



New Era
Partners™

Generating Synthetic FSMA 204 Traceability Data for Mock Investigations

Andrew Kennedy

THE TRACEBACK PROBLEM

Romaine lettuce · 2019 E. coli O157:H7
Outbreak · 474 illnesses · 215 hospitalizations ·
25 deaths · No lot code at point of sale



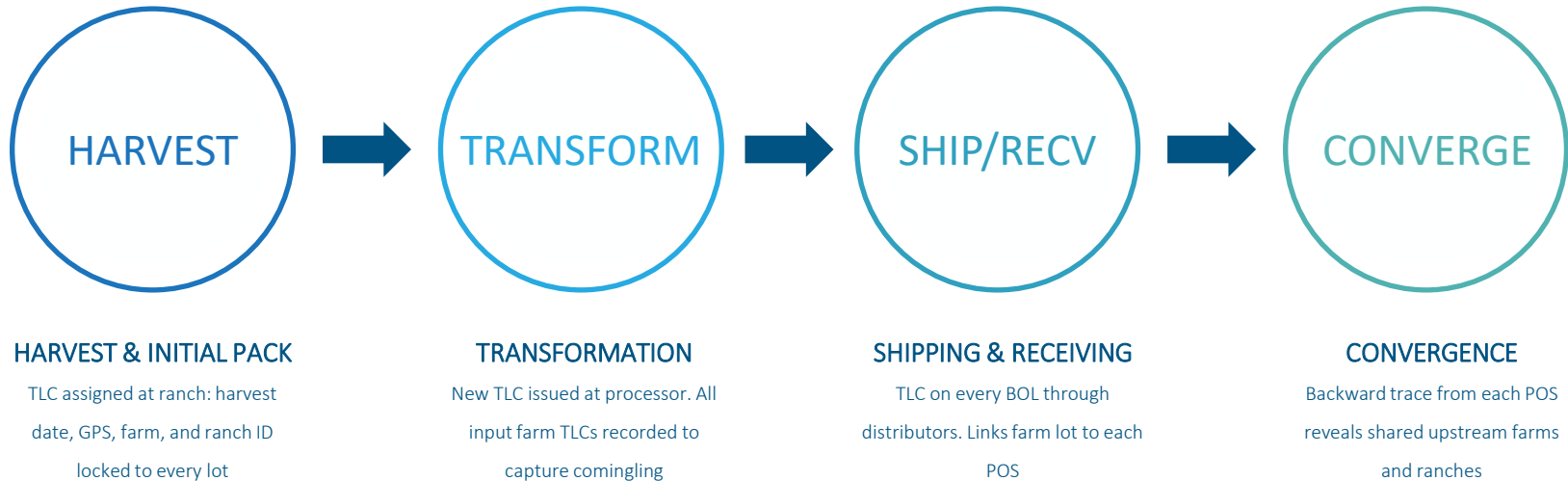
Source: An Overview of Traceback Investigations and Three Case Studies of Recent Outbreaks of Escherichia coli O157:H7 Infections Linked to Romaine Lettuce

<https://doi.org/10.4315/JFP-21-112>



FSMA 204: CRITICAL TRACKING EVENTS

CTE records link every lot from ranch to point of sale



CONVERGENCE ANALYSIS: SYNTHETIC DATA PROMPT

- As a FSMA 204 subject matter expert, I want to create a presentation on how to generate synthetic traceability data for mock investigations.
- The project includes a PDF with a research report written by Kari Irvin of FDA. Table 4 has an example of a traceback diagram for a 2018 Romaine Lettuce Outbreak Investigation with locations investigated and linked by traceability data.
- In the attached template I created examples in the Harvest and Shipment tabs of what synthetic data should look like.
- Based on this information, create a spreadsheet with synthetic data including Harvest, Initial Packing (in Fields), Shipping, Receiving, and Transformation (at Processors).
- Traceability Lot Codes (GS1 GTIN-14 and Lot Number) should connect locations to each other.
- Traceability Lot Code Source should be the name of the field location where the romaine was harvested and initially packed or transformed from whole head romaine into bagged romaine (washed and chopped).
- This information will be fed into software designed to visualize the supply chain and determine convergence.
- Use the research report to provide additional details that may help you to create more realistic synthetic data.
- Output should be an xlsx file.
- The first tab should provide a narrative of how you created the file and how it should be used by analysis software to map it and analyze it.

