Project Title: Food Product Tracing Technology Capabilities and Interoperability
PI: Dr. Jennifer Cleveland McEntire
Project Start Date: September 2010 **End Date:** August 2012
Period covered in this report: 5/1/2011 – 8/31/2011

1. Aims of the project

The aims of the project are to

- Compare and contrast the capabilities of different traceability solutions currently in the market.
- Identify the potential opportunities and challenges to enable interoperability between the different traceability solutions.
- Refine and test IFT-coined product tracing terms, "Critical Tracking Events" (CTEs) and "Key Data Elements" (KDEs).

2. Aims (tasks, milestones, deliverables) addressed in this progress report (May – Aug, 2011)

The following tasks, milestones and deliverables are addressed in this progress report:

- *Milestone*: Expert Panel Meeting 2
- *Deliverable*: Compare and contrast technology providers' capabilities

3. Research conducted and results obtained – include description of research protocol and summary of data collected in tabular form.

A model of the real world supply chain dataset collected was created and a virtual contaminant was introduced to create a fictitious outbreak scenario. The simulation kept track of which lots of the ingredients and finished products were contaminated and which ones were not. This allowed us to compare and contrast the responses provided by the individual traceability solution providers to evaluate the rate of accuracy as well as identify any false positives/false negatives.

The solution providers were given the test data set in early March and had about 1 week to enter the data into their technology solution. Once data entry was completed, each solution provider was given the first scenario, and we recorded their time to respond. Once the first scenario was completed, the second scenario was forwarded to them and we again recorded their time to respond. We evaluated the traceability solutions on three success metrics (successful data entry, response to Scenario 1 provided, and response to Scenario 2 provided) and three time metrics (time to complete data entry, time to respond to Scenario 1, and time to respond to Scenario 2).

The scenarios used for this test are as follows:
<u>Scenario 1:</u>
A retail location has asked that you trace back to determine all ingredients that were used to create batch/lot number 18812849, and then trace forward to determine all the product lots that used these ingredients.

- Report back to IFT as soon as your company has traced batch/lot number 18812849 *both* forwards *and* backwards in the supply chain. (This time will be used as one measure to determine your company's capability.)
- Report back to IFT on *all* ingredients associated with batch/lot number 18812849. (List of unique ingredient batch lot numbers used to create this lot)

- Report back to IFT on *all* product lots in the supply chain that used any of the ingredients associated with batch/lot number 18812849. (List of unique product batch lot numbers at the retail level that used those ingredients listed in 2.)

Scenario 2:

Track all products consumed between February 11 and February 17 at the following retail location #740479272733 that have been linked to a foodborne illness outbreak of tacos.

- Report back to IFT once your company has identified the likely source(s) of contamination (list of unique batch lot numbers by product).
- Report back to IFT after attempting to find all points of commonality in the supply chain where contamination may have occurred (list of unique location identifiers and potential contamination date ranges).

Report back to IFT with any lot number(s) of product that may need to be recalled to prevent further contamination (list of unique batch lot numbers that may be affected by the contamination).

