PULSENET WEB-PORTAL: TRUTH OR URBAN LEGEND??

Dave Boxrud

Minnesota Department of Health

On behalf of the Web-Portal Users Group

As a participant in PