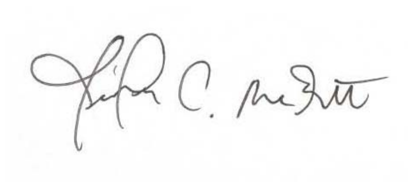
INSTITUTE OF FOOD TECHNOLOGISTS

Institute of Food Technologists/Food and Drug Administration
Contract No. **223-04-2503**

Analysis and Review of Topics in Areas of Food Safety, Food Security, Food Processing, and Human Health

Task Order No.7 Final Report: *Revised*
Tracing Systems: An Exercise Exploring Data Needs and Design



12/01/09

Jennifer C. McEntire, Ph.D.
Research Scientist & Manager,
Science and Technology Projects

Date

Executive Summary

The U.S. Food and Drug Administration (FDA) under a five year contract (# 223-04-2503) with the Institute of Food Technologists (IFT) issued a task order for a product tracing exercise exploring data needs and design. Specifically, IFT was to organize and implement a mock trace-back/trace forward exercise, and evaluate the mock trace-back/trace forward exercise, including key issues and lessons learned in platform design, technology needs, data gaps, and data management. IFT was also tasked with reviewing the utilization of similar technology platforms to store and manage industry product tracing data on a national scale for fresh produce and its applicability to other food items. FDA may use this information to further explore the flow of information and information needs as they relate to technologies which may be used to track movement of food products back to the original source and forward through the supply chain to increase FDA's ability to more rapidly and precisely track and trace the origin and destination of contaminated food items.

The foundation for this work began in August 2008, following the *Salmonella* Saintpaul outbreak, with the "Harvard University Executive Session on Food Safety" convened by the Harvard Kennedy School of Government. The Harvard Kennedy School brought together federal and state food safety officials, leading information technology firms, and several stakeholders in the tomato supply chain, including trade association and industry representatives from various segments of the chain to explore issues of the *Salmonella* Saintpaul outbreak and possible solutions in an "Executive Session on Food Safety". The Executive Session on Food Safety met four times before this task was issued to examine the current strategy used to trace food backwards and forwards. Session attendees agreed that trace-backs use many resources, including often lengthy amounts of time, and that it would be ideal to reduce the amount of time needed to trace food back to find points of commonality, and trace forward to where the food product was distributed. It would also be ideal if industry data and technologies already existed that could lend themselves to improved trace-back/trace forward exercises. All Executive Session participants had the same overall goal to increase and assure food safety in the US, while maintaining consumer trust and confidence, but they often did not work closely together to achieve these goals outside of a crisis situation. Working at a deeper level of collaboration allowed participants to brainstorm an exercise that would explore existing industry data and a tool for expedited trace-back/trace forward exercises.

Over the course of its four meetings, the Executive Session on Food Safety arrived at four matters it felt should be investigated in a mock trace-back/trace forward exercise:

> **1) Ease of Participation and Use**. The Session wanted to understand the "ease of participation and use" for industry to submit and store data and for government to use the data.
>
> **2) Illuminate the Supply Chain**. The Session wanted to understand whether and how, using available data and visualization software, industry working with government could, upon request, illuminate designated supply chains sufficient to conduct trace-back/trace forward exercises.

3) Expedite Investigations. The Session wanted to understand whether the approach taken could expedite trace-back/trace forward investigations, and what the implications and requirements are for an improved trace-back/trace forward process.

4) Collaboration. The Session wanted to understand whether and how government and industry could come together to better understand the governments' trace-back/trace forward process and requirements; provide appropriate industry data to aid in a trace-back; determine information technology tools that could improve a trace-back; and initiate a mock trace-back/trace forward exercise using these resources to see if they allowed industry and government to better work together and result in faster trace-back/trace forward investigations.

IFT was contracted in June 2009 to address some of the issues raised by the Executive Session. From June to September 2009, IFT, with subcontractors (Harvard, TIBCO, and Microsoft), FDA, some original Executive Session members and others, worked to finish framing out the mock trace-back/trace forward, finalizing data needs and system preparations, and creating scenarios to test in the exercise, and conducted the actual mock trace-back/trace forward exercise.

Large amounts of existing tomato industry data from many points in the supply chain were collected to conduct a mock trace-back/trace forward exercise for both open and "closed" supply chains. Historical data from a two-week period in early November 2008 was used, and the initial data elements to be captured and the geographical area data represented were determined prior to this task commencing in June 2009. The original data set was expanded with data from repackers, distributors and retail necessary to complete the entire open supply chain. Some companies were consulted numerous times since the data determined to be necessary evolved over the course of the task. Quality assurance of data was necessary after capture to ensure that data were available for all fields required for this task and that the data appeared to be accurate.

Standardization of data fields such as dates, quantities, and addresses was also necessary, in addition to establishing linkages for both internal and external traceability primarily using lot code and repack numbers. Only then were data that had been quality assured, standardized, and were considered complete able to be uploaded to the visualization tool and used for the mock trace-back/trace forward exercise. Not all data collected could be used for this task. A data acquisition tool was created for this task which integrated the industry data in a spreadsheet database and provided storage of the data prior to export to a visualization tool.

Lot and repack run numbers could be viewed in the spreadsheet database and linkages could be found between the various lot and/or repack numbers using date and other data elements. The definition of a lot often varies by company and not all lots are assigned a number or code, but typically shipped tomatoes receive an outgoing lot number and are assigned a new lot number upon arrival at the repacker, with that lot number amended each time the tomato is graded or sized. This "repack run number", which may be unique to the tomato industry, provides the pedigree of the original lot as received from the shipper, the processes undertaken to finish the product for the customer, and the lot ID for the final product shipment to the customer. Once linkages between lots were established using various data elements, product movement for a specified time period could be visually traced in the visualization tool for the portion of the tomato supply chain that data were available from. Visualization allowed for faster determination of points of commonality in the chain.

The mock trace-back/trace forward was able to very rapidly illuminate points of commonality in tomato supply chains when data were able to be uploaded to the visualization tool. Limitations to the exercise include that it tested limited tomato industry data sets. Aspects such as real time data, other food products, a broader geographical region, import data, and a complete supply chain weren't fully tested. Significant time and effort was spent on the front-end for data collection, quality assurance of data for gaps, and standardization of select data fields where necessary. This exercise was able to show the potential to expedite trace-backs, based on data availability, capture, and readiness, but continued exploration and work needs to be done to conduct the trace-back/trace forward process more quickly and smoothly in real time.

The exercise also showed the value of collaboration to reach a shared goal. Collaboration was necessary to better understand all of the data that were collected for this task, and to better utilize both industry and government resources. Government/industry collaboration before, during and after food safety issues is essential to pool assets such as industry's data and business knowledge, government's investigative knowledge and legal authority, and technologies, in an effort to protect public health. Technology is a critical facilitator of collaboration, and is readily available, although training, costs, and other requirements must be taken into account when new technologies are considered.

IFT also compared the system used for the mock trace-back/trace forward exercise with 10 other product tracing systems. Information was gathered from 10 traceability technology providers who shared details on their various systems for comparison with the system used in this task. While this task focused on tomatoes, all systems report that they could be used for virtually all food products. None of the 10 systems used for comparison were tested by IFT, and it would be worthwhile to test the capabilities of other systems in additional pilots, and to test the applicability of the systems, including the one used for this exercise, to other segments of the food industry.

The results of this exercise should be viewed in context. It tested limited tomato industry data sets so that aspects such as real time data, other food products, a broader geographical region, import data, and a complete supply chain weren't fully tested. Additional pilots on tomatoes, as well as other produce lines or other sectors of the food industry, could allow for continued learning. Determining what data are available outside the tomato industry would help to determine key data elements needed to establish linkages among the data. Required data elements that allow for linkages can lead to improved data acquisition, quality assurance, storage and use, which in turn will lead to an improved trace-back/trace forward effort. Applicability of this and other existing product tracing technologies to other sectors of the food industry and expanded geographical areas could also be explored, as well as applicability to various size companies.

The mock trace-back/trace forward exercise was able to show there is potential to expedite trace-backs by visualizing supply chains to find points of commonality based on data availability, capture, and readiness. The significant time and effort needed on the front-end of this task for quality assurance and standardization of data could be eliminated potentially making trace-back/trace forward exercises even quicker if key data elements were identified and required to be maintained in a standardized form for product tracing. Identification and standardization of key data elements that industry should record and submit to FDA upon request to show linkages along the entire supply chain, as well as continued collaboration among industry, government and other food safety stakeholders are essential for improved trace-back/trace forward exercises and improved overall food safety. However, to assure collaboration and to identify and standardize key data elements more information is needed. Further solutions must be tested in order to scale up the findings from the mock trace-back/trace forward for national implementation, such as protocols, standards and agreements for industry/government collaboration before, during, and after food safety issues.

Table of Contents

List of Tables and Figures

Acronym List

BBOS - Blue Book Online Services
BT Act – Bioterrorism Act of 2002

CDC – U.S. Centers for Disease Control and Prevention
COOL - Country of Origin Labeling
CROMERR - Cross-Media Electronic Reporting Rule

DC – Distribution Center
DOD – Department of Defense

EPA – Environmental Protection Agency
EPCIS - Electronic Product Code Information Services
ETL - Extract, Transform and Load

FDA - U.S. Food and Drug Administration

GTIN – Global Trade Identification Number (a GS1 standard)

ID – Identification
IFT – Institute of Food Technologists
IT – Information Technology

NDA – Non-Disclosure Agreement

PACA - Perishable Agricultural Commodities Act
POC – Proof of Concept
POS – Point of Service/Sale

QA – Quality Assurance

RFID – Radiofrequency Identification

SOA - Service-oriented architecture
SPS – SharePoint Portal Services

TO6 – Task Order 6 (IFT conducted a review of food, feed and non-food industry product tracing practices in the U.S. and other countries, including systems and technologies used for product tracing and provided recommended best product tracing practices and a cost/ benefit review.)
TO7 – Task Order 7 (this task)

US – United States
USDA – US Department of Agriculture

XML - Extensible Markup Language

Background

This task order was used to explore the flow of information and information needs for the U.S. Food and Drug Administration (FDA) as they relate to technologies which may be used to track movement of food product back to the original source and forward through the supply chain, to increase FDA's ability to more rapidly and precisely track the origin and destination of contaminated food items.

The *Salmonella* Saintpaul Outbreak

Starting in April 2008, the Centers for Disease Control and Prevention (CDC) and FDA worked with state and local regulatory agencies and health departments, international authorities, and food industry groups to investigate a large, multi-state outbreak of *Salmonella* Saintpaul. Over 1,440 people were infected with the same genetic fingerprint of *Salmonella* Saintpaul in 43 states, the District of Columbia, and Canada between April and August 2008 (Figure 1). At least 286 persons were hospitalized, and the infection may have contributed to two deaths. The CDC investigation showed that tomatoes and peppers were possible sources of contamination and in June 2008 FDA advised consumers not to eat raw red plum, red Roma, and red round tomatoes, and products containing the raw, red tomatoes *unless* the tomatoes were from FDA's list of states, territories, and countries where tomatoes were grown and harvested that were not associated with the outbreak. FDA advised consumers that it was safe to eat cherry and grape tomatoes, and tomatoes with the vine still attached. However, consumers still did not trust that these tomatoes or other tomato varieties that hadn't been implicated were safe to eat. In July 2008, FDA lifted its advice to avoid raw red plum, Roma, and round tomatoes but damage had already been done to the tomato industry, which estimates it lost at least $100 million in sales. FDA again worked closely with CDC and state and local health departments to determine if implicated raw jalapeño and serrano peppers might be linked to the outbreak. Jalapeño peppers were traced back to a distributor in the US that received fresh produce from growers and packers in Mexico and samples of jalapeño peppers in one of their US warehouses contained *Salmonella* Saintpaul that matched the strain found on jalapeño peppers from an infected person's home. Samples of serrano peppers and agricultural water from a farm in Mexico also both contained the same genetic fingerprint of *Salmonella* Saintpaul as the outbreak. In late August 2008, the CDC announced that the *Salmonella* Saintpaul outbreak appeared to be over, and that jalapeño and serrano peppers grown in Mexico and associated with the *Salmonella* Saintpaul outbreak were no longer in circulation in the US market. Based on the CDC announcement, FDA then lifted its advice to consumers to avoid eating jalapeño and serrano peppers grown, harvested or packed in Mexico.

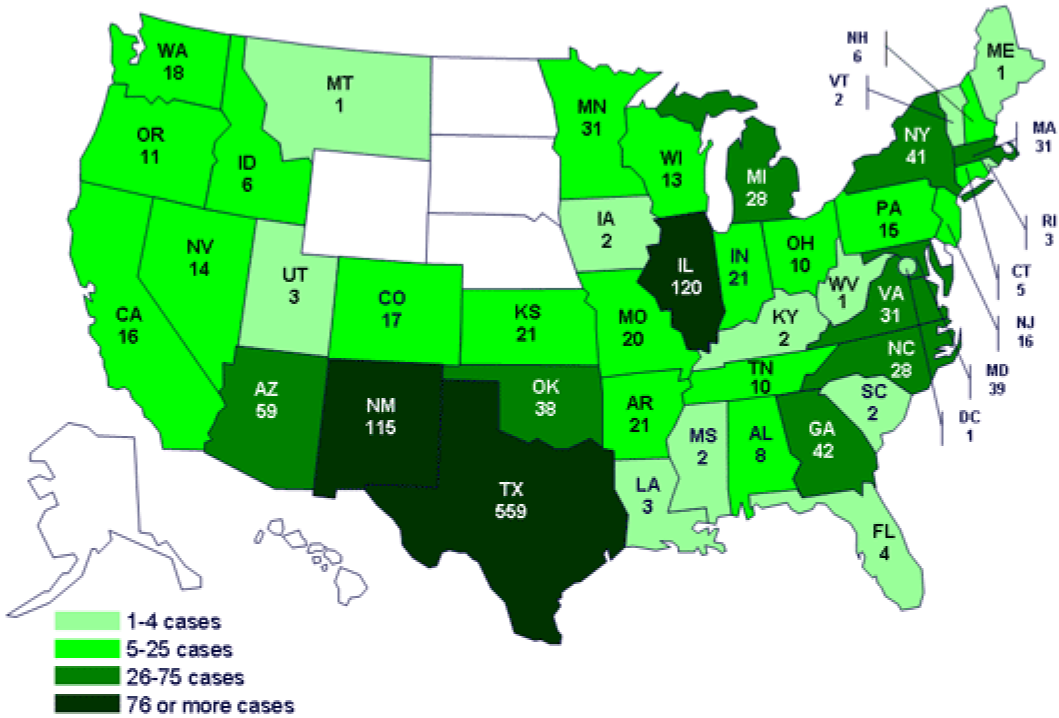


Figure 1 - Cases of *Salmonella* Saintpaul, by state, as of August 25, 2008

Harvard University Executive Session on Food Safety

Harvard approached FDA in the summer of 2008 to consider an Executive Session to explore key food safety issues, and food safety leaders at FDA agreed that this approach was worth pursuing. The 'executive session' methodology was developed at the John F. Kennedy School of Government in the 1970s to help cross-boundary collaborators jumpstart change on important public policy problems by reexamining strategy, gaining consensus on best next moves, and moving ideas to action in the form of proofs, pilots and tests.

FDA decided that the tomato industry would be a good focus for the Executive Session because of the recent 2008 *Salmonella* Saintpaul outbreak, and because tomatoes are primarily grown in two states, California and Florida, it may represent a more manageable supply chain of growers, shippers, packers, and distributors. FDA believed strongly that state regulators should be involved in the Executive Session because of their vital role in outbreaks. FDA also felt that tomato industry executives from various points in the supply chain, such as retail and foodservice, should take part in the Executive Session, as well as information technologists, who could offer insight on what role technology may play in any solutions discussed. (See Appendix A for Executive Session members.)

The Executive Session met four times: August and October 2008 at Harvard University; online by WebEx meeting in December 2008; and at FDA headquarters in January 2009 to reexamine the current strategy used to trace food backwards and forwards. At the August 2008 Executive Session meeting, it was discovered that mistrust and other hurdles to collaboration existed between industry and government due in part to the difficulties that arose during the *Salmonella*

Saintpaul outbreak. Many significant obstacles are faced by industry and government during an outbreak, such as unawareness of each others' resources and capabilities, which may make it difficult to work together. However, both industry and government found that they have a shared commitment to the safety of the US food supply and wanted to find solutions that would return trust and consumer confidence to the marketplace. It was discussed that possible solutions should maximize prevention, achieve continuous protection, and allow government and industry to work collaboratively to resolve food safety problems quickly once detected. Members of the Executive Session decided to focus on and try to improve the trace-back/trace forward investigation, since this process often requires a great deal of time and resources.

Improvements envisioned included increased speed and accuracy of government to trace-back and prevent foodborne illness outbreaks, which would possibly minimize illness by locating contaminated sources faster, reduce costs to market, and reduce lengthy, resource-intensive government investigations. Faster trace-backs would help government concentrate its limited investigative resources on high-probability supply chain segments sooner. Collaboration between government and industry could help to achieve a faster trace-back and to utilize the existing resources of both industry and government, such as data and other information.

At the October 2008 Executive Session meeting, tomato industry executives shared data collected during commercial transactions with Executive Session members. Much, though not all, one-step up/one-step back data, such as who shipped tomatoes, who received them, and when, in detail at times down to the lot number, is routinely collected in the course of industry commercial transactions. These data typically aren't shared with FDA unless specific requests are made, although FDA doesn't have a comprehensive awareness of what data exist unless they can take significant time upfront to have discussions with each company that is possibly involved to fully understand their internal tracing system and what data may be available. Therefore, FDA had never viewed these data together before and it appeared that the data could be useful for trace-back investigations. However, data are maintained by many different companies in the supply chain, often in different formats, and are in some instances treated as proprietary information.

Supply chains may be either open or more "closed". In closed chains, the customer (typically the retailer or foodservice corporation) typically have product requirements and relationships that extend past "one step back" and therefore may have better awareness of their supply chain. Due to industry practices such as first-in, first-out rotation programs at distribution centers, "one step forward" records may not be maintained and therefore few points of service may currently have data that could aid in a trace-back. Therefore, the hard-slog work on the front-end of trace-back investigations likely is still necessary. Investigators will continue to have to trace-back from the point of service, piecing together the trace-back from the front-end of the chain using currently available data. With existing data however, information technologists felt that advanced visualization tools might help reveal supply chains more clearly, whether on a trace-back or trace forward basis.

Based on these observations, the Executive Session identified five hypotheses they felt should be investigated further.

1. In situations where smaller points of service may be implicated who may not have all necessary data, the trace-back can still take less time if the rest of the industry has data and is able to show its supply chains, waiting for the top-down investigation to reach it. After manually tracing back to that point in the supply chain, visualization can then be used to more quickly complete the trace-back and then trace forward.
2. "Closed" supply chains (typically a point of service that determines which suppliers, distributors, and so on are trade partners and maintain data for the entire supply chain) could gain similar benefits by being able to provide data and therefore visualize their entire supply chain.
3. With increased awareness by government of what data industry maintains and what data would be useful in a trace-back, industry can provide necessary data electronically upon request to federal and state authorities, perhaps reducing some of the data burden.
4. Federal and state authorities may be able to conduct a trace-back investigation more quickly with the proper data visualized, allowing them to focus on implicated supply chains.
5. Collaboration between industry, government and other stakeholders is necessary before, during and after food safety issues, along with the proper agreements in place, the necessary data, and technology to visualize the data, for improved trace-back/trace forward investigations.

The Executive Session still needed to better understand the accessibility and usability of industry data. Participants agreed to gather historical data from two weeks in November 2008 showing transactions across a "closed" chain supporting a major chain restaurant, and a dispersed or "open" chain which supported diverse small-firm points of sale. Working with TIBCO and Microsoft information technologists, data were acquired, converted to spreadsheet format (often manually), and then exported to the Spotfire visualization software.

At the December 2008 Executive Session meeting, Executive Session members met through a web-based meeting to view the exported data via Spotfire visualization software. Industry leaders had struggled but succeeded in gathering data from most portions of the supply chains. In many cases, these were maintained in industry systems, such as accounting systems. Members learned that they were able to store available data together using off-the-shelf spreadsheet software and then export the standardized data to the Spotfire visualization software to better show areas of the supply chain that data were available for. Furthermore, it appeared that available industry data and the visualization tool could help state and federal authorities conduct faster trace-back/trace forward investigations. Significant issues remained in the design, development and implementation of the concept developed. A notable issue was the extensive up-front work to quality assure data received before they could be converted into spreadsheet format and then imported to the visualization tool. However, the Executive Session felt that the capability was worth exploring further.

The Executive Session convened at FDA headquarters in Maryland in January 2009 to discuss next steps. The Executive Session determined four areas that it hoped the mock trace-back/trace forward exercise would investigate:

1) Ease of Participation and Use. The Session wanted to understand the "ease of participation and use" for industry to submit and store data and for government to use the data.

2) Illuminate the Supply Chain. The Session wanted to understand whether and how, using available data and the visualization software, industry working with government could, upon request, illuminate designated supply chains sufficient to conduct trace-back/trace forward exercises.

3) Expedite Investigations. The Session wanted to understand whether the approach taken could expedite trace-back/trace forward investigations, and what the implications and requirements are for an improved trace-back/trace forward process.

4) Collaboration. The Session wanted to understand whether and how government and industry could come together to better understand the governments' trace-back/trace forward process and requirements; provide appropriate industry data to aid in a trace-back; determine information technology tools that could improve a trace-back; and initiate a mock trace-back/trace forward exercise using these resources to see if they allowed industry and government to better work together and if they could lead to faster trace-back/trace forward investigations.

Current U.S. Regulations Related to Produce Tracing[1]

It is important to understand current regulations governing produce, as well as other food products, to have a greater understanding of the current recordkeeping requirements food companies have. A goal of the mock trace-back/trace forward exercise was to use existing records from companies, to not request more from companies than they were already doing and to show that these data do exist and are usable for a successful trace-back/trace forward. Some regulations with existing recordkeeping requirements include the Bioterrorism Act of 2002; the Reportable Food Registry; the Perishable Agricultural Commodities Act (PACA); Country of Origin Labeling (COOL); and the Organic Food Production Act of 1990. There are also mandatory state regulations that may require members of the produce supply chain to maintain information that could be used for product tracing.

FDA/IFT Task Order 7: Mock Trace-Back/Trace Forward

Task Order 7 to conduct a mock trace-back/trace forward exercise was issued to IFT in early June, 2009, and brought together the various stakeholders needed to conduct an exercise that may lead to faster, more efficient trace-back/trace forward exercises.

[1] Institute of Food Technologists (IFT), Contract No. 223-04-2503 for FDA. Task Order 6, Traceability in Food Systems. Vol. 1, Technical Report. 2009.

