

Rule-Based Complex Event Processing for Food Safety and Public Health

Monica L. Nogueira and Noel P. Greis

The University of North Carolina at Chapel Hill

Barcelona, Spain, 21 July 2011

CENTER for LOGISTICS and DIGITAL STRATEGY
FRANK HAWKINS KENAN INSTITUTE of PRIVATE ENTERPRISE
WWW.CLDS.UNC.EDU



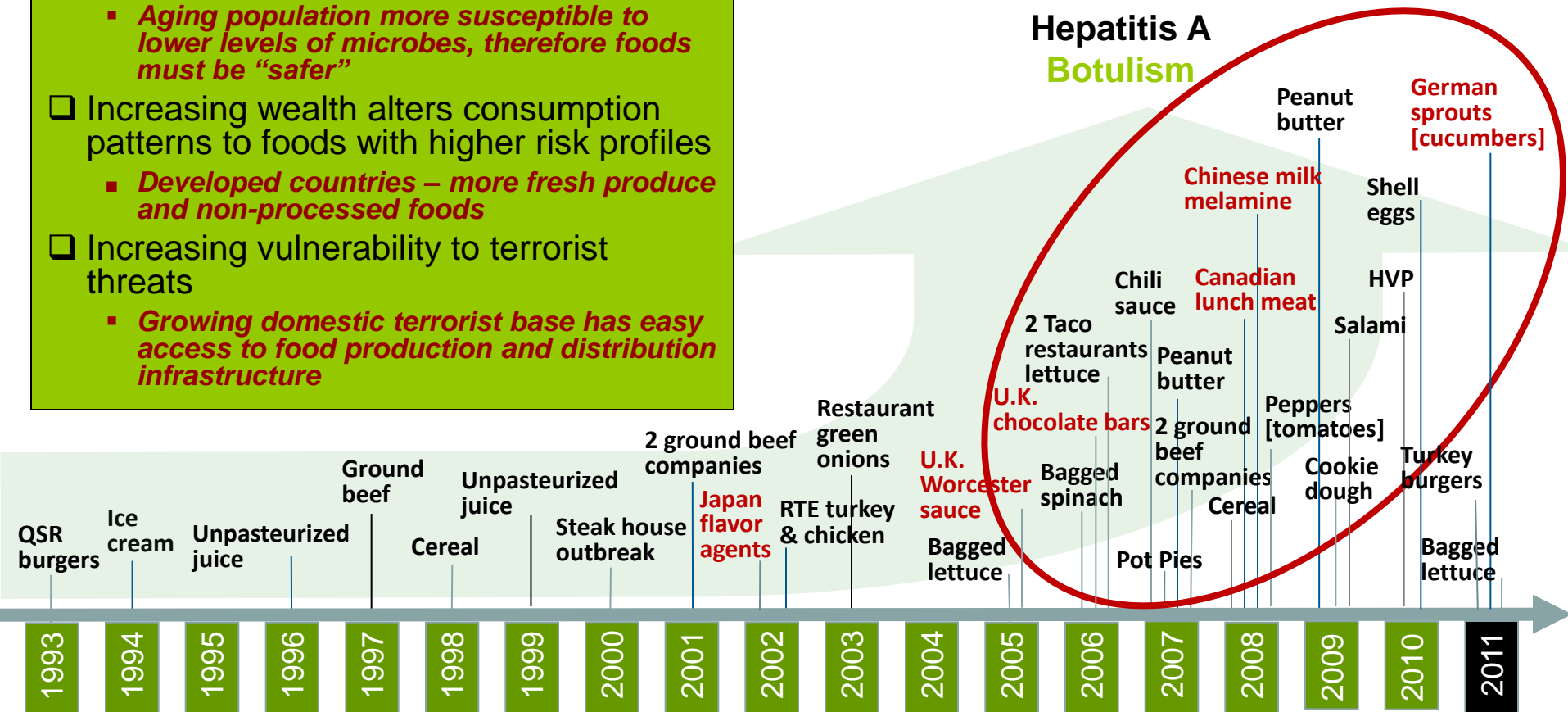
UNC
KENAN-FLAGLER
BUSINESS SCHOOL

FOOD SAFETY & PUBLIC HEALTH

Growing Global Challenges

- ❑ Demographics changes alter food safety vulnerability
 - **Aging population more susceptible to lower levels of microbes, therefore foods must be "safer"**
- ❑ Increasing wealth alters consumption patterns to foods with higher risk profiles
 - **Developed countries – more fresh produce and non-processed foods**
- ❑ Increasing vulnerability to terrorist threats
 - **Growing domestic terrorist base has easy access to food production and distribution infrastructure**

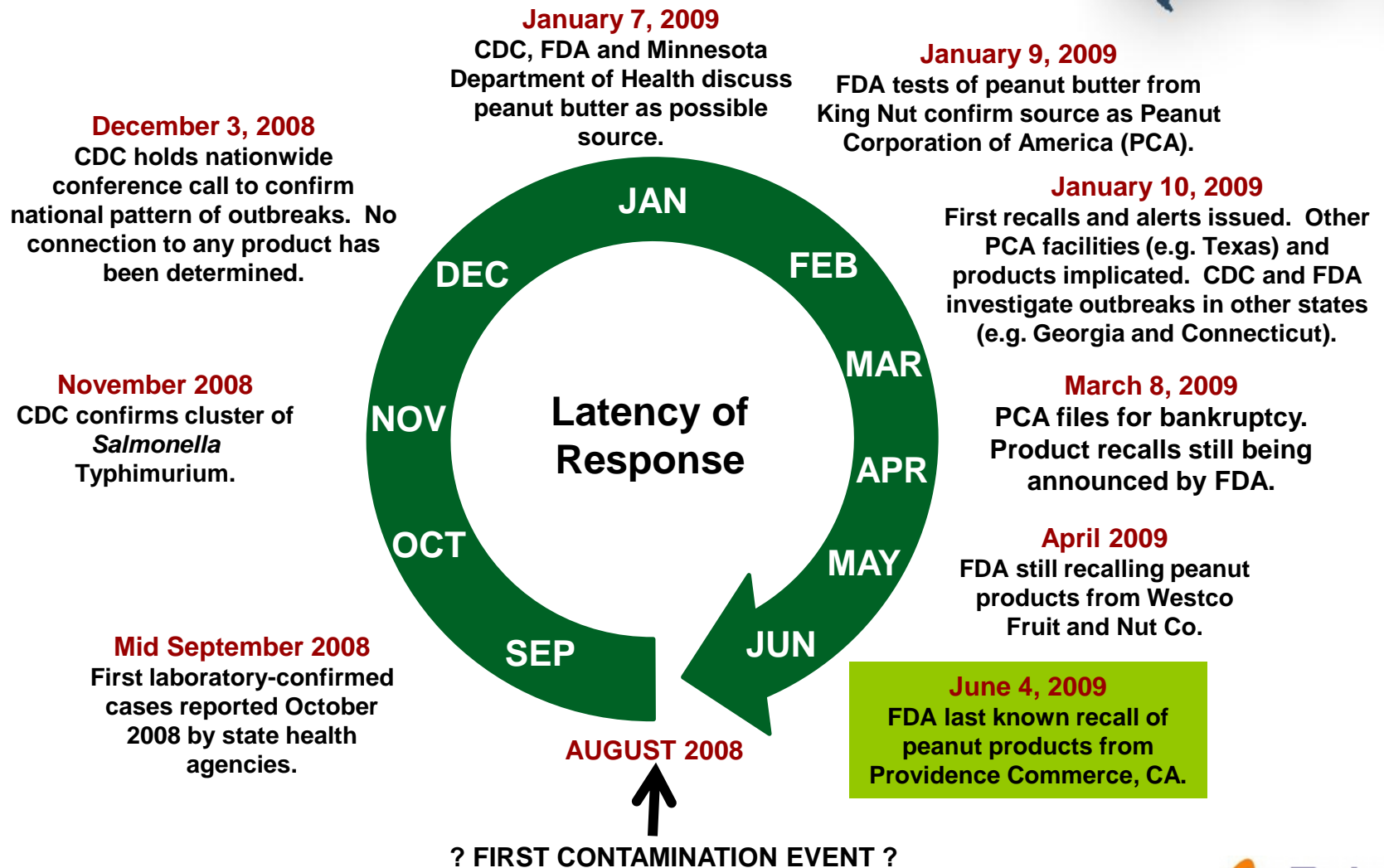
Salmonella
Listeria
E. coli O157:H7
Hepatitis A
Botulism



