The NCFOODSAFE Project
Developing a Foodborne Events Data Integration and Analysis Tool for North Carolina (NCFEDA)

Noel Greis and Monica Nogueira
Center for Logistics and Digital Strategy
Kenan-Flagler Business School

Pia MacDonald and Rachel Wilfert
North Carolina Center for Public Health Preparedness
Gillings School of Global Public Health

University of North Carolina at Chapel Hill

November 4, 2010
Food Safety Challenges

Underlying premise

• Our national security depends on a safe and secure food supply—whether the contamination is *unintentional* or the result of a *terrorist act*.

• Currently, **15 federal agencies** collectively administer at least **30 laws** related to food safety.

• Meeting this challenge depends on:
  
  – **New process models** reflecting new stakeholders (consumers, private sector)
  
  – **New, non-traditional data sources** that support real-time monitoring and response; and
  
  – **New informatics tools** that enhance our ability to collect, interpret and disseminate electronic information across organizational boundaries.

---

NO SINGLE PATTERN OF FOOD CONTAMINATION

A 2005 study from Stanford University estimated that the addition of only **4 grams of botulinum** into a milk production facility could cause as many as **50,000 fatalities within 24 hours** [Wein and Liu, 2005].

The recent **PCA peanut butter event** was the **largest food contamination event in US history** based on insurance claims. Estimates of nearly $1B have been suggested including $500M due to a 25% decline in peanut sales—as well as loss of consumer confidence in the government’s ability to protect its citizens.
What One Hundred Million Calls Reveal About New York City

Source: Wired Magazine
What Five Hundred Calls Reveal about North Carolina Food Safety in 2009
New Stakeholder Model for Food Safety
Building Collaborative Information Processes
Four Main Capabilities of NCFEDA Bridge

**Data Fusion.** Combine relevant information into a single situational picture (New and nontraditional data sources; private sector data).

**Visualization.** Provide a visual representation of data as a problem-solving technique.

**Analytics.** Find patterns in data to speed up the process of identifying source, scale and scope of contamination.

**Collaboration.** Enable anytime, anywhere exchange of information in (near) real time among all stakeholders.

**NCFEDA GOALS**

Reduce latencies in the surveillance and response to foodborne disease and contamination and

Improve preparedness of North Carolina against threats to food safety and human health.
Reducing Latency Of Response
2008-2009 PCA Peanut Butter Contamination

**Milestone 1**
What Pathogen, Bacteria, or Agent is Causing the Contamination?

- Mid September 2008
  - First laboratory-confirmed cases reported October 2008 by state health agencies.
- November 2008
  - CDC confirms cluster of *Salmonella Typhimurium*.
- December 3, 2008
  - CDC holds nationwide conference call to confirm national pattern of outbreaks. No connection to any product has been determined.

**Milestone 2**
What Food is Associated with the Contamination?

- January 7, 2009
  - CDC, FDA and Minnesota Department of Health discuss peanut butter as possible source.
- January 9, 2009
  - FDA tests of peanut butter from King Nut confirm source as Peanut Corporation of America (PCA).

**Milestone 3**
What Food Product (s) is Associated with the Contamination?

- January 10, 2009
  - First recalls and alerts issued. Other PCA facilities (e.g. Texas) and products implicated. CDC and FDA investigate outbreaks in other states (e.g. Georgia and Connecticut).
- March 8, 2009
  - PCA files for bankruptcy. Product recalls still being announced by FDA.

**Milestone 4**
Locating and Removing All Contaminated Products (and Derivative Products) from Retail (and other) Shelves.

- April 2009
  - FDA still recalling peanut products from Westco Fruit and Nut Co.
- June 4, 2009
  - FDA last known recall of peanut products from Providence Commerce, CA.
Building Situational Awareness
Leveraging (Near) Real-Time Information

**NEAR REAL-TIME: Emerging Situation Alert**
- NCDA&CS: Consumer Complaints
- CONSUMER: Web 2.0, Twitter
- NCDPH: NC DETECT

**Outbreak Detection & Investigation**
- NCDPH: NC EDSS
- NCDENR: BETS Table

**Inspections & Recall**
- NCDA&CS: Firms DB - Food Manufacturing Facilities Inspections
- NCDENR: Media file - Food Serving Facilities Inspections (restaurants/schools/nursing homes)