A trace-back is a technique to track food items, including fresh produce, back to their source (growers, packers, etc.). Trace-backs can often be quite complicated because of the varying forms of recordkeeping and various types of information recorded by food companies, to name a few potential issues. A system to identify the source of fresh produce cannot prevent the occurrence of a microbiological hazard that might lead to an initial outbreak of foodborne disease, but the ability to identify the source of a product through trace-back can serve as an important component of good agricultural and management practices intended to prevent the occurrence of food safety problems. A trace forward traces a food product from its origin throughout the supply chain through its final distribution to points of service, and if conducted for an ingredient, can show finished products that the ingredient was added to and where they were distributed. Trace-backs for contaminated products or ingredients are used to show points of commonality along the chain and to illuminate where products or ingredients may have first become contaminated and trace forward to show where the products or ingredients went. This may allow a recall of a targeted product to prevent an outbreak or may prevent an outbreak from growing.

Improving FDA's ability to trace a contaminated product back to the source of production would allow the agency to conduct more rapid and thorough investigations and to do so more efficiently. It would also allow producers to more precisely identify the source of a problem in order to improve production practices and could help narrow the scope of recalls by more quickly identifying the specific plant or country of origin. Reducing the time required for a trace-back/trace forward may better protect public health, can help reduce the economic hardship affected industries face, and maintain consumer confidence in the US food supply following an incident.

From June to September 2009, IFT, with Harvard, TIBCO, and Microsoft under subcontract (see Appendix B for Key TO7 Personnel), FDA, some original Executive Session members and others met to frame out and execute a mock trace-back/trace forward exercise using industry data from two weeks in November 2008, a data collection tool to facilitate the movement of existing industry data to a common database, and visualization software.

Scope of Work (As Provided to IFT by FDA)

1. IFT shall organize and implement a mock trace-back/trace forward exercise, in which FDA and other subject matter experts will participate, utilizing a collaboration platform to share data from various sectors of the tomato industry to establish whether the data sets and technology platform allow for expedited electronic trace-back/trace forward of tomatoes.
    - Data from the producer, packer, distributor, and retail sectors should be utilized in the exercise.
    - The data available on the platform should allow for multiple trace-back/trace forward scenarios, ranging from simple to complex.
2. IFT shall provide a report (both hardcopy and electronic copy on diskette or CD-ROM in Microsoft Word compatible format and HTML format) of its evaluation of the mock trace-back/trace forward exercise, including key issues and lessons learned in platform design, technology needs, data gaps, and data management.

3. IFT shall provide recommendations about the utilization of similar technology platforms to store and manage industry product tracing data on a national scale for fresh produce and its applicability to other food items.

Approach to the Scope of Work

Approach to Scope of Work 1. IFT organized and implemented, using subcontracts to Harvard University Kennedy School, Microsoft, and TIBCO, a mock trace-back/trace forward exercise, in which FDA and other subject matter experts (including trade association, industry, and state representatives) participated, using collaboration to share data from various sectors of the tomato industry and technologies to visualize a mock trace-back/trace forward of tomatoes. Representatives from IFT, Harvard, TIBCO, Microsoft, FDA, some original Executive Session members and others met on three separate occasions: June 24, 2009 to finish framing out the mock trace-back/trace forward and building the scenarios; July 15-16, 2009 to finalize data/system preparations and identify any further needs for the upcoming mock trace-back/trace forward; and, July 22, 2009 for the actual mock trace-back/trace forward. A fourth meeting was held by the Executive Session and FDA after the task had concluded and thus is considered separate from this task. Attendees at every meeting included: (*listed alphabetically by organization*) Ed Beckman (CA Tomato Farmers); Ana Hooper (Darden); Sherri McGarry, Ingrid Zambrana, Sandra Hanson (FDA); Zach Tumin (Harvard); Sarah Ohlhorst (IFT); Susan Conway, John Ylinen, Josh Wall (Microsoft); and, Tim Wormus, Sean Conners, Dave Athey (TIBCO). Data collection, standardization and quality assurance, storage, and upload into the visualization software were necessary for the mock trace-back/trace forward, and the approach to each step is explained in more detail below.

Data Collection (Microsoft):
Prior to IFT being issued this task in June 2009, tomato supply chain representatives from the retail, grower and producer end who had been involved with the earlier Executive Session meetings volunteered to provide and gather data. Existing data were collected from a two-week period in early November 2008 from a "closed" supply chain supporting a major chain restaurant, and a dispersed or "open" chain which supported diverse small-firm points of sale. The restaurant supply chain is considered closed because the point of service (restaurant) determines which suppliers, distributors, and so on are trade partners, such that they exert some level of control more than just one step back. Data collected had to be representative of the tomato industry and sufficiently connected so that supply chains as far forward as possible were represented. Expanding on the original data set, tomato industry members and the restaurateur worked with industry partners to acquire data for the repacking sector, as well as the distribution and retail segments of the supply chain necessary to complete the supply chain. The restaurant representative contacted various points in their supply chain, such as distributors, growers, and so on, and asked them to submit data for this exercise. Data covered suppliers from CA, FL, and Mexico, and distribution throughout the entire US.

A grower representative and other members of the Executive Session reached out to points in the open supply chain such as repackers and distributors, and asked them to submit data for this exercise. Data collected came from California and Florida; data from Mexico growers was provided through their US distributors; and, the repacker data included product produced in

Canada and Mexico. To more easily manage the amount of data from large distributors and retail sectors that deal with product all over the US, Executive Session members decided to limit the focus of the exercise to the Texas repacker locations to make the exercise manageable and to mirror the initial stages of the 2008 *Salmonella* Saintpaul outbreak. The geographic distribution of the repackers included: Texas, Louisiana, Oklahoma, and Illinois.

Data elements to be collected had also been previously determined by Executive Session participants and reviewed by FDA. Although they differ slightly for each sector, they generally include company name, city, state, lot identifier, and product name/description (Table 1 shows data elements collected). Blue Book number was also a data field collected for this task. Blue Book conventions were used to ensure that participants were correctly and uniformly identified, assuming Blue Book listings are accurate and that actual locations are listed, rather than a company's headquarters. Approximately 10,000 produce companies are listed on the Blue Book Online Services (BBOS). The company has a comprehensive database with a search engine that provides real-time information on produce sellers, buyers, transportation and supply firms that are located in the US, Canada, Mexico, and other international locations. There is no charge to list basic company information in the Blue Book so many companies take advantage of this service and become "registered users." Basic company information includes name, address, phone and fax number, e-mail address, website, as well as certain licenses, classification and commodity information. Additional information can be added to a listing for a charge. Registered users are assigned a 6-digit Blue Book number.

Table 1. Data Elements Captured for Task Order 7

Grower:	**Packing house:**	**Repacker:**	**Distributor:**	**Point of Service:**
Name	Name	Name	Name	Store #
				Address
City	City	City	City	City
State	State	State	State	State
Blue Book #	Blue Book #	BB # (if any)		
Lot ID	Lot ID	Input/ Output Lot #	Input/ Output Lot #	Input Lot #
		Repack #	Repack #	
Harvest Date	Pack/Ship Date	Ship Date	Ship/ Receive Date	Receive Date
	Product Description	Product Description	Product Description	Product Description
	Product Code	Product	Product	Product
	Quantity Shipped	Quantity	Quantity	Quantity
	Pounds	Pounds	Pounds	Pounds
		Shrink	Shrink	Shrink

Repack run number was also a data element collected for this task. Repacking may be one of the most confusing components of the fresh tomato supply chain. Tomatoes are typically assigned an initial lot number upon arrival at the repacker, with that lot ID amended each time the tomato is graded or sized. This may include multiple grading runs. Therefore, the repack run number provides the pedigree of the original lot as received from the shipper, the processes undertaken to finish the product for the customer, and the lot ID for the final product shipment to the customer. It should be noted that this data element used for internal product tracing may be unique to the tomato industry, where repacking for quality occurs on a regular basis, and this term may not be found for other produce lines or food items, which may make internal product tracing more difficult for other segments of the food industry.

Data collection, for this exercise, focused on spreadsheet templates. Although many user-friendly aspects were envisioned for the data collection tool, data collection for this exercise was primarily manual, and not electronic or automated. Data were collected by providing points along the supply chains with the data collection tool, which was a spreadsheet with required data elements labeled, and having companies fill in the required data elements, using data from existing accounting systems or any other means of data storage the company had. Data were often entered manually into the spreadsheet template before being submitted to Microsoft. Microsoft then entered data from spreadsheets into a larger database for further quality assurance, before data from the database was uploaded to the visualization software. Figure 2 below shows a screen shot of the visualization software linking back to a supply chain database Microsoft developed for this task.

The goals for the data collection tool include a tool one that can be used by any business regardless of the type of electronic (or manual) data storage they used. The web-based tool was designed to consume a locally produced spreadsheet (or other file types such as XML) and produce an output file that can be automatically consumed by the database. Once data are in the database they can be utilized for any number of output reporting methods. The data collection tool was designed to accept comma delimited (spreadsheet) output from the various accounting systems, as this was the easiest method for the participating companies and some already have data in a spreadsheet format. The tool is also able to take data from XML files, although no companies provided data in this format. All the potential options for data collection were explored for this exercise, but the information technologists built out only those methods necessary for the mock trace-back/trace forward exercise.

Data Quality Assurance (Microsoft):
Although data elements were identified, no standard for submission of data was set, so quality assurance and standardization of the data was necessary. Quality assurance of data was necessary after capture to ensure that all fields required for this task had data and that the data appeared to be accurate. Standardization of data fields such as dates, quantities, and addresses was also necessary for entry to the spreadsheet and then visualization tool. Microsoft thoroughly reviewed all data received by hand to clarify and eliminate gaps or variances in the data. For example, companies may record date and/or time, address, quantity or other data fields in different ways and one standard method of displaying data had to be selected and used for all data to be entered into the database in the same way for display in the visualization exercise.

Graphical & Tabular View of Chain

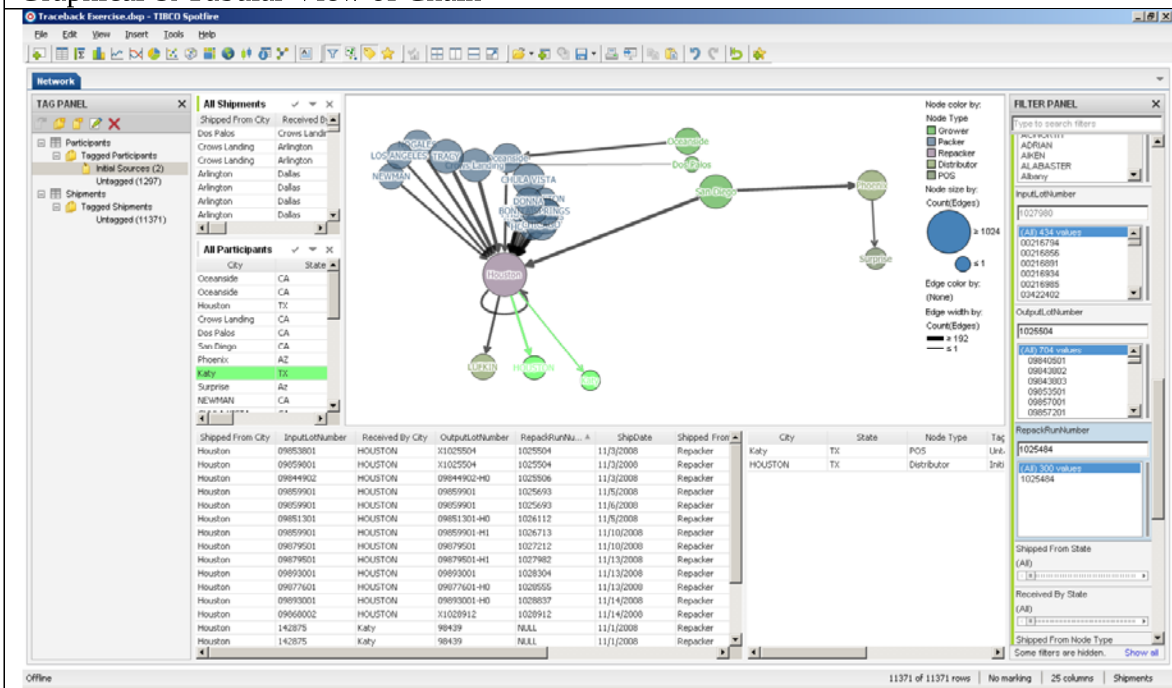


Figure 2. Microsoft screen shot of database information, linked to visualization software

Once in the database, quality assurance and standardization of the data was necessary to link diverse individual transactions by various lot numbers (incoming lot numbers to repack run numbers, then to outgoing lot numbers) across multiple segments of the supply chain. Validation of data integrity was conducted manually by Microsoft for the immediate purposes of this exercise. For the global, scalable needs of the food industry, the data collection tool was designed to help any user train their IT system to "dump" data from fields in its unique database (which were peculiar to the accounting, purchasing or logistics systems), into a single standardized set of fields. For example, no matter how an IT system might store an address, it would always export via the tool into the fields, "### Street, City, State, Zip." Once standardized, and once the system was "trained" to how data would be entered, the tool could import those data any number of times (from any number of time periods) into a master database, on demand (without 'retraining' the tool). Source data must be introduced to the data collection tool and a company only needs to train the tool once; after that, data conformation routines should be automatic.