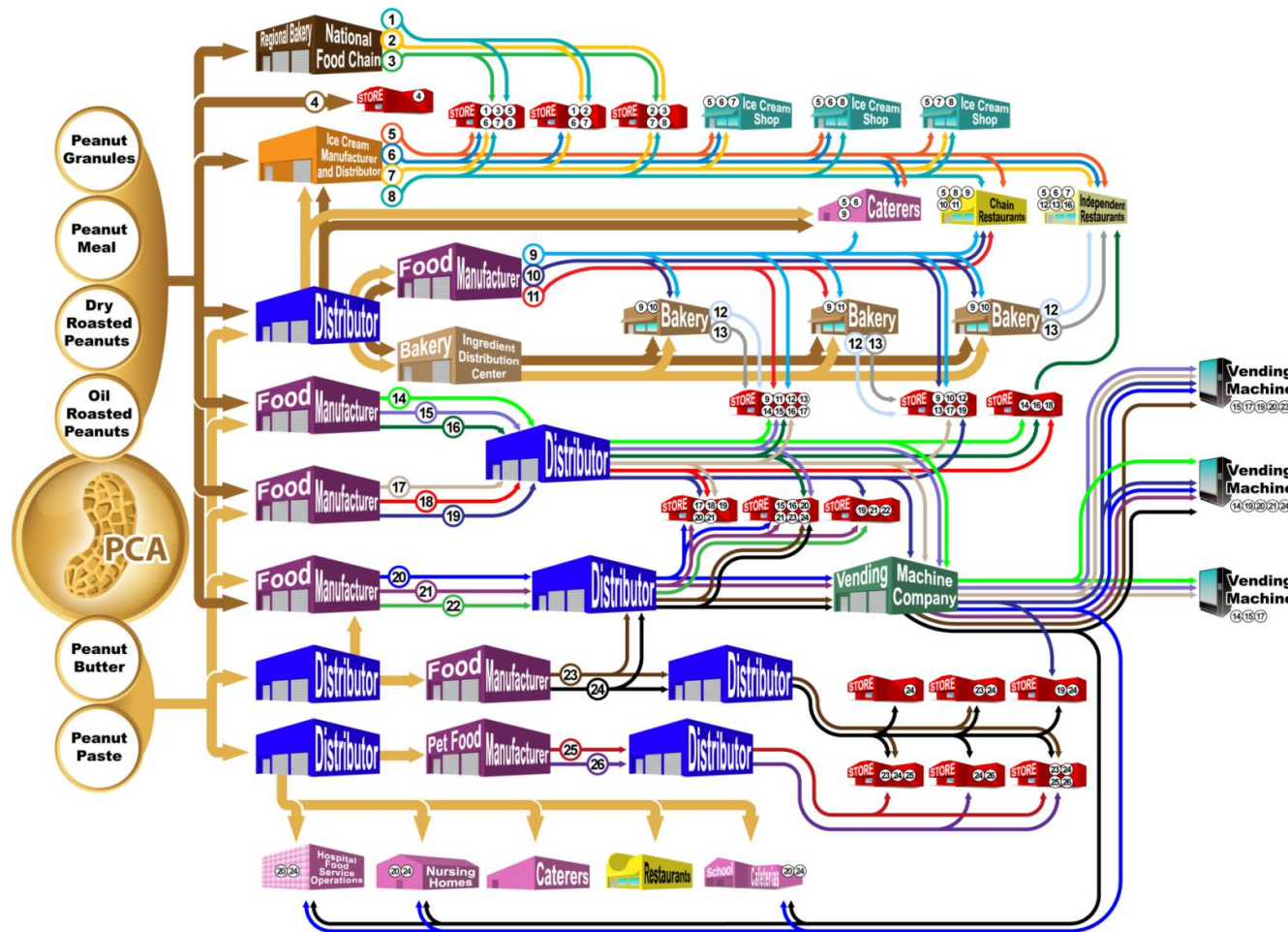
Source: Adapted from Ecolab

PEANUT BUTTER CONTAMINATION

Foodborne Disease Communication Wheel



PEANUT BUTTER DISTRIBUTION CHAIN



Food Chain Challenges:

- **Size**
- **Complexity**
- **Lack of information**
- **Lack of visibility**
- **Lack of traceability tools**
- **Lack of analytical tools**
- **Lack of food safety regulation**
- **Inadequate food safety policies & standards**