HOW CLAUDE BUILT THE SYNTHETIC DATA

01

STEP 1: DESIGN THE SUPPLY CHAIN

Map farms, ranches, processors, distributors, and POS locations to a real outbreak scenario (Irvin et al. 2021, Fig. 4 & 6)

02

STEP 2: ASSIGN LOT CODES

Generate GS1 GTIN-14 + Lot Numbers per ranch per harvest day. TLC Source = originating field name

03

STEP 3: POPULATE CTE SHEETS

Build Harvesting, Initial Packing, Transformation, Shipping, and Receiving tabs with pipe-delimited input TLCs

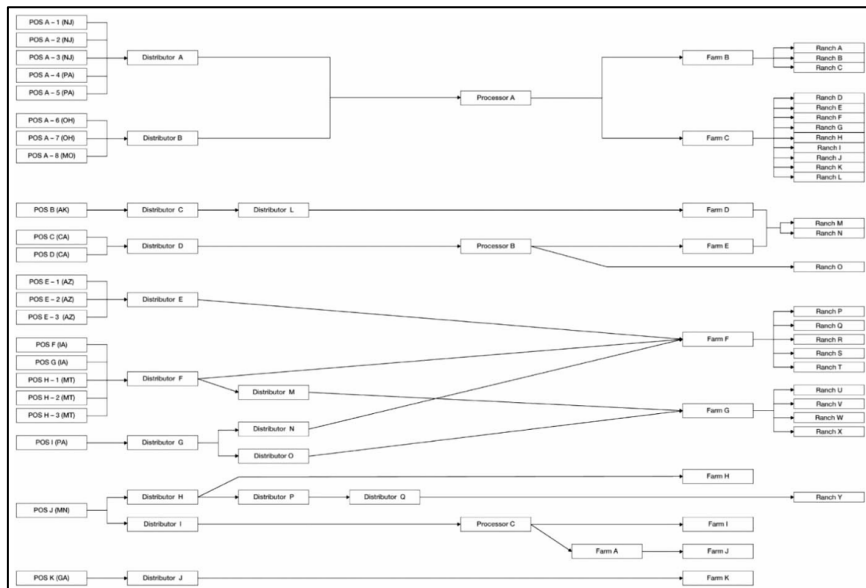
04

STEP 4: VALIDATE CONVERGENCE

Confirm backward tracing from each POS reaches Farm A, and POS F never traces to Farm A