Data Presentation (TIBCO):
Data were imported from spreadsheets into Spotfire's visualization software tool. With the data standardized, quality assured and imported into spreadsheets, the Spotfire visualization tool could acquire them. TIBCO Spotfire software offers a visual, interactive environment to perform ad hoc analysis and to rapidly capture, share, and author analytic applications. Spotfire is used to analyze data containing both qualitative and quantitative variables and perform such tasks as discovering relationships, examining distributions, and spotting trends and outliers.

Using received/shipped data from tomato transactions, Spotfire provided a visualized depiction of the network relationships and supply chain transactions contained in the data. Specifically, TIBCO assisted in configuring guided analytic applications; demonstrated proper use of the configured application and mentored users; and, prepared technical documentation for illustrating how to use and deploy the application.

The primary visualization capability used for the mock trace-back/trace-forward exercise was Spotfire's network graph, although Spotfire offers other visualization capabilities such as bar charts, pie charts, line charts, maps, tree maps, and scatter plots, that each address analytic questions. In addition, Spotfire provides filters that are automatically configured to represent data variables. Adjusting the filters changes the views of the visualizations based on the resulting selected data.

A network graph consists of nodes and edges. In the mock trace-back/trace forward, the nodes depicted the different locations such as the growers, distributors, packers, re-packers, and points of sales. The edges identified relations between the nodes. In Spotfire, a mouse click on a node displays both the records of data that identify the node and those that identify all of the node's inbound and outbound relationships. A mouse click on an edge displays the records that identify the relationship between nodes. Figure 3 below shows a screen shot of the trace forward used in the mock trace forward exercise. The trace forward shows the path of potentially contaminated tomatoes from a Nogales, Arizona repacker (dark green circle) to all other locations product had been shipped to. Additional screen shots from the TIBCO visualization software used in the mock trace-back/trace forward exercise can be found in Appendix C.

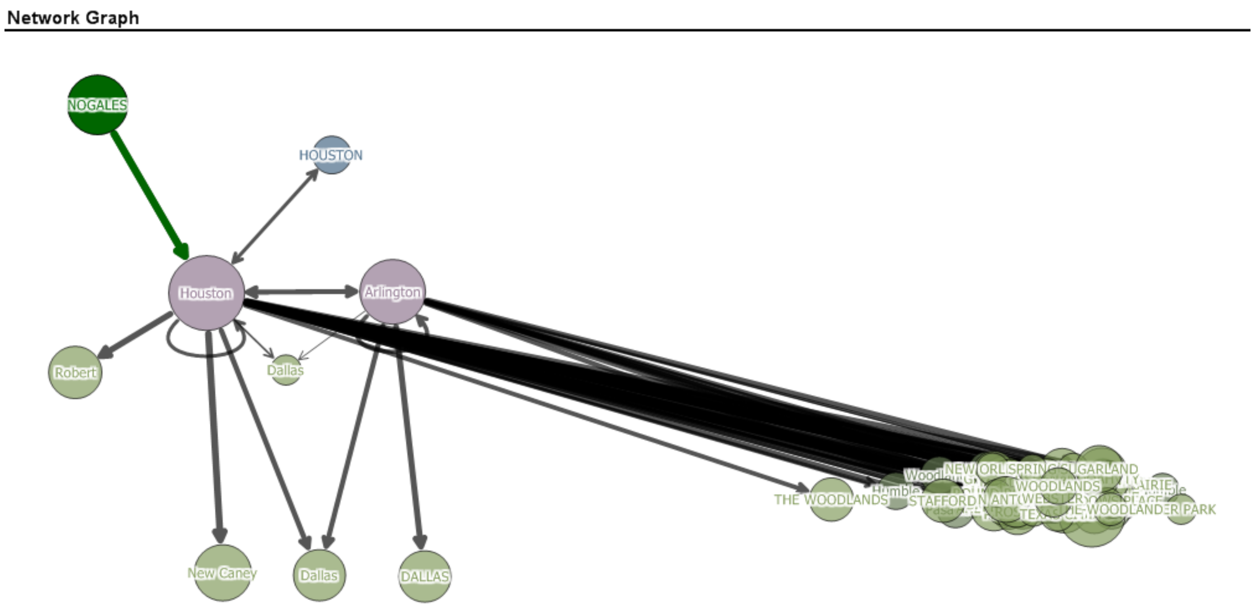


Figure 3. Trace forward shown via TIBCO’s visualization software

At the mock trace-back/trace forward exercise, FDA subject matter experts were asked to lead those present in a walk-through of various trace-back/trace forward scenarios using the data acquired. The trace-back/trace forward scenarios included an open and closed supply chain. The FDA subject matter experts were assisted by a "guide", who was a produce industry subject matter expert and also familiar with trace-backs and trace forwards. In addition, an IT expert was used to maneuver the visualization for the data, responding to the directions of the others involved.

Approach to Scope of Work 2. IFT shall provide a report (both hardcopy and electronic copy on diskette or CD-ROM in Microsoft Word compatible format and HTML format) on its evaluation of the mock trace-back/trace forward exercise, including key issues and lessons learned in platform design, technology needs, data gaps, and data management.

Summary reports provided to IFT by Harvard, TIBCO and Microsoft were utilized for this final report, provided by IFT to FDA. This final report also includes IFT's sole input on the mock trace-back/trace forward exercise and the comparison of product tracing technologies, described below.

Approach to Scope of Work 3. IFT shall provide recommendations about the utilization of similar technology platforms to store and manage industry product tracing data on a national scale for fresh produce and its applicability to other food items.

IFT conducted a comparison of the platform used to evaluate the data in Task Order 7 (herein known as TO7 platform) against other product tracing technology platforms in order to determine the effectiveness of the TO7 platform and other platforms considered, and to provide information that can be used to develop recommendations on criteria that leads to an effective technology platform.

IFT planned to examine approximately eight to twelve technology platforms (including the platform created through this task), and evaluate them based on predetermined factors. IFT previously conducted an in-depth review of product tracing technologies that are currently being used or that may be used by food companies in the U.S. or abroad for FDA as part of Task Order 6. IFT referred to information collected previously in Task Order 6 to select seven to eleven platforms to compare against the TO7 platform. Due to the expedited nature of TO7, IFT gathered information from 10 technology solutions providers that we had worked with previously to compare against the systems used in this task. Although there are many companies that currently provide product tracing technology solutions, IFT ensured that the limited number of platforms selected for comparison are representative of the wide array of options available.

IFT considered numerous factors when comparing product tracing technology platforms, such as the breadth, depth, and precision of tracing technologies from source to point of sale for all food or feed products in the food continuum, including products that may or may not have a label and/or lot code associated with them and ingredients that go into multiple finished products. Priority evaluation was given to the ability to track fresh produce. The tool used to gather data from companies can be found in Appendix D. IFT also compared factors such as:

- The ability of each platform to store data, including how much data can be stored and for how long;
- The management of all data;
- The ability of the system to accept data in multiple formats;
- How data are shared and how access is enabled or restricted;
- If product tracing extends from points of service to points of processing and production through point of sale;
- The precision of the platform to locate a source of contamination, product's movements or characteristics and more;
- How quickly information (source of contamination, etc.) can be provided and communicated; and,
- The applicability of the platform to national and international food industries.

IFT used the Task Order 6 "Traceability (Product Tracing) in Food Systems" expert panel to ensure that critical factors that may lead to an effective product tracing technology platform were considered. The expert panel reviewed the proposed factors and other platforms proposed by IFT to determine final comparison criteria before it was collected. IFT used comparative views of alternative platforms to provide an overall assessment of the platforms used in the mock trace-back.

Results of the mock trace-back/trace forward exercise

Scope of Work 1. Large amounts of data were collected for this task. Companies that data were collected from were identified by industry experts participating on the task, or the industry experts themselves supplied data for their company. Data acquired for this task were in excess of 25,000 confidential business records of tomato transactions from the targeted two weeks in early November 2008. These data specifically came from four California growers/shippers who shipped 214,657 cartons (5.5 million pounds) during that period, as well as additional growers from Florida, Mexico, and other areas. Repackers with operations in six states contributed data, with 8290 records from two Texas locations used in the exercise. Destinations for the product included six wholesalers, two repackers, a variety of independent retailers, government commissaries, and small foodservice outlets, along with four retail distribution centers, two broadline foodservice distribution centers, one fast food chain distribution center, and one restaurant chain distribution center. Almost equal amounts of data were used from a restaurant's more closed supply chain.

While essential data are routinely captured in commercial transactions, industry data does not currently exist in a state of "ready usability". It is sometimes considered proprietary and held by various segments of the supply chain, and therefore not available at the same time or in the same format across all segments. Data currently captured and maintained by firms for commercial transaction often has common elements which can be standardized for export to databases, but this requires significant time and effort to quality assure and standardize data to make it useable for the purposes of a trace-back.

Data capture tools currently exist or can easily be developed to facilitate the export of data from existing industry systems such as accounting systems or spreadsheets, into a larger shared database. With appropriate agreements in place, a data collection tool can greatly facilitate the sharing of commercial transaction data, help to quality assure and standardize data, and ready it for import to a visualization tool.

The data collection process requires standardization and quality assurance of data at many steps. Not all data that was captured by the project could be used, as gaps or breaks existed in certain supply chain segments where data had not been collected, causing some available data to lose their utility for the purpose of trace-back/trace forward. Consistency and accuracy of data is not assured without checks and audits. Security of data is also necessary to assure, although industry appeared willing to participate in this task since it was understood to be for the betterment of its business interests.

Industry participants in the mock trace-back/trace forward exercise who contributed data reacted very positively to the process and tools, with one stating: *"Our initial data collection for the pilot took over 2 days (for each pass) to collect, including administrative and IT time working with the information technology team. Once the data collection tool was built, the data were aligned, translated and conformed in minutes."*

Once placed in the visualization tool, data can be viewed simultaneously and various lot numbers can be used to view the movement of tomato lots over diverse supply chain segments, even as pallets are broken apart, cartons repacked, and further distributed. The exercise team was able to use the Spotfire visualization tool, together with the Microsoft generated database, to gain insight to the tomato supply chains and the potential links in the supply chain, and inquire further into particular transaction records of interest. Various segments of tomato supply chains were able to be linked together sufficient to visualize trace-back and trace forward from grower-shippers to and from distribution centers. For the purposes of this exercise DCs were considered to be the "points of sale" since that was a data gap.