Source: U.S. Centers for Disease Control and Prevention

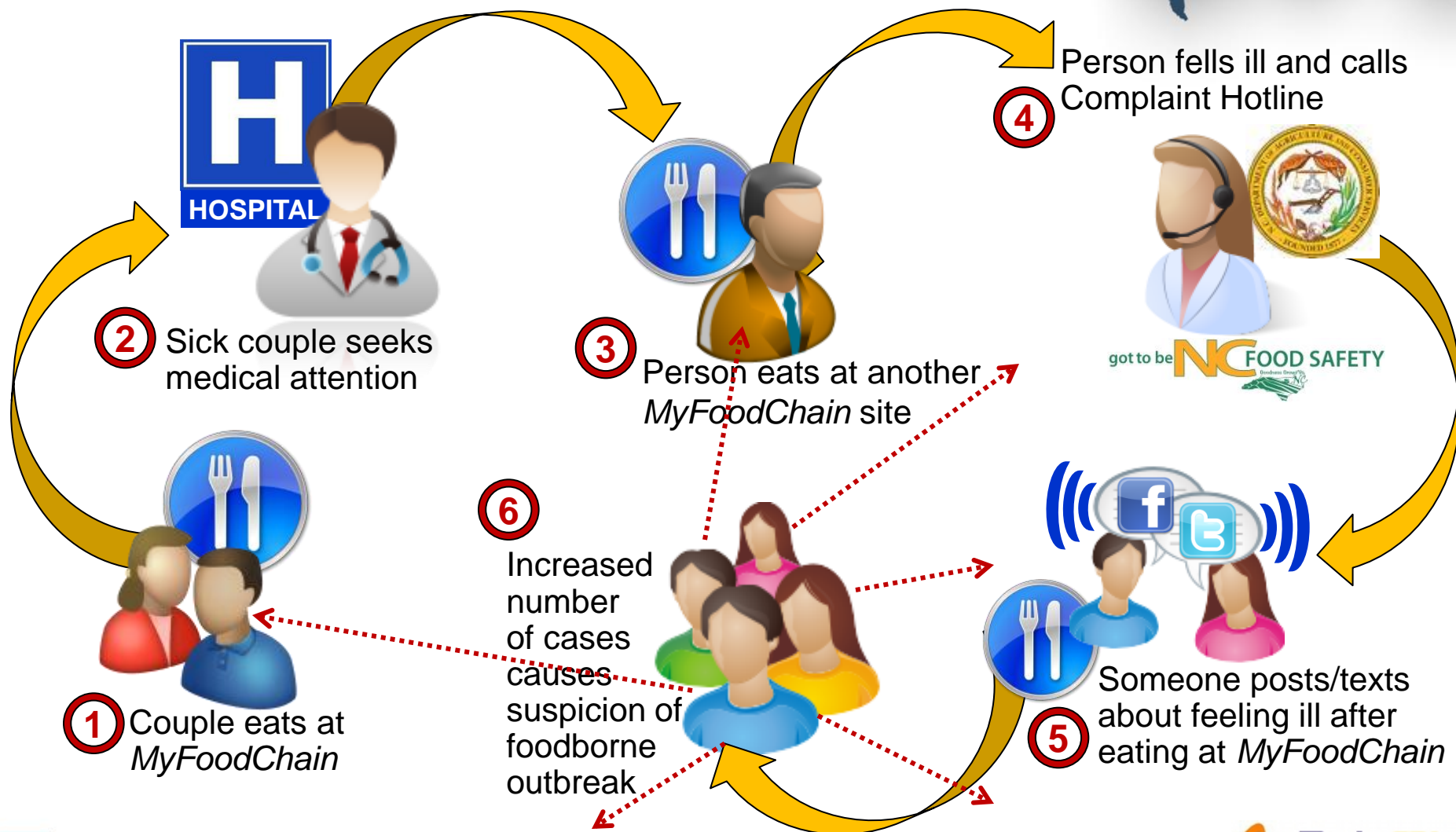
MOTIVATION

- Data silos - separate information systems maintained & managed by distinct governmental agencies and firms
- Lack of data interoperability - heterogeneous data formats
- Data incompleteness, noise, ...
 - ✓ Public sector - restrictive regulation due to privacy concerns
 - ✓ Private sector – limited access due to competitive advantage concerns
- ...