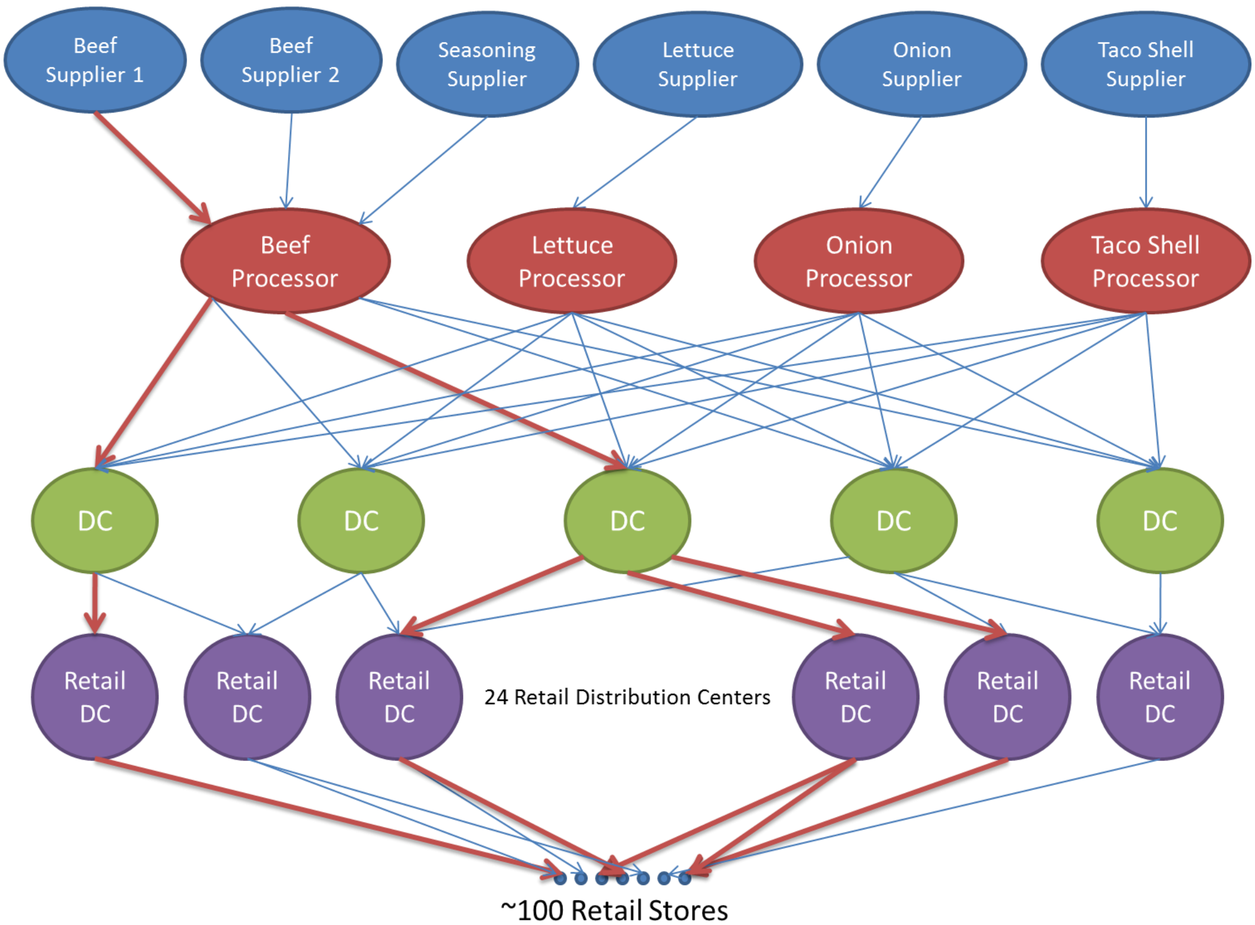


Figure 1: Introduction and monitoring of a virtual contaminant (red arrows represent flow of contaminated products) in the supply chain (blue arrows represent flow of all products)

Table 1: Capabilities Test - Success Metrics (√ = completed, X = not completed, N/A = not attempted)

Solutions Provider	Data Entry	Response to Scenario 1	Response to Scenario 2
1	√	√	√
2	√	√	√
3	√	√	√
4	√	√	√
5	X	N/A	N/A
6	X	N/A	N/A
7	√	X	N/A
8	X	N/A	N/A
9	√	√	√

Table 1 summarizes the results of the success metrics. The first metric we tested for was the successful entry of the simulated supply chain data set into each technology provider's system. One provider served as a "practice test" to ensure the data were able to be managed. Since no major changes were made to the data set or the scenarios following the feedback received from the practice test, the practice test results were treated the same way as the test results for the other eight technologies, resulting in the nine data sets shown in Table 1. As shown in Table 1, six of the nine participants successfully completed the data entry phase of this test. To verify successful data entry for these six, we requested a demo or screenshot of our data set in each system's user interface. Only four of the six were able to provide us with this verification. Of the six that completed data entry and moved on to the scenarios phase of the test, five were able to provide us with the responses we requested.