Using the TIBCO Spotfire visualization software, a trace-back was visible by identifying a node or group of nodes at the end of the supply chain. Using filters settings in Spotfire and the view provided by the network graph, various nodes or groups of nodes were able to be rapidly identified, visually marking a point in the supply chain with commonalities. This visualization can help target the node or nodes where investigational resources should be deployed to determine the potential source(s) of contamination. In the cases where multiple nodes are identified, the clusters of illness could be traced back to a point in the supply chain where the product converged to a single source. For example, clusters of illness at multiple points of sale were traced back to a single distributor. Government and industry members were able to better understand the supply chain based on its visualization. They could identify all the major segments of the supply chain to the extent that the data were available - grower/shippers, packers, repackers, distributors/points of sale. With some manipulation, they were able to easily identify them in a network display, and see the web of transactions connecting them. In a live demonstration, information technologists were able to click on any single transaction as government or industry experts requested, or any particular network node, to display the underlying (alpha/numeric) transaction and identity data from the spreadsheet.

A trace forward was visible by selecting a node near the beginning of the supply chain. By following the edges in the network graph from the selected node to its connected nodes and their subsequent connected nodes, the path along the supply chain is easily identified. As in the trace-back, the Spotfire filters facilitate a view of the steps in the trace forward. Spotfire's filtering capabilities provided the capability to quickly show the trace of a cluster of illnesses in the supply chain. Every variable in the data automatically has a dynamic filter in Spotfire. Therefore, a view can be immediately updated to show only those nodes and edges that correspond to a filter setting, such as a range of dates or selection of cities.
Collaboration and a shared mission were key to the mock trace-back/trace forward exercise, as they help develop trust and respect between all stakeholders. Industry commercial data are not self-explanatory, often even to other industry members, and collaboration is necessary to help make sense of all the individual pieces of information that are collected. Industry and government often do not have insight to each other's processes, but collaboration can help them to more quickly understand and reach a shared goal. Although significant legal issues may arise, industry and government can collaborate without violating or compromising their trusts. Government/industry collaboration before, during and after food safety issues is essential to pool assets such as industry's data and business knowledge, government's investigative knowledge and legal authority, and technologies in an effort to improve national food safety. Technology is a critical enabler of collaboration, and is readily available, although training, costs, and other requirements must be taken into account when new technologies are considered.

Key Challenges

Although many data gaps are identified below and various other issues arose and were dealt with, certain challenges that occurred during this task stand out.

One challenge was a feeling of "scope creep" among some participants. This had much to do with the timing of new ideas for the project, in addition to the continued education of industry on the process of an FDA trace-back. A goal from the Executive Session meetings was to conduct a broad exercise to see what could be learned and what was possible with data storage and analysis tools. The more specific scope of work requested by FDA for this task included a trace-back/trace forward for a complete supply chain, so the need to identify and capture additional data from distributors required to run a complete trace-back/trace forward for an entire open supply chain became a focus that was only partially complete the day of the mock exercise.

Once data for the extended scope were captured, a full trace-back/trace forward for the entire open supply chain could not be run due to a lack of data fields that would link each step of the trace-back/trace forward path. There were linkages between the majority of the supply chain, but data collected from distributors linked back to a different repacker than was used for this task, not allowing the supply chain to extend past repackers' shipments out to distributors (to POS) and excluding the distributor data from use in the mock trace-back/trace forward exercise. This was an oversight due to a lack of understanding of all the data required for an accurate trace-back, but the lack of time before the mock trace-back/trace forward did not allow for a correction of obtaining data from the distributor that did link to the repackers used in the mock trace-back/trace forward.

Other challenges occurred with data collection and quality assurance. Data collection was a challenge in that some companies were hesitant to share data and doubted the security and confidentiality of their records. Non-disclosure agreements were required by some companies. There was a lack of understanding of key data elements that were needed from specific points along the supply chain, and therefore the lack of data fields to link each step in the trace-back/trace forward. At times companies had to be consulted numerous times to gather accurate data from correct data fields and to confirm data following quality assurance. Significant quality assurance of data and standardization of data fields such as dates, quantities, and addresses was necessary before data could be uploaded into the visualization software. QA looked for inaccuracies in data such as errors in human entry to ensure that all fields required for this task had data and that the data appeared to be accurate. Requiring standardized formats for key data elements that enable product tracing could alleviate the need for so much time and effort required during data collection and quality assurance.

The trace-back/trace forward exercise still occurred with the on-hand expertise and visualization capabilities available, and provided those involved with the exercise valuable information. The trace-back began at the repacker, rather than at a point of service, and trace-forward could only be conducted to the distributor (not POS) but this was still a valid test. The visualization systems helped demonstrate a capability for analysis of data so that insight gained can be immediately put to action.

The mock trace-back/trace forward was a success, in that it showed the value of industry and government working in collaboration, sharing data when possible, and achieving faster, visual trace-back/trace forward results. However, given that this was only an exercise of limited scope, there were portions of the system that were not tested. These included other food items, other technologies, imported food products, and a broader geographical area. These additional “lessons learned” are noted below.

Lessons Learned

Platform Design

Once data were compiled, checked for quality assurance, and standardized if necessary, the data were able to be viewed in spreadsheet format and with the visualization tool. Government and industry members were better able to understand the tomato industry’s supply chain and dealings using the visualization tool. They could identify major segments of the supply chain through data in the spreadsheet or the visualization software. They were able to easily identify points in the chain in the visual network display, and see the web of transactions connecting one to another. In a live demonstration, information technologists were able to click on any single transaction as government or industry experts requested, or any particular network node, and display the underlying (alpha/numeric) transaction and identity data from the spreadsheet. Training would be required for FDA or others to use the visualization tool, but it appears easy to maneuver and appears to allow for ease of data analysis. The tools used for this task are very similar with what other product tracing technologies offer, in terms of security, access, ease of use, and so on. Although the technologies used in the exercise are similar to what other technologies offer, the various technologies currently have no interoperability to communicate with one another if all were used for product tracing.

Technology Needs

Hardware needed for this task included an internet connection and computer, and software for this task was provided by TIBCO and Microsoft. A 3-year license for 5 users of the TIBCO visualization service and limited use of a Spotfire server (with 1 year of maintenance) is estimated at $59,404 (which includes one training class). Microsoft roughly estimates the software licensing to deploy the system into production will cost between $700,000 and $800,000 for up to 500 users. This estimate takes into account the costs for the server and client side software, as well as costs to make the system securely available across the internet. Once the final requirements are assessed and there is a better understanding of the user base that would need to access the system (i.e. software they may already have licensed) a more accurate estimate or quote, along with a bill of materials, could be compiled for FDA.

The use of these technologies by smaller companies must be considered. Some smaller companies may not use an electronic system at all, or they may have limited resources and staff to devote to hardware, software, and other needs for data capture, availability, and readiness to use if this or other pilots were to move forward to national implementation.

Another consideration is the use of primarily electronic records in a trace-back/trace forward. As part of the Bioterrorism Act, FDA requires most food companies to maintain one-step up and back records for up to two years, depending on the shelf-life of the food. The Act allows companies to maintain records in either paper or electronic form as long as they are available for inspection and copying by FDA within 24 hours of an official request. The admissibility of electronic records in judicial proceedings arose during discussions of the Task Order 6 expert panel, and is applicable to this task as well. Whether a particular document is admissible in a court of law depends on many factors, including whether it may be authenticated as to source. Other federal agencies allow the use of electronic documents for reporting purposes, and are able to use the actual electronic documents or information from them in enforcement proceedings. For example, the Environmental Protection Agency (EPA) allows electronic reporting under the Cross-Media Electronic Reporting Rule (CROMERR), which went into effect on January 11, 2006.[2]

Data Gaps

Data were at times somewhat difficult to obtain as a few companies worried about the security and confidentiality of their data, and non-disclosure agreements had to be in place before these companies were willing to provide their information**.** If this tool were to be used on a larger scale, data collection and quality assurance must be further discussed, as the methods used for data collection and QA for this task would not scale-up to a large number of companies submitting data**.**

[2] Institute of Food Technologists (IFT), Contract No. 223-04-2503 for FDA. Task Order 6, Traceability in Food

Data collection represents a challenge for the industry as well as FDA. A major part of this challenge is how to provide data. Identifying the correct data fields needed for this task was a challenge that required input from the exercise team, FDA and industry participants. Data gaps such as inaccuracies and blank data fields were found through quality assurance and required some participants to resubmit data numerous times. Some data were not used for this task if these data gaps were not filled prior to the task, or if data could not show linkages between points in the supply chain. The key component that is likely to change industry by industry is the data field that helps to establish the chain of custody from the grower to the point of service; in the case of the tomato industry, this is often the repack run number. Key to the enablement of a more effective trace-back/trace forward will be the definition of core (common areas across all industries) and unique data fields, and standardization of how these data fields are recorded, and submitted to FDA upon request.

It is important to note that original data collection began before the task was issued to IFT and was not originally intended to meet the scope of work FDA set forth in the contract. Some data limitations resulted from this.

Although farm to point of service data were available from the closed chain for the mock trace-back/trace forward, data collected from distributors in the open supply chain from November 2008 could not be used in the mock trace-back/trace forward. This data linked back to a different repacker than was used for this task, and was unable to show linkages in the supply chain past repackers' shipments to distributors (to POS). Therefore, the mock trace-back/trace forward exercise for the open supply chain was developed considering retail distribution centers as point of services, and as such, the available point of service data were not employed in the exercise.

Data from the closed supply chain covered distribution across the entire US, but the Executive Session members decided to limit the focus of the open supply chain mock trace-back/trace forward to Texas repacker locations to make data more manageable. The geographic distribution to points of service from the repackers included: Texas, Louisiana, Oklahoma, and Illinois. Since the trace forward goes to other states, the Executive Session members felt that this was still a good representation of supply chains, although repacker data used was geographically limited. The assumption is that this system will be able to examine real-time data in the future and will be able to process data from multiple segments of the industry covering multiple states, countries, etc., at the same time but this was not tested in this exercise.

The mock trace-back/trace forward exercise was able to take product to the U.S. border, but did not cross over the border, although some data from other countries was collected. In actual fresh produce trace-backs conducted by FDA, they have crossed international borders and require a tool that can do this as well. There was not an opportunity to show how this system could handle international trace-backs, but one can assume that with proper recordkeeping, maintenance and uploading, a trace-back across the border would still be possible with this system and would flow the same as domestic trace-backs that were tested.

Because of limitations discussed above, the multiple trace-back/trace forward scenarios tested were primarily simple, not complex as requested in FDA's scope of work, #1. These issues underscore the need for collaboration on the development of a standardized format for collection of data from grower to point of service. Learnings such as this from this exercise can be applied to improve future pilots.

According to the technology companies involved with this task, they can use any form of data and data can be easily uploaded into their systems. While this is true in part, the exercise showed that standardization and QA of data is an important issue that must be conducted before the system can work to its fullest and this must be considered for scalability of the exercise. Data gaps were identified requiring Microsoft to thoroughly review all data received and work to clarify and eliminate gaps. For example, QA provided an opportunity to confirm possible inaccuracies in data and fill blank data fields. Also, companies may record date and/or time, address, or quantities in different ways and one standard method of displaying data had to be selected and used for all data to be entered into the database and be used in the visualization exercise.

Another consideration is that since much of FDA's current trace-back/trace forward processes are manual, they should be analyzed before optimum technology improvements can be recommended. Currently FDA investigators often go to a site and physically identify potential data and bring the physical documents (copies) back for analysis. This can involve thousands of individual records across numerous business establishments. Current technology processes are designed to take advantage of the deep in-depth knowledge and expertise of the FDA staff. One goal of any future technology solution should be to empower FDA staff's skills, allowing them more time to do analysis and investigation as opposed to data collection functions.