Main challenge: how to access, process and interpret more events more quickly, thus reducing the time, scale, and scope of an emerging event

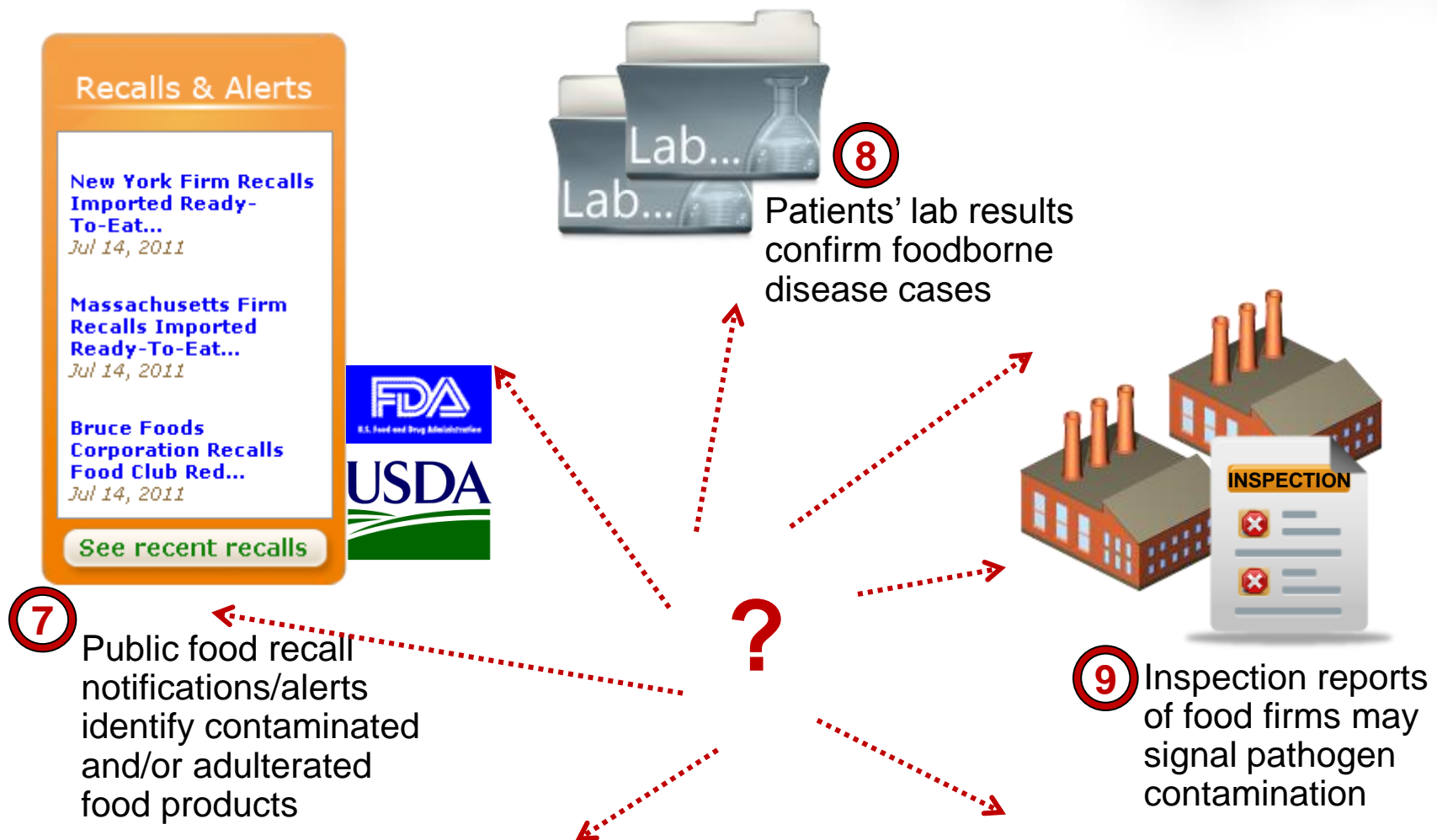
FOOD-RELATED EVENT INFORMATION

"Hints" of an Emerging Outbreak



FOOD-RELATED EVENT INFORMATION

“Hints” of an Emerging Outbreak



EVENT MODEL

- *Event* – acquisition of a piece of information that is significant within domain of interest.
 - ✓ Types: Simple or complex (materialized)
- *Simple events*
 - ✓ *Atomic* events – distinct spatio-temporal identity relevant to the determination of a complex event.
 - ✓ *Molecular* events – atomic events “linked together” by evidence.
- *Complex or materialized events*
 - ✓ Events inferred by engine’s rules evaluation of occurrence of other simple events.
- *Event stream* – sequence of simple events received by CEP system and assigned a timestamp & geostamp.