Table 2: Capabilities Test - Time Metrics (in days)

Solutions Provider	Data Entry	Response to Scenario 1	Response to Scenario 2
1	7	<1	<1
2	8	<1	2
3	8	3	<1
4	8	7	3
5	N/A	N/A	N/A
6	N/A	N/A	N/A
7	9	X	N/A
8	N/A	N/A	N/A
9	11	7	5

Table 2 summarizes the results of the time metrics used to compare the capabilities of the systems. The time metrics are useful for evaluating the scalability and usability of each technology. Scalability is derived from the time it took to enter the non-standardized data set into the system (ranges from 7 to 11 days). Usability is derived from the time it took to query the scenario and provide the results (ranges from a few hours to a week). The time was measured in days it took for the participants to respond after we supplied them all the information they would need to successfully execute each of the three phases listed in Table 2.

The following table summarizes the accuracy of the responses received from the solution providers when compared with the simulated flow of contamination in our model (for the complete dataset and the actual flow of contamination, see attachment: *Simulated Data.xlsx*):

Table 3: Accuracy of responses (number of lots identified / actual number of lots affected)

Solution Provider	Scenario 1 Trace forward			Scenario 2 Traceback
	Accuracy	False positives	False Negatives	Accuracy
1	N/A*	N/A*	N/A*	6/6
2	200/200	0	5	6/6
4	130/200	6	70	N/A*
9	200/200	0	0	6/6

*For all the data fields that have "N/A", either the solution provider was unable to complete that phase of the test, or the results were in a format where further analysis was not possible (for example a geographic map of the flow of contamination or responses were provided in PDF format where time and resource constraints prohibited extraction of data and further analysis).

IFT manually mapped the flow of products, ensuring accuracy, to determine the "correct" routes, lot numbers, dates, etc. that should have been returned by the technology provider analyses.

(1) **Solution Provider 1**: For scenario 1, this solution provider successfully traced back the contaminated lots to their origin. However, the format that was used for presenting the trace forward results prohibited further analysis. Thus, its accuracy was not evaluated. For scenario 2, the solution provider successfully identified the likely sources of contamination with 100% accuracy.

(2) **Solution Provider 2**: For scenario 1, this solution provider successfully traced back the contaminated lots to their origin. For tracing forward, they identified 5 lot numbers related to 90% lean beef trimmings that were not included in the actual contamination data. Manual searching for those lot numbers did not relate them to the source of contamination. The solution provider successfully identified the correct contaminated lot numbers for 50% lean beef trimmings and the seasoning. For scenario 2, the solution provider successfully identified the likely sources of contamination with 100% accuracy.

(3) **Solution Provider 4**: For scenario 1, this solution provider successfully traced back the contaminated lots to their origin. They also successfully traced forward the contaminated lots to the distribution centers for 90% lean beef, 50% lean beef, and seasoning. However, their results at the retail level did not fully match those manually identified. The summary of their accuracy is given below:
Seasoning: total number of lots in the manual response: 80; total number of lots in their response: 59; numbers matching: 56
90% lean beef: total number of lots in the manual response: 40; total number of lots in their response: 18; numbers matching: 18
50% lean beef: total number of lots in the manual response: 80; total number of lots in their response: 59; numbers matching: 56

This solution provider was unable to provide a response for scenario 2

(4) **Solution Provider 9**: For scenario 1, this solution provider successfully traced back the contaminated lots to their origin. They also successfully traced forward the contaminated lots to the retail centers for 90% lean beef, 50% lean beef, and seasoning with 100% matching with actual contamination flow. For scenario 2, this solution provider "only" linked one ingredient (taco shell) to the location ID identified in this scenario (740479272733). For the selected ingredient, they successfully traced back to the likely source of contamination. However, it is not clear why other ingredients were not listed (i.e., diced onion, shredded lettuce, seasoned beef)

Expert Panel Meeting #2:
An in-person meeting with the expert panel was conducted on July 14, 2011 in conjunction with IFT's first Traceability Research Summit, to discuss the results from the baseline capabilities test and the face-to-face interviews conducted in the previous quarter, as well as the year 2 project plan based on the lessons learned from year 1. A summary of the test results along with an analysis was shared with the expert panel (see attachment: *NCFPD Product Tracing Phase 1 SME Update.docx*). The following individuals participated in the meeting:

- Bruce Welt, University of Florida
- Kurt Collins, UFPC
- Ben Miller, Minnesota State Department of Agriculture
- Brenda Lloyd, UFPC
- Tejas Bhatt, IFT
- Bill Hardgrave, Auburn University
- Amir Mokhtari, RTI (via conference call)

4. Analysis and interpretation of results

The following factors contributed to the failure of the 3 participants to complete the data entry phase:

- Scope of the system was on internal traceability (within four walls of a facility)
- The system enabled communication between supply chain partners during an investigation, but did not store any data itself
- The data set was not in a standardized format

In addition, one participant was unable to respond to the scenarios because some data elements required by their system were not available within the data set (like the name of the employees that handled the food product). It was apparent from this test that not all traceability solution providers had the capability to participate in an interoperability framework proposed for Year 2, and even of those capable, we concluded from the time metrics that not all traceability solutions were built with scalability and usability in mind. For a truly interoperable framework and the timely resolution of an outbreak investigation, including trace-back and recall, the technologies need to be fast and responsive. The time metric helped us further narrow down the prospective list of candidates who had the baseline capabilities to participate in Year 2 of this project. At this point, solution providers 1, 2 and 9 look most promising as participants for the interoperability study in Year 2, as they were able to handle our data set and respond quickly to the scenarios. When discussing the successes and failures of the various participating technology solutions providers in the face-to-face meetings, several common overarching challenges emerged that have an impact on interoperability and whole-chain traceability. Concerns about confidentiality of data, protection of formulation information, and potential loss of competitive advantage create a corporate resistance to more actively engaging in whole-chain traceability. This concern over privacy (as well as cost) was mentioned as the reason some small companies within the food system still use a paper-based system for tracking and tracing. Several traceability solutions can accept faxes or scanned paper-based records and use optical character recognition (OCR) to recover key data elements (KDE) and critical tracking events (CTE) from them. Cost was another major factor cited as preventing small to mid-sized firms from adopting electronic whole-chain traceability. To minimize the cost overhead, several solution providers have designed their systems to work directly with existing internal systems at these firms.

It was clear from these discussions that for any interoperability framework to be successful, it needs to ensure minimal changes in information technology (IT) at the small to mid-sized firms and be capable of importing data via a variety of methods, including:

- Excel
- Extensible Markup Language (XML)
- Web services with batch uploads
- Manual form-based uploads
- Integration with enterprise resource planning (ERP) systems like SAP
- Faxed form using OCR
- Extract-Transform-Load (ETL) approaches

Some solutions use a centralized database approach in which every KDE for each CTE is sent to and collected by the technology. This kind of a centralized approach has some serious scalability and performance issues when attempting to interoperate with a system as complex as the food system where there is a need to store and analyze large data sets.

To mitigate the negative effects of a centralized system, some companies have used cloud/distributed computing *for their technology*. However, distributed systems have the challenge of uptime and maintenance (reliability), as well as higher costs (of implementation and ownership of IT infrastructure). Still, this fragmented approach works well if all members of the supply chain use the same technology solution provider, but fails to enable true interoperability due to lack of standardization. The lack of standardization not only applies to the type and scope of data collected, but also standards related to data sharing and communication.

Discovery services are being developed and introduced to overcome this challenge. Such services act as links or interfaces between two different traceability solutions and provide a standardized way for them to communicate and share data (for example, electronic product code information services or EPCIS). However, for this approach to be effective, a widespread adoption of those standards needs to be in place, but currently is not. Finally, the majority of the participants noted that pedigree is an inefficient method to accomplish traceability.

Whatever the technical approach, each of the nine systems evaluated could be categorized into one of the following three scopes from a traceability perspective:

- Some solutions work only for internal traceability (inventory and warehouse management systems)
- Some solutions work only for external traceability
- Some solutions work for both internal and external traceability.