Although the mock trace-back only examined data from the tomato industry, there is also an assumption that the system will apply to other fresh produce and non-perishable food items with appropriate recordkeeping, maintenance, and upload**.** However, one must assume that all companies receiving a data request would be able to provide complete records quickly, but this exercise did not go beyond the tomato industry and it is not known if the same types of records are available from other segments. If companies cannot produce the data for all fields identified, the track-back/trace forward is made more time consuming and difficult.

Data Management

The mock trace-back/trace forward examined supply chains with data collected from many sectors, including producers, packers, distributors, and retail sectors**.** Microsoft completed data collection, coordination and upload for all sectors for this task, but companies will be required to coordinate and upload their own data in the future, if this system were to be used**.** Where data will be submitted to and who will manage the data is something that must be considered.

Non-disclosure agreements were requested by some companies providing data for this task. Security and privacy of data must be assured, and future requirements for collaboration between industry and government to upload and view data using this system must be explored, as there may be legal constraints involved.

Results of Technology Platform Comparison

Scope of Work 3. The results of IFT's evaluation of eleven technology platforms (including the platform used for this task) are outlined below. IFT gathered information from 10 traceability solutions providers to compare with the systems used in this task. It should be noted that the information from the traceability solutions providers is self-reported and as such, IFT cannot verify the accuracy or validity of the responses. The tool used to gather information from traceability solutions providers is found in Appendix D. It is also important to note that assumptions specified for the tool were that 1) FDA has specified what data are needed for analysis during an outbreak, and 2) All members of a supply chain are willing to contribute data to your system. There are many product tracing solutions available and the 11 systems reviewed here represent only a small sample, but IFT did attempt to ensure that the limited number of platforms compared below were representative of the wide array of options available.

Company's descriptions and product tracing offerings varied greatly – some offer cold-chain management, others Software as a Service, perishable product tracing, one a non-profit standardization service and so on. While some offer primarily product tracing solutions, others include product tracing with their primary service, and others provide a venue for companies to share information, but do not further evaluate the tracing this shared communication provides. In all instances companies retain ownership of data even after it's submitted to traceability solutions providers. Data are either housed on the food company's internal system or submitted to an outside server. In one instance data were kept on an RFID tag, as well as an internal system. The majority of food companies send data to an outside server, which is usually maintained through the traceability solution provider and may be located at their facility. The systems used for TO7 would allow companies to retain ownership of their data after it's submitted to an outside server (however, details for use beyond this exercise, such as who would maintain the server and location of the server were not determined).

Two of the 10 companies house data on Electronic Product Code Information Services (EPCIS) repositories. EPCIS is an EPCglobal tool for capturing and sharing Electronic Product Code™ (EPC)-related information between trading partners. EPCglobal develops EPC standards and other tools to support the use of Radio Frequency Identification (RFID). The goal of EPCIS is to enable dissimilar applications to leverage data via EPC-related data sharing, both within and across enterprises. Although EPCIS offers an EPC standard, companies can use the interface without using their standards. Companies can use EPCIS by downloading the package online at no cost and integrating EPCIS with their existing system. The interface is able to link two or more systems together to communicate and share information by aggregate events. Broader than RFID, EPCIS offers a way to share massive amounts of visibility data. The EPCIS Validation Portal is a self testing service, accessible through the internet. The tool provides a fast and simple way to visualize the data content of clients' EPCIS information, and checks for errors. It analyzes the data to ensure a uniform vocabulary and other standards across the supply chain. EPCglobal has also been developing a Core Business Vocabulary which provides companies with a mechanism to define industry terms. Through these repositories, industries can set standards and share information. EPCglobal drives a standardized approach which is scalable for large and small companies and has no proprietary recommendations (does not require use of GS1 standards to use EPCIS).[3]

All systems, including those used for TO7, could work for all sectors of the food industry, and all are able to receive data from sources outside the US. Many of the 10 traceability solutions providers (other than those piloted in this task, which are not currently commercially available) have international users or systems that are currently operating in other countries.

The majority of the companies who responded (80%) can accept both non-electronic and electronic records into their system. The same holds true for the systems used for TO7, although data for this exercise was submitted via spreadsheet and manually entered into a database. Most companies can accept data from any system and in any format, although the lack of standardization may limit interoperability of systems or users of one system. The tools tested for TO7 can accept data from any source. Source systems are generally capable of unloading data in one (or more) formats (such as comma delimited [spreadsheet], XML, etc.) that the tool can then accept. Although companies can generally accept data in any format, companies may require a few data elements to be entered in a standardized format or may convert data to a standardized format themselves to ease analysis. For example, date may be standardized to ensure all dates are entered as MM/DD/YYYY by requiring companies to enter dates as such or by altering submitted date information to meet this standard. Likewise, certain data elements captured for this task had to be standardized during entry into the database. Many companies expressed the ease standardization of data provides and expressed a need for standardization across systems to increase interoperability. Interoperability is important to be able to show total supply chain product tracing and has yet to be achieved for the various systems companies use to communicate with one another.

Although some technology companies dictate which data elements are required, generally they can accept any data elements and often leave it up to the user to determine which elements they would like captured. One exception was a company that assigns a unique lot ID to products from a food company and this is the only data element that must be captured. The systems used for TO7 could accept any data elements, although required data elements were predetermined for this task.

Most data are captured via an electronic data interchange of some sort, such as upload, email, e-fax or scanning. Paper records can be manually entered or scanned and read by the system. Web data entry was the most common method for data entry, as it only requires users to have a reliable internet connection and computer. Only two companies reported physical mediums to record and share data other than a computer. The hardware, software and communication capabilities needed to support the TO7 exercise were designed using commodity hardware which are widely available and software which may already be licensed by FDA or other companies. Who manually enters data often depends on who houses the system – the food company or the technology company. If the technology company is to manually enter data, they receive paper records by mail, fax, Fed-Ex, export and so on. Many technology companies allow for input of non-electronic records into their system to allow smaller trading partners who may not have electronic recordkeeping to still participate, although this method does require more work and

[3] www.epcglobalinc.org/standards/epcis/epcis_1_0_1-standard-20070921.pdf

has more room for error. The potential for human error (transposition of numbers and other human errors) is high with manual entry and transposition errors may still occur if a computer reads a scanned version of the paper document. Also, if records must be sent to someone else for manual entry there is a slight lag in having "real-time" data available through the system.

As data are entered onto a system, they typically become instantly available to users, so most respondents felt that their system had data available in real time. The majority of companies have quality assurance checks or filters of data built into their systems to decrease human or other errors in data. If some of the automatic checks for accuracy aren't met, an alert might go to the data submitter or the system may flag data elements for follow-up by an operator. These automatic checks/filters may include a standard format for entry of data fields, such as date, or not allow blank or duplicate data.

All companies have some way to secure data in their system, allowing the user to determine who can view what data. The systems often provided user log-ins, passwords, and so on for secure access. For most, systems can be used by all with user names and passwords from within a company, even those in other facilities. Users can also determine if supply chain partners or government will be provided with a password to view specific data companies decide to provide them with access to. In a few instances, the system did not enable supply chain partners to view data. In other instances, the system provided regulatory agencies with their own special access capabilities, such as in the event of an outbreak. All technologies, including those in TO7, also reported the ability to "hide" or blind proprietary information, with the exception of the company that assigns a unique identifier and doesn't require any other company data, including proprietary data. Proprietary data can be hidden or blinded in other systems often because the user has the ability to determine who can view specific data.

End reporting capabilities differed for companies, but many store data in XML format so that they can be exported to any system. Almost all had the capability to provide various reports from their system detailing the data submitted. Systems varied in their ability to do more with the data, such as determine points of convergence. Some companies report being able to identify points of convergence and others report that with additional, more advanced software (that is currently available) their system would be capable of this. All technology companies reported that their system can show product movement throughout the entire supply chain, although of interest is that two companies have included technologies to show this movement on a map. As reported, these systems seem most similar to the visualization software piloted in this task.

All companies indicated that they would be very willing to participate in a pilot to conduct a mock trace-back/trace forward of a specific industry segment, and many are already conducting pilots on their own. One company mentioned a number of ongoing pilots with large US retailers; another is conducting numerous pilots, including a state-wide farm-to-point of service trace-back pilot; and yet another has scheduled a real time pilot for farm-to-point of service trace-back/trace forward of product from a foreign country going to a US retail store. It should be noted that the pilots mentioned by traceability solutions providers were in various stages: completed, ongoing, or envisioned. In general, traceability solutions providers seem very interested in sharing their company's offerings with FDA and others. Also, IFT spoke to 58 food companies for TO6, including produce growers, packers, shippers, and many offered their assistance for future product tracing pilots as well.

There are also numerous off-the-shelf technologies that companies may tailor to meet their product tracing needs, if they do not have the resources to create and implement their own product tracing system in-house. These systems were not explored for TO7, but research during TO6 found that many food companies have adopted a warehouse management system or accounting system that captures currently required product tracing data, to varying extents. Import of data from accounting systems was originally planned for this task, but instead data were moved to spreadsheets for entry into this task's database. However, some smaller companies may not have resources for these systems. It should also be considered that many industry systems lack interoperability along various points in the supply chain, and do not lend themselves to product tracing for the entire sector.

Conclusions and Next Steps

The mock trace-back/trace forward exercise was able to show there is potential to expedite trace-backs by visualizing supply chains to find points of commonality based on data availability, capture, and readiness. The exercise tested limited data sets so that aspects such as real time data, other food products, a broader geographical region, import data, and a complete supply chain weren't tested to their full extent. Future work on other sectors of the food industry may help determine if data are missing or more incomplete than what the industry leaders assisting with this exercise have access to. Other existing product tracing technologies should also be tested. Applicability of this and other tools to other sectors of the food industry and other increased geographical areas should also be explored, as well as applicability to various size companies. For future pilots, necessary data must be clarified and gathered early on. Further discussion in many areas is needed to consider moving this tool beyond this exercise, including required data elements; how data would be captured, quality assured, and stored on a larger scale; governance and agreements for security and privacy of data; and so on.

There are a number of possible next steps that could be taken to build upon the lessons learned from this mock trace-back/trace forward exercise. After continued exploration and work have occurred, the capabilities researched for the mock trace-back/trace forward exercise could be utilized to implement a national trace-back/trace forward exercise for tomatoes. Mock trace-back/trace forward pilots may also be conducted for other segments of the food industry, including additional produce lines, as well as pilots to study the capabilities of additional product tracing systems. None of the 10 additional product tracing systems compared to the systems used for this task were tested by IFT, and it would be worthwhile to test the capabilities of other systems in additional pilots, and to test the applicability of the systems to other segments of the food industry. Many steps taken for this mock trace-back/trace forward will be necessary for future pilots, such as identification and evaluation of key data issues pertaining to other food items will be necessary. While some involved with the mock trace-back felt pilots for other segments of the food industry would transfer limited resources from establishing a national trace-back/trace forward system for tomatoes, others believed it is more appropriate to look at extensibility.

Identification and standardization of key data elements that industry should record and submit to FDA upon request to show linkages along the entire supply chain for improved product tracing is necessary for improved trace-backs. The significant time and effort needed on the front-end of this task for quality assurance and standardization of data could be eliminated potentially making trace-back/trace forward exercises even quicker if key data elements were identified and required to be maintained in a standardized form for product tracing. Standardization of how industry must record and provide product tracing information to FDA could also lend itself to interoperability of various product tracing systems. Also, continued collaboration among industry, government and other food safety stakeholders is essential to build upon the collaboration and its positive impacts shown in this exercise and lead to improved trace-back/trace forwards and overall food safety. However, to assure collaboration and to identify and standardize key data elements more information is needed. Further solutions must be tested and assured in order to scale up the discoveries from the mock trace-back/trace forward for national implementation, such as protocols, standards and agreements for industry/government collaboration before, during, and after food safety issues. Also knowing what industry data are available outside the tomato industry would help to determine key data elements, which can lead to improved data acquisition, quality assurance, storage and use, which in turn will lead to an improved trace-back/trace forward effort.