SEMANTIC MODEL

- **E1: Medical record** of patient with foodborne disease
- **E2: Lab test results** confirming a patient's foodborne illness
- **E3: Cluster of ill patients** due to common pathogen (Molecular event)
 - ✓ Cluster record – cluster id, “patient#1 id,” pathogen, county, #cases
 - ✓ Cluster illness record – cluster id, patient id, patient county
- **E4: Complaint hotline record** of consumer calls
 - ✓ Complaint caller record – call id, caller county, food product, illness status, #cases
 - ✓ Complaint food operator record – call id, operator location
 - ✓ Complaint food product record – call id, food product, FDA food code, date of manufacturing
- **E5: Public food recall** notifications
 - ✓ Recall record – recall id, company, food product, reason, #cases, #geo-areas affected
 - ✓ Recall area record – recall id, geo-area affected
- **E6: Inspection reports** of food firms
- **E7: Microblog messages** about food illness

ANSWER SET PROGRAMMING

- A rule is a statement of the form:

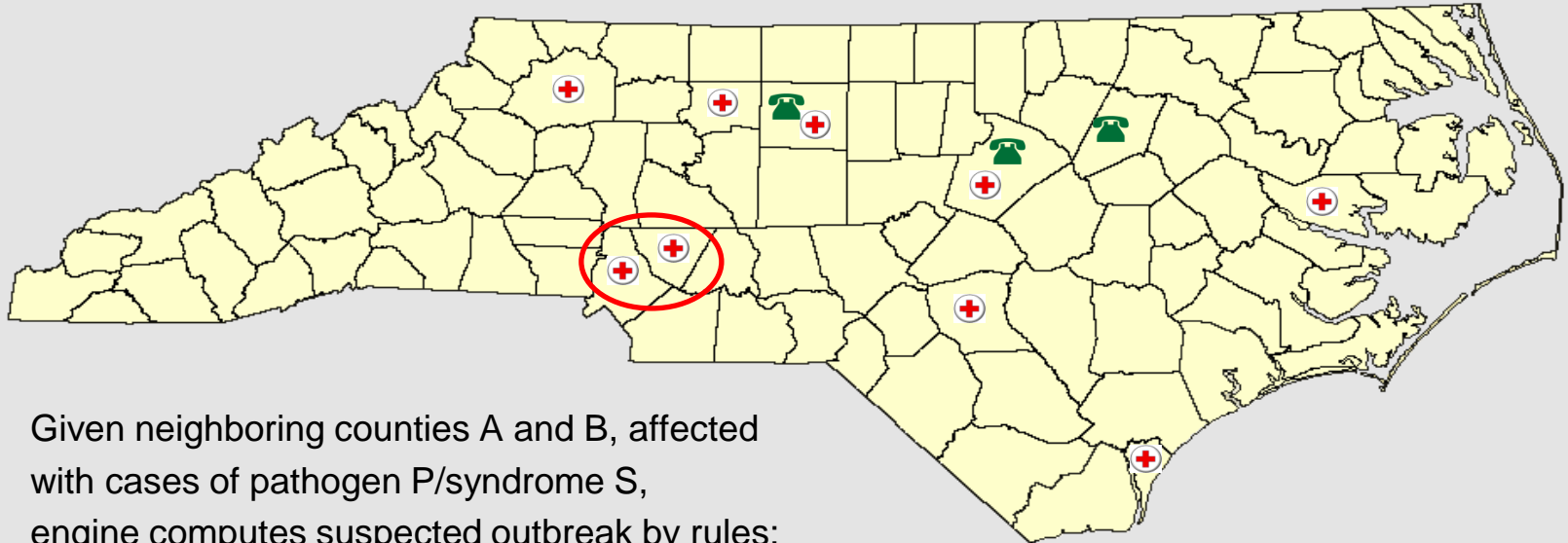
$$h_1 \vee \dots \vee h_k \leftarrow l_1, \dots, l_m, \text{ not } l_{m+1}, \dots, \text{ not } l_n.$$

where h_i 's and l_i 's are ground literals, not is a logical connective called “negation as failure” or “default negation,” and symbol \vee corresponds to the disjunction operator.