**Routine Trend Analysis & Prevention**
- All data sources above
- NC CATCH
- NC EDIS
NCFEDA Architecture

Relevance Engine

Analysis Triggers

Visualization Tool

Fused EventData

Business Rules
Policy/Process Rules
Predictive Analytics

Stakeholder Dashboards
- NCDA&CS Dashboard
- NCDENR Dashboard
- NCDPW Dashboard

Geocoded data

Stakeholders data

Reported Illnesses
Recall Notices
Consumer Complaints
Inspection Violations
Microblog Messages
Data Fusion for Situational Awareness and Common Operational Picture

Facilities Inspection Records
Consumer Complaints
Laboratory Reports
NC DETECT
NCEDSS
Recalls
Twitter
Google
News, Reports
Industry Violations
FDA Reportable Food Registry

Public Sources
Private Sector
NC DOC

NCEDIS
NC PH
USDA
NC DENR
NCDA&CS

Common Operational Picture
Relevance Engine & Sample Analytics

Linking “Related” Events Across Agencies

Relevance Engine

Web-based Data

Structured Data

Unstructured Data

What is the likelihood that “Event X” is related to .....?

ELI RATINGS

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No Likelihood</td>
</tr>
<tr>
<td>1</td>
<td>Possible Likelihood</td>
</tr>
<tr>
<td>2</td>
<td>Low Likelihood</td>
</tr>
<tr>
<td>3</td>
<td>Some Likelihood</td>
</tr>
<tr>
<td>4</td>
<td>Moderate Likelihood</td>
</tr>
<tr>
<td>5</td>
<td>Significant Likelihood</td>
</tr>
<tr>
<td>6</td>
<td>High Likelihood</td>
</tr>
<tr>
<td>7</td>
<td>Highest Likelihood</td>
</tr>
</tbody>
</table>

Engine Toolkit

Statistics, Data Mining, Text Mining, Logic, Rulesets, Pattern Recognition, etc.
Cluster of Reported Illnesses Identified by NC Public Health Surveillance System

**PATHOGEN:** P  
**START DATE:** MM/YYYY  
**COUNTRIES:** L1, L2, ...

**Step 1:** Search **Recalls** that match rule:  
IF RECALL issued one month prior to MM/YYYY date AND cites pathogen P AND recalled product was distributed in NC or nationwide, THEN return recall text (to allow access to recalling company, product name and distribution states).

**Step 2:** Search **Complaints** that match rule:  
IF COMPLAINT CALL received one month prior to MM/YYYY date OR consumer resides in one of the counties in the cluster L1 or L2 or... AND/OR illness caused by pathogen P AND/OR recalled product found by RECALL search rule AND/OR recalling company found by RECALL search rule, THEN return complaint call text.

**Step 3:** Search **Violations** that match:  
IF CODE VIOLATION occurred one month prior to MM/YYYY date AND/OR facility located in one of the counties in the cluster L1 or L2 or... AND/OR pathogen P found in facility AND/OR company manufactures recalled product found by RECALL search rule, THEN return code violation text.

**Step 4:** Search **Microblogs** that match:  
IF MICROBLOG MESSAGE posted one month prior to MM/YYYY date AND/OR blogger resides in NC (or mentions county L1 or L2 or...) AND/OR illness caused by pathogen P AND/OR blogger cites recalled product or recalling company found by RECALL search rule, THEN return message text.

**TRIGGER DATE ISSUED:** MM/DD1/YYYY  
FDA recalls product Q manufactured by company C due to contamination with pathogen P. The product was distributed to following states: S1, S2, NC, ...

**CALL DATE:** MM/DD2/YYYY  
Consumer called to report illness confirmed as caused by pathogen P after eating product Q from company C. Consumer resides in county L2.

**RESULT VIOLATION DATE:** MM/DD3/YYYY  
Inspected facility F tested positive for pathogen P on MM/DD3/YYYY. Company F manufactures product Q and is located in NC county L2.

**DATE POSTED:** MM/DD4/YYYY  
Blogger reported illness confirmed as caused by pathogen P after eating product Q from company C at restaurant R. Blogger resides in NC and cites county L1.
A BRIEF DEMONSTRATION

- Location: Offices of NCDA&CS
- Timeline: Start 9AM on November 4, 2010
- Duration: 3 days in 3 minutes
- Event: Possible *Salmonella* contamination
- Situational Awareness [Endsley, 1995]:
  - **Perception:** North Carolina Common Operational Picture
  - **Sensemaking:** Placing Emerging Event in National Context
  - **Projection:** Trend Analysis Based on History

CAUTION: FEDA is still a work in progress!