Appendix A: Harvard University Executive Session on Food Safety Participants

David Acheson	US Food and Drug Administration (through July 31, 2009)
Marion Aller	Florida Dept. of Agriculture and Consumer Services
Edward Beckman	California Tomato Farmers
Reginald Brown	Florida Tomato Exchange
Jeff Farrar	California Department of Health Services
Gary Fleming	Produce Marketing Association
Bruce Harris	Microsoft
Ana Hooper	Darden Restaurants, Inc.
Martin Ley	Del Campo
Ellen Morrison	US Food and Drug Administration
Janice Oliver	US Food and Drug Administration (through July 31, 2009)
Jeffrey Shuren	US Food and Drug Administration
Steven Solomon	US Food and Drug Administration
Thomas Stenzel	United Fresh Produce Association
Tim Wormus	TIBCO, Spotfire Division
Frank Yiannas	Wal-Mart Stores
Zachary Tumin	Harvard Kennedy School

Appendix B: Key TO7 Personnel

Staff Personnel
Dr. Jennifer McEntire served as the Program Development Analyst for this task. In this role, she had broad oversight and accountability for all aspects of the task. Additionally, she worked to execute subcontracts and ensure that the subcontractors provided deliverables on a timely basis. Jennifer is a food microbiologist and manages IFT's grants and contracts program.

Sarah Ohlhorst, MS, RD, served as the Project Director for this task. She was responsible for day-to-day activities, including maintaining the proposed timeframe. She communicated with the various subcontractors, and collected and compiled information to provide in the monthly reports and final report to the Agency. She also contributed to the development of the mock trace-back scenario, attended the demonstration exercise, contributed to the analysis of those findings, and led the development of recommendations about the utilization of similar technology platforms to store and manage industry product tracing data on a national scale for fresh produce and its applicability to other food items. Sarah is a registered dietitian with an MS in food science, and has been a key contributor to many IFT task orders.

Non-staff Personnel
As indicated above, several subcontractors were used to accomplish this task. The individuals designated to work on this task were:

Microsoft:
Susan Conway is a Architect/Project Manager with over 15 years of experience bringing IT solutions to government (DoD and Civilian) and commercial customers. She is responsible for solution strategy for Business Intelligence, Knowledge Management, Productivity Optimization, MOSS and database/data warehouse projects. She is experienced at aligning solutions to agency strategy, architecture implementations and designing Proofs of Concept (POC's). She works extensively with customers and technology partners to develop relationships and strategies for successful solution development.

Kevin daCosta is a Consultant II, specializing in UI design and implementation of .NET solutions, SPS webparts and enterprise templates.

TIBCO:
Tim Wormus: In his role as an Analytics Evangelist for Spotfire, a division of TIBCO Software Inc., Tim Wormus is responsible for tracking and analyzing Analytics and Business Intelligence trends, as well as helping Spotfire customers develop novel uses for Spotfire analytic methods.

Joseph Taylor is the Spotfire Division Director of Professional Services for North America.

Jack Callahan is a Spotfire Senior Consultant.

Dave Athey is a Principle Services Consultant for Spotfire.

Harvard Kennedy School:

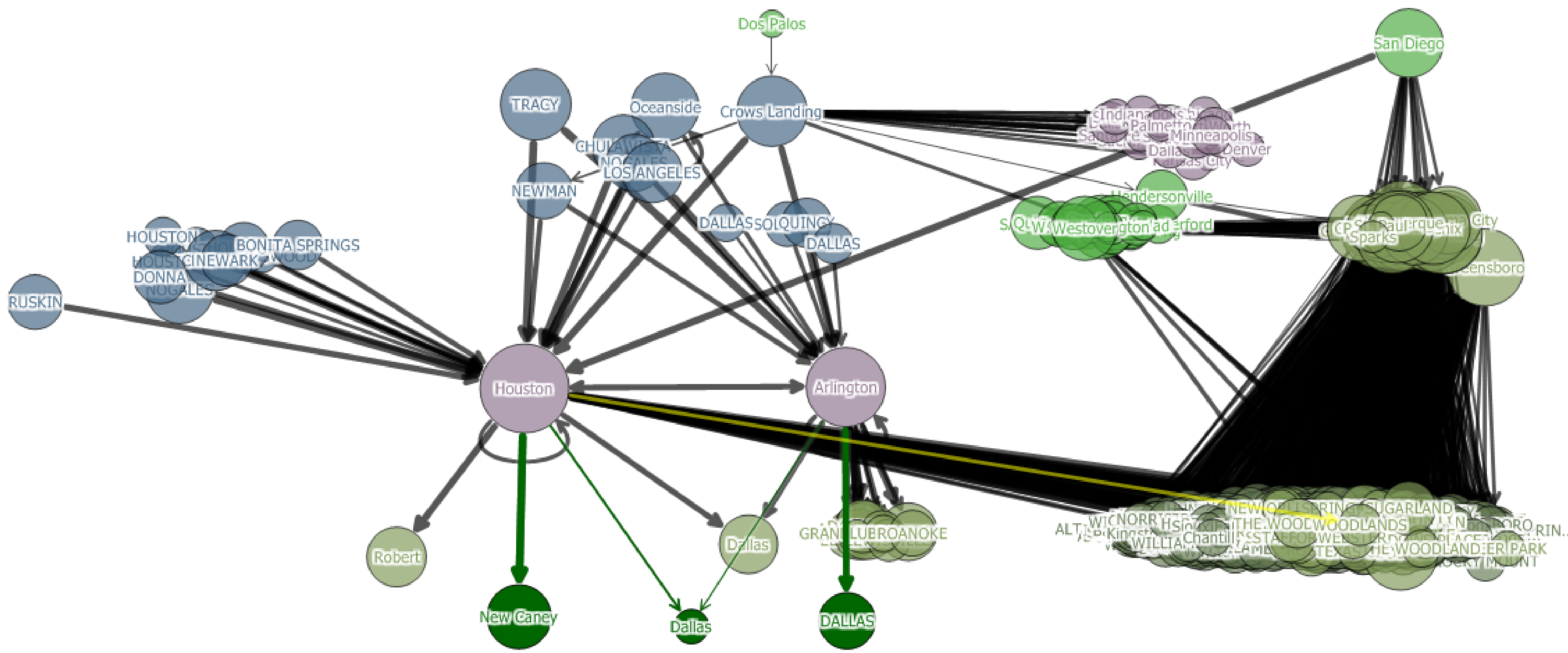
Zachary Tumin served as the Project Manager.

Mark H. Moore served at the Principle Investigator.

Experts:
During the conceptualization of this work, the Harvard Kennedy School worked to engage several stakeholders in the tomato supply chain, including trade association representatives, state personnel, industry, etc. This group performed a separate function from typical IFT panels and as more data were collected throughout the task, contributors were allowed to sit in on Task Order 7 meetings and sometimes calls. This group was more fluid than panels used in other tasks.