- Rule's intuitive meaning:

A reasoner believes at most one of the h_i 's if it believes all l_1, \dots, l_m and has no reason to believe any of the l_{m+1}, \dots, l_n .

DETECTING EMERGING CLUSTERS

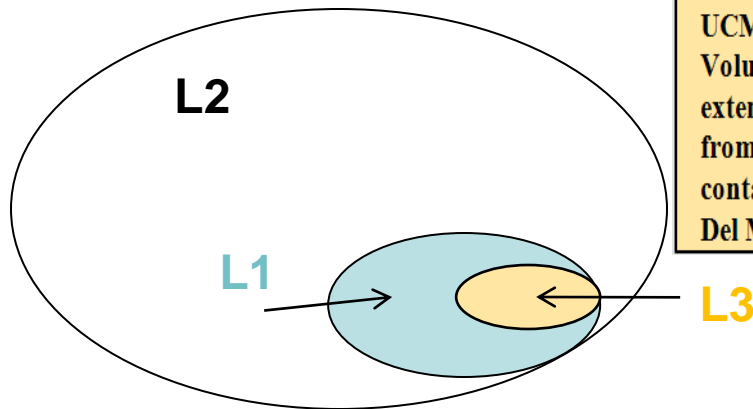


$suspccluster(A,B,P,S) \leftarrow neighbors(A,B), minreached(A,P,S), minreached(B,P,S).$

$suspccluster_illness(A,B,Id,P,A) \leftarrow suspccluster(A,B,P,S), P \neq S,$
 $patient_illness(Id,H,M,AmPm,Day,Mon,Y,A, Sys,P).$

$susp_outbreak(A,B,P) \leftarrow suspccluster(A,B,P,_).$

LINKING CLUSTERS TO FOOD RECALLS



UCM230780^11/4/2010^11/4/2010^Del Monte Fresh Produce N.A., Inc. Extended Nationwide Voluntary Cantaloupe Recall^ Del Monte Fresh Produce N.A., Inc ("Del Monte Fresh") extended today to nationwide its voluntary recall of certain cantaloupes grown in and shipped from Arizona. The affected product is being recalled because lab tests show cantaloupes are contaminated with Salmonella.^Cantaloupes^ Del Monte Fresh^Cantaloupes^Salmonella^ Del Monte Fresh Produce N.A., Inc.^nationwide^illnesses reported

- Suspected cluster is linked to recalls affecting the geographic region where cluster is located.
- Recalls of food distributed directly to a state or a more specific region are of higher interest.

more_specif_susprecall_linked(R1, A, B, F1, M1, L1) ←
susprecall(R1, A, B, F1, M1, L1), *susprecall* (R2, A, B, F2, M2, L2),
subregion(L1, L2), R1 != R2, not *other_more_specif*(A, B, L1, R2).

other_more_specif(A, B, L1, R2) ←
susprecall(R2, A, B, _, _, L2), *susprecall*(R3, A, B, _, _, L3),
subregion(L3, L1), R2 != R3.

EVIDENCE SET & EVENT EVIDENCE INDICATOR

- **Evidence Set** - set of linked events that provide evidence of the materializing of a complex event. For example,

```
evidence(A, B, P, S, R, F2, M, L, F1, FC, T) ←  
  suspcall(C, A, B, F1, FC, T), nccounty(T),  
  suspcluster(A, B, P, S),  
  susprecall(R, A, B, F2, M, L), type_of(F2, FC).
```

