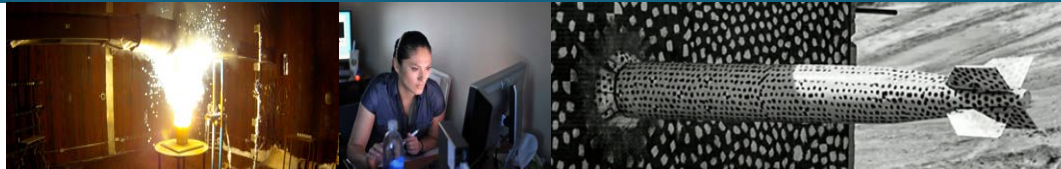




Sandia  
National  
Laboratories

# Practical Data Analytics for CBP



Steve Morrison  
Sandia National Laboratories  
[semorri@sandia.gov](mailto:semorri@sandia.gov)



Sandia National Laboratories is a  
multimission laboratory managed and  
operated by National Technology &  
Engineering Solutions of Sandia, LLC, a  
wholly owned subsidiary of Honeywell  
International Inc., for the U.S. Department  
of Energy's National Nuclear Security  
Administration under contract DE-  
NA0003525.

# Introduction



Erin the bartender and the fantasy football league

- “You know football, I know math. Let’s work together.”

Data scientists don’t understand CBP’s world. CBP often doesn’t understand how to apply data science. Collaboration is key.

This talk shares several lessons learned over years of working in data analytics with CBP.

# 1. You need data to do data analytics.

## Example #1: Border crossing anomaly detection

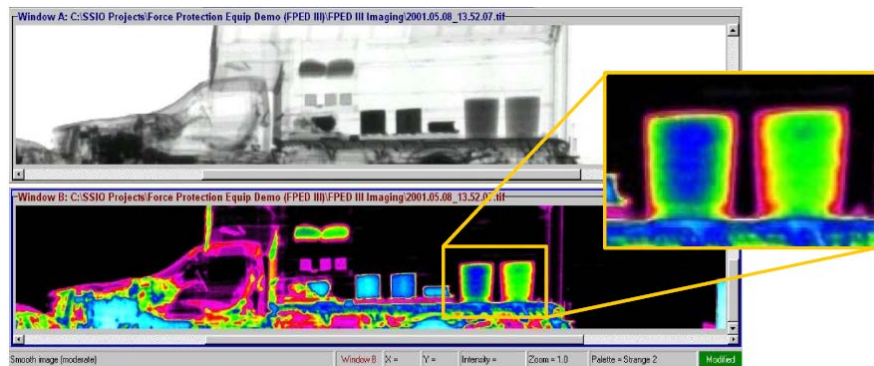
- We wanted to look for patterns/anomalies, but learned that south-bound crossing data is not even recorded.



$$T_{J \rightarrow I} = \sum p(i_{n+1}, i_n^{(k)}, j_n^{(l)}) \log \frac{p(i_{n+1} | i_n^{(k)}, j_n^{(l)})}{p(i_{n+1} | i_n^{(k)})}$$

## Example #2: Automated threat recognition modules

- Supervised machine learning needs ground truth, but no database of seizure results exists.



## 2. Correlating multiple data sets may yield much more interesting results than analyzing a single data set.

### Example #1: Correlate phone data with border crossing data

- Phone records or border crossing data can be interesting by themselves, but much more interesting when we correlate the information.
- Who are people calling after they cross the border? If they have a pattern but then deviate, who are they calling?

### Example #2: Quality of evidence analytic

- Data triage based on presence of person's data showing up in multiple data sets

- PN: 1603903XXXX (Unpursued)  
QE: 1.426
- PN: 162420XXXX (Arrested)  
QE: 1.418
- PN: 1904481XXXX (Arrested)  
QE: 1.416
- PN: 1425486XXXX (Arrested)  
QE: 1.415
- PN: 160533XXXX (Arrested)  
QE: 1.414

Call Date: 02/14/1999

	Dialed No.	Call Time HR:MN:SC	Call Duration HR:MN:SC	Cell Site
1	4432539023	10:26:55 PM	00:00:28	L651C
2	4432539023	10:26:55 PM	00:00:28	L651C
3	incoming	10:26:55 PM	00:00:28	TPA44
4	4432539023	09:35:59 PM	00:00:28	BLTM2
5	incoming	09:35:59 PM	00:00:28	TPA44
6	41C [REDACTED]	09:04:56 PM	00:00:57	L651C
7	incoming	07:48:43 PM	00:00:16	L655B
8	301 Nisha	07:17:30 PM	00:10:14	L608C
9	incoming	06:28:30 PM	00:09:01	L655B
10	incoming	06:09:53 PM	00:00:20	L608C
11	[REDACTED]	06:09:06 PM	00:00:09	L608C
12	[REDACTED]	01:18:46 PM	00:01:10	L601E
13	[REDACTED]	11:28:11 AM	00:11:48	L602C
14	[REDACTED]	11:09:53 AM	00:00:40	L608A
15	4432539023	11:06:34 AM	00:00:04	BLTM2
16	incoming	11:06:34 AM	00:00:04	TPA44
17	4432539023	03:08:13 AM	00:00:05	BLTM2
18	incoming	03:08:13 AM	00:00:05	TPA44