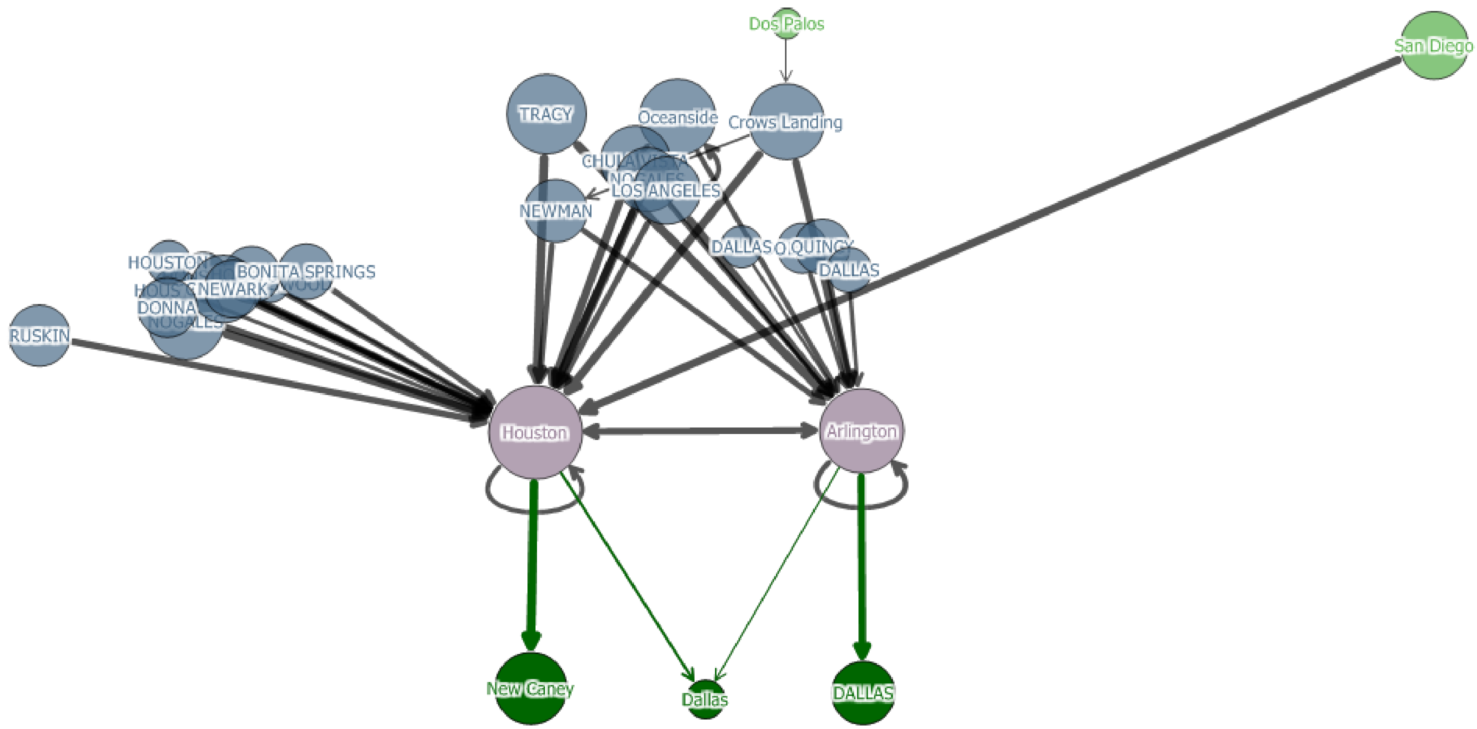
Appendix C: TIBCO Screen Shots

Network Graph



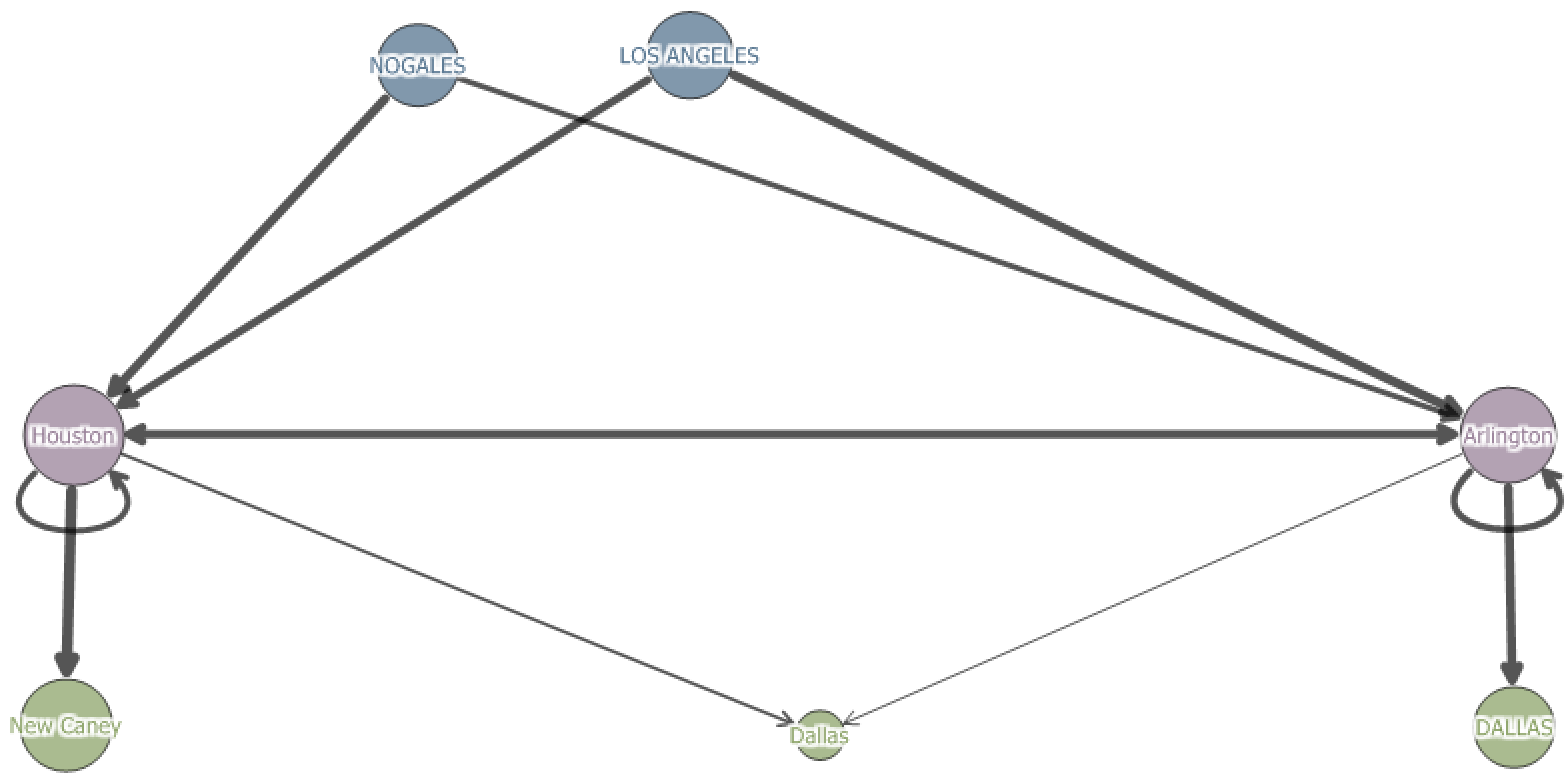
This screen shot shows an overview of network data gathered in November 2008. This includes growers (green), packers (blue), repackers (pink/purple), distributors (grey), and points of service (dark grey). For the purposes of this trace-back, distributors were considered to be points of service. These network diagrams are pulled from a database of thousands of individual transaction records.

Network Graph



This screen shot traces a cluster of illnesses back to tomatoes from three growers (dark green). Two are located in Dallas, TX and one is located in New Caney, LA. The network of packers, repackers, and distributors for these three growers are visually identified as well.

Network Graph



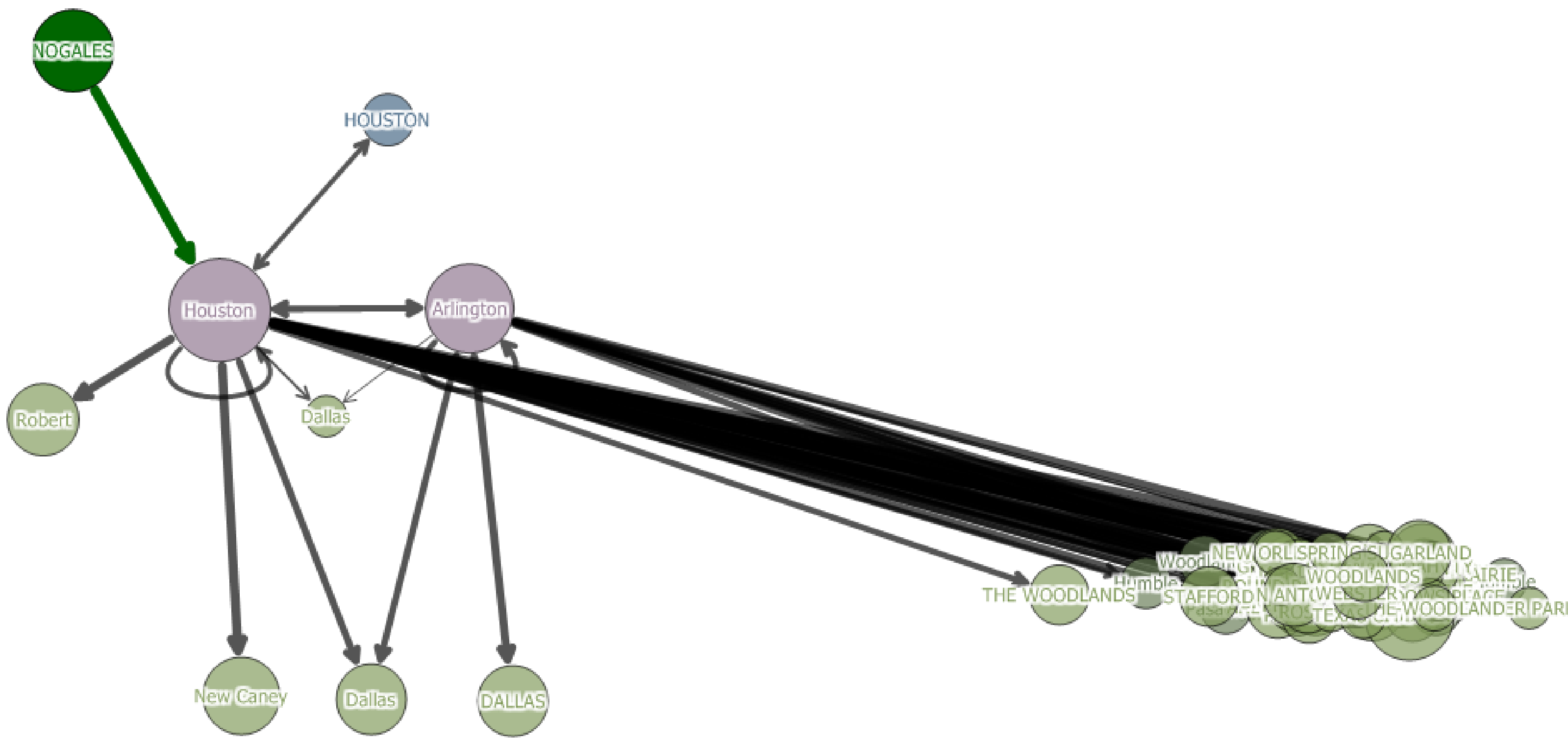
This screen shot removes all suppliers who, assuming the data are correct, were not involved with the three implicated sites. This allowed the trace-back exercise team to identify two suppliers, one in Nogales, AZ and one in Los Angeles, CA who supplied tomatoes which were sent to all three locations implicated in the cluster of illnesses.

Table

Shipped From Bl...	Shipped From City	Shipped From St...	Shipped From N...	InputLotNumber	OutputLotNumber	RepackRunNum...	Weight	Shi...
169587	Houston	TX	Repacker	09823101	09823101	1024664	1350.00	50
169587	Houston	TX	Repacker	243480	09840101	1024666	123.72	NULL
169587	Houston	TX	Repacker	243480	09840101	1024666	346.43	NULL
169592	Arlington	TX	Repacker	243480	09840101	1024666	123.72	NULL
169592	Arlington	TX	Repacker	243480	09840101	1024666	346.43	NULL
169587	Houston	TX	Repacker	09809301	X1024666	1024666	128.74	5
169587	Houston	TX	Repacker	09840101	X1024666	1024666	123.72	5
169587	Houston	TX	Repacker	09809301	X1024666	1024666	360.49	14
169587	Houston	TX	Repacker	09840101	X1024666	1024666	346.43	14

By identifying repack run numbers (e.g. 1024666) which were common to the illness clusters, it was possible to determine which lots of tomatoes were implicated. Repack run numbers could then be used to trace-back. Two suppliers were identified: one in Nogales, AZ and one in Los Angeles, CA. Industry insight supplied the information that the Los Angeles supplier was a repacker of Nogales tomatoes. Therefore, the Nogales supplier appeared to be implicated in the cluster of illnesses. The network visualization showed there was a link between the two suppliers identified in the trace-back. A further probe of the underlying data allowed the trace-back exercise team to reasonably determine the Los Angeles supplier was likely not the source of the illness cluster, and further investigate the Nogales supplier.

Network Graph



Trace forward from implicated Nogales supplier shows where contaminated product may have been shipped to via visualization software.

Appendix D: Tool Used to Gather Information from Traceability Solutions Providers

IFT Task Order 7: Tracing Systems: an Exercise Exploring Data Needs and Design

As part of IFT's 5-year contract with FDA, IFT was tasked with facilitating a mock trace-back/trace forward exercise. IFT, through subcontracts with the Harvard Kennedy School, Microsoft, and Tibco, worked to develop a platform to submit, store and analyze existing industry data, and create a spatial visualization of the supply chain to facilitate a foodborne outbreak investigation.

As part of the task, IFT is charged with providing recommendations regarding utilizing similar technology platforms for the purpose of analyzing data to facilitate a foodborne outbreak investigation. We are requesting your assistance with this portion of the task, so that we may compare existing product tracing platforms/systems with the one conceptualized through this task. Please provide information on the topics below to Sarah Ohlhorst (sohlhorst@ift.org) *by Friday, August 21*, or let her know if you have any questions.

Assumptions:

- FDA has specified what data are needed for analysis during an outbreak
- All members of a supply chain are willing to contribute data to your system

1. Where does data reside and who has ownership once it's been submitted to your system – the company or a third-party.
2. Can your system accept non-electronic records.
3. What data elements (ex. lot number, ship date) does your system use to demonstrate current product tracing capabilities.
4. Must data be submitted in a standardized format (e.g., everyone uses Julian date) and must data be submitted through a standardized system (e.g., electronic spreadsheet).
5. How is quality assurance of data managed (e.g., ensuring the accuracy of data, and identifying potential inaccuracies with a "red flag").
6. Are data shown in real-time.
7. What are the analytical end-reporting capabilities of your system. (e.g., Can data be mapped to a standardized system, such as an electronic spreadsheet, for analysis or will FDA be able to analyze data within this system.)
8. How is access to data enabled or restricted for:
    a. multiple facilities within the providing company.
    b. supply chain partners not using this system.
    c. a regulatory agency.
9. Can your platform show product movement throughout the supply chain (assuming that all data are provided).

    a. If not, which portion of the supply chain does your platform service.
10. Is the system able to determine points of convergence within the supply chain (e.g., identifying a common source in the supply chain where contamination might have occurred).

Optional Topics

1. What are the hardware and software requirements to use your system.
2. Can the platform accept and manage data from sources outside the U.S.
3. Does the platform work for all sectors of the food industry (produce, packaged goods, distribution, retail, foodservice).
4. Do you control security of data.
    a. If yes, how so.
5. Are there capabilities to “hide” or blind proprietary portions of data.
6. Would you be willing to participate in a pilot to conduct a mock trace-back/trace forward of a specific segment of the food industry.