CONVERGENCE ANALYSIS: SYNTHETIC DATA

Input to Claude



Output from Claude

	A	B	C	D	E	F	G
FSMA 204 - Shipping CTE-5 Figure 6 Synthetic Dataset - All Supply Chain Stages							
Supply Chain Stage	Traceability Lot Code (TLC)*	Product Description*	Brand	Qty Shipped*	UOM*	Ship Date*	
1a-Farm—Aggregator	GTIN.10011111100001.LOT.RA-20191101-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	195	CS	2019-11-02	Sunrise V
1a-Farm—Aggregator	GTIN.10011111100001.LOT.RA-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	196	CS	2019-11-03	Sunrise V
1a-Farm—Aggregator	GTIN.10011111100001.LOT.RB-20191101-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	184	CS	2019-11-02	Sunrise V
1a-Farm—Aggregator	GTIN.10011111100001.LOT.RB-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	178	CS	2019-11-03	Sunrise V
1a-Farm—Aggregator	GTIN.10011111100001.LOT.RC-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	170	CS	2019-11-03	Sunrise V
1a-Farm—Aggregator	GTIN.10011111100001.LOT.RC-20191103-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	164	CS	2019-11-04	Sunrise V
1a-Farm—Aggregator	GTIN.10011111100001.LOT.RD-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	182	CS	2019-11-03	Sunrise V
1a-Farm—Aggregator	GTIN.10011111100001.LOT.RD-20191103-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	176	CS	2019-11-04	Sunrise V
1b-Aggregator—Processor	GTIN.10011111100001.LOT.RA-20191101-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	195	CS	2019-11-04	Central VA
1b-Aggregator—Processor	GTIN.10011111100001.LOT.RB-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	196	CS	2019-11-05	Central VA
1b-Aggregator—Processor	GTIN.10011111100001.LOT.RB-20191101-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	184	CS	2019-11-04	Central VA
1b-Aggregator—Processor	GTIN.10011111100001.LOT.RB-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	178	CS	2019-11-05	Central VA
1b-Aggregator—Processor	GTIN.10011111100001.LOT.RC-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	170	CS	2019-11-05	Central VA
1b-Aggregator—Processor	GTIN.10011111100001.LOT.RC-20191103-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	164	CS	2019-11-06	Central VA
1b-Aggregator—Processor	GTIN.10011111100001.LOT.RD-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	182	CS	2019-11-05	Central VA
1b-Aggregator—Processor	GTIN.10011111100001.LOT.RD-20191103-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	176	CS	2019-11-06	Central VA
1-Farm—Processor	GTIN.10011111100001.LOT.RE-20191101-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	156	CS	2019-11-02	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RE-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	152	CS	2019-11-03	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RF-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	166	CS	2019-11-03	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RF-20191103-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	160	CS	2019-11-04	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RG-20191102-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	190	CS	2019-11-03	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RG-20191103-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	184	CS	2019-11-04	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RH-20191103-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	174	CS	2019-11-04	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RH-20191104-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	168	CS	2019-11-05	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RI-20191103-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	180	CS	2019-11-04	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RJ-20191104-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	174	CS	2019-11-05	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RJ-20191103-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	150	CS	2019-11-04	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RJ-20191104-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	144	CS	2019-11-05	Sunrise V
1-Farm—Processor	GTIN.10011111100001.LOT.RK-20191104-001	Romaine Lettuce, Whole Head, 50-lb Ca	SunValley Brand	162	CS	2019-11-05	Sunrise V



Source: An Overview of Traceback Investigations and Three Case Studies of Recent Outbreaks of Escherichia coli O157:H7 Infections Linked to Romaine Lettuce

<https://doi.org/10.4315/JFP-21-112>

CLAUDE'S PROMPT TO PERPLEXITY COMPUTER

I asked Claude to help me create a prompt to build the app in Perplexity

Prompt: Build an FSMA 204 Supply Chain Traceback Analysis & Visualization Tool

CONTEXT

I have an Excel workbook (`FSMA204_Synthetic_Traceback_Data.xlsx`) containing synthetic food traceability records formatted to the FDA's Food Safety Modernization Act (FSMA) Section 204 standard. The data simulates a fictional multistate foodborne illness outbreak linked to romaine lettuce from the Salinas Valley, California (modeled after real 2019 outbreak investigations).

I need a **Python web application** that ingests this file, builds a supply chain graph, detects convergence (the common source across multiple illness locations), and visualizes the full traceback interactively.

GOAL

Build a **Streamlit web application** (single `app.py` file with a `requirements.txt`) that:

- Accepts the Excel file via drag-and-drop upload
- Parses all 5 data tabs and builds a directed supply chain graph
- Identifies and ranks convergence points across all Points of Sale (POS)
- Displays an interactive, node-link supply chain visualization
- Generates a traceback investigation summary report



RESULT: PERPLEXITY APP, TRY IT OUT



FSMA 204 Supply Chain Traceback

Entity-level supply-chain graph, convergence analysis, TLC ledger, and auto-generated investigation summary for FSMA Section 204 traceback investigations.

Controls

Uses the bundled synthetic Salinas Valley workbook by default. Upload another FSMA 204 workbook to rerun the traceback.

Workbook (.xlsx)

Choose File No file chosen

TIMEFRAME

Start date

10/14/2019

End date

10/26/2019

MAP OPTIONS

- Show ranch nodes
- Show farm nodes
- Highlight contaminated TLCs

Color nodes by

Entity Type

Minimum POS coverage: 50%

Traceback for POS

(none)

Run traceback analysis

Reset to sample workbook

34

Nodes

42

Edges

15

POS endpoints

9

Contaminated TLCs

153

Ledger rows

Supply Chain Map

Convergence Analysis

TLC Traceback Ledger

Investigation Summary

Supply Chain Map — entity-level DAG (hover for TLC detail)