TOTAL USAGE: 00:37:13



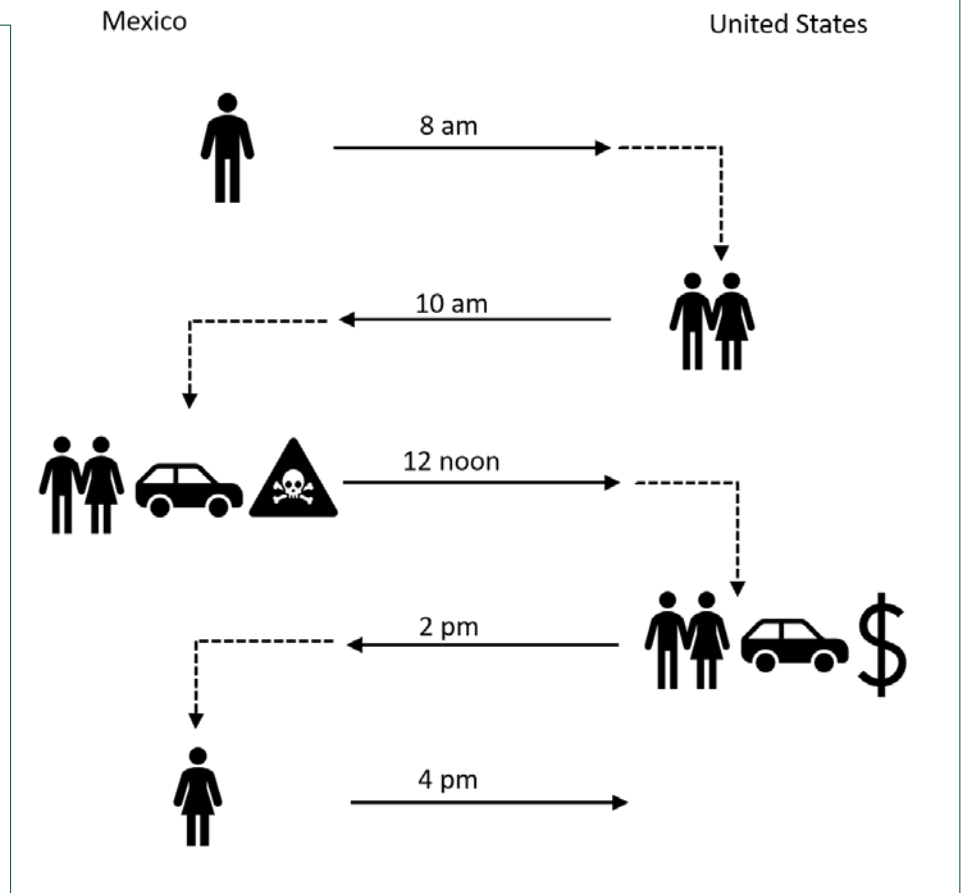
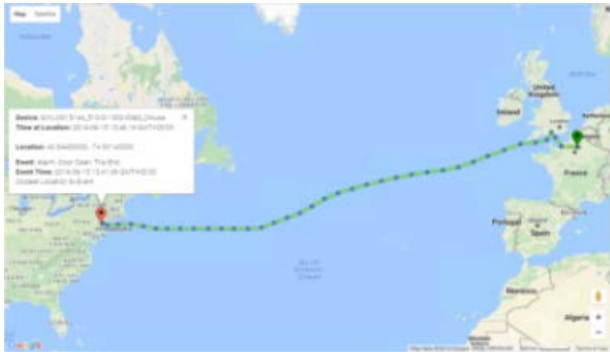


### 3. Data analytics are more powerful when agencies work together to share data.

Example #1: Homeland Security Investigations identifies patterns of behavior based on post-seizure analysis. Why not use those patterns to interdict?

#### Example #2: GPS tracking data from industry

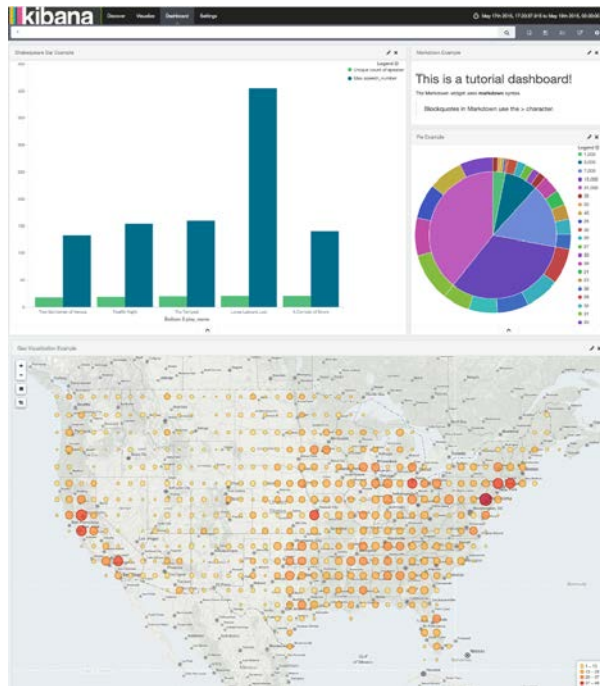
- Could be used to identify suspicious behavior in the supply chain



## 4. Analytics are no good if you can't create a usable tool for the end user.

### Example #1: Drug seizure analysis

- User: “How do I search?”
- Developer: “You can query the ElasticSearch node using REST API calls through the Kibana dashboard developer tool window.”
- User: “Huh?”



### Example #2: Third-party software, IP, licensing fees

- Software applications can lock the customer into a dependence on a third party





## 5. Analytics are no good if you don't use the results.

### Example #1: In-bond electronic seal pilot

- Truck went outside designated corridor between Laredo and Detroit, but no policy was in place to handle the violation.



### Example #2: Multi-energy portal installations

- When portals are installed in pre-primary, what shifts will we see in commercial commerce at the port and neighboring ports?





Questions?