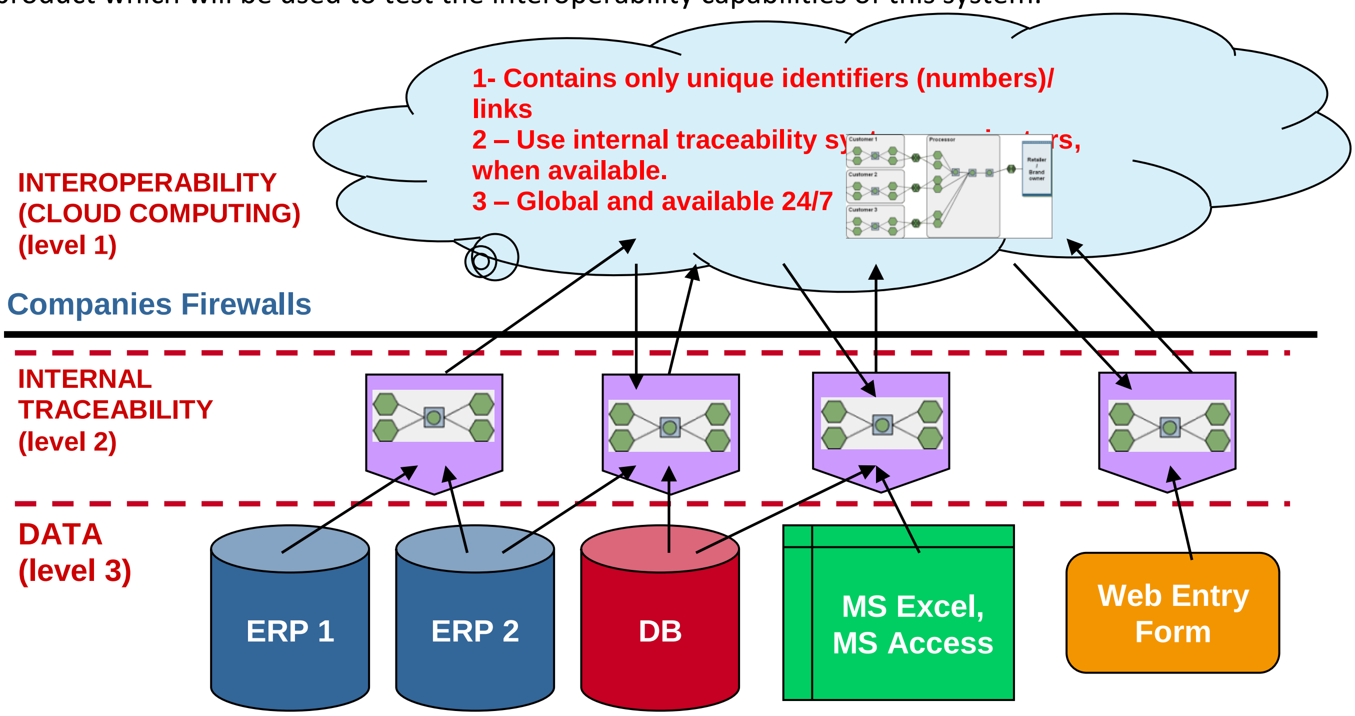
For Year 2, while recognizing these variances in traceability technology capabilities, we propose engaging only those who currently have a whole-chain traceability component in their system.

5. Pitfalls or difficulties encountered and alternative methods to resolve the problems

None

6. Future plans/next steps

Based on the lessons learned from the Year 1 test of baseline capabilities, we have a modified approach to test for interoperability of the traceability solutions (outlined in the figure below). The subject matter experts agreed upon this approach during their second meeting (July 2011) to first enable communication between third party solution providers and then test the effectiveness of their interoperability capabilities. The two traceability research summits hosted by IFT along with its funding partners have conceptualized an architecture for an interoperable traceability system which will be used for this project. We will work with three companies that successfully completed the baseline tests to modify/tweak their systems to enable interoperability. We are also collecting the real world supply chain data for a canned tuna product which will be used to test the interoperability capabilities of this system.



7. Additional references (not previously cited)

Additional Reference Materials:

- *"NCFPD Product Tracing Phase 1 SME Update.docx"*
- *"Simulated Data.xlsx"*