retailer or food service operator) and the unavailability of more complete trading data generated during a non-outbreak period.

3. Materials and methods

3.1. Description of the integrated simulation tool

The CanGRASP food safety tool was developed such that the ArcGIS™ (Esri) application is tightly coupled with the Arena® (Rockwell Automation) simulation tool (Hashemi Beni et al., 2011). An interface was customized to interact with the users, the database and the simulation tool to enable the interchange of data back and forth between the ArcGIS and Arena components. The parameters required for a simulation are input by the users via the control screen of the interface (Fig. 1a and b). Combining these parameters with the other information that Arena reads directly from the database, a simulation is run. In the Arena simulation, lettuce packages are defined as discrete entities that are transported through the modeled supply chain. Using the Analyse and Map Risk feature of the interface (Fig. 1c), CanGRASP tools can