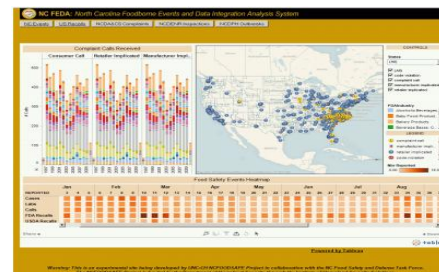
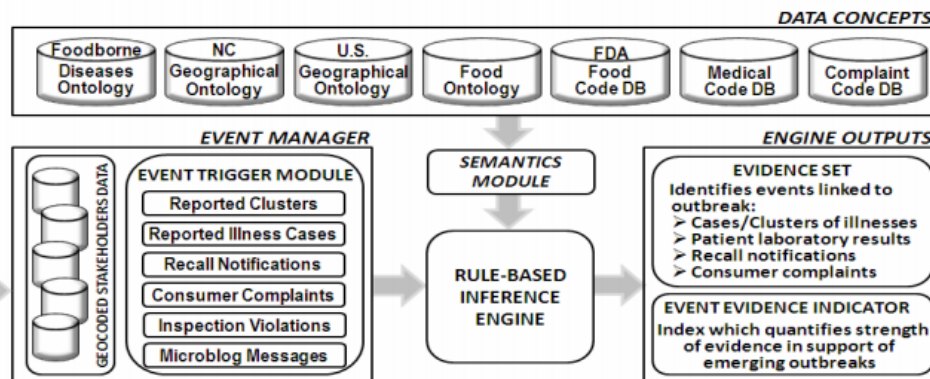
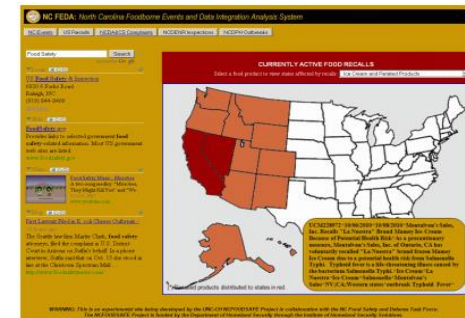
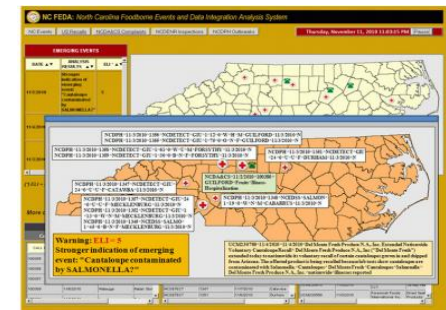
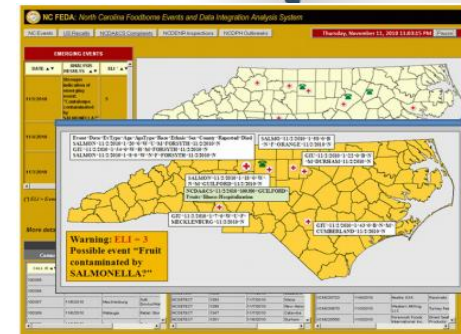
Intuitively, this rule means that there is evidence that complaint call C is connected to a materialized cluster of illness P affecting counties A and B, if call C can be linked to recall R of food F2 as it is of same type as food F1.

- **Event Evidence Indicator (EVI)** - *measure of strength of evidence supporting the conclusion of an emerging complex event. Ranking ranges from 0 to 7.*
 - *Based on the number and strength of relationships connecting events in the evidence set*
 - *Weighted summation of EVI components calculated for the subsets formed when linking pairs of different types of events*

NORTH CAROLINA FOOD EVENT DATA INTEGRATION & ANALYSIS SYSTEM

Engine consists of ASP program of:


- inference rules for food safety domain;
- new and stored facts describing situation being monitored;
- DLV solver.



- Arrival of new events triggers computation of new stable models of program to determine whether there is an emerging outbreak event occurring.

CONCLUDING REMARKS

- Framed outbreak detection problem as a complex event
 - ✓ structured event data (e.g. case information)
 - ✓ unstructured event data (e.g. recall or complaint data).
- Developed semantic models that extract meaningful information from unstructured text data that can serve as event triggers.
- Used ontologies and rules to discover semantic links between events that provide evidence of an emerging outbreak event.
- Used ASP to identify events that comprise the evidence set.
- Introduced new concept: *Event Evidence Indicator* that quantifies strength of evidence in support of an emerging event as a basis for response by public health officials.
- Implemented these concepts in the NCFEDA prototype.



Food Safety—Emerging Public-Private Approaches: A Perspective for Local, State, and Federal Government Leaders (Greis & Nogueira, 2010)

IBM Center for The Business of Government : www.businessofgovernment.org/

NCFOODSAFE: New Informatics Tools for Latency Reduction (Greis & Nogueira, 2010 & 2011)

Institute for Homeland Security Solutions: <https://www.ihssnc.org>

NCFEDA: Building Situational Awareness for Safer Food (Greis & Nogueira, 2011)

Center For Logistics And Digital Strategy: www.kenaninstitute.unc.edu/clds

Application of Answer Set Programming for Public Health Data Integration and Analysis (Nogueira & Greis, 2011) ARES2011 – In print.

Contact: monica_nogueira@unc.edu or noel_greis@unc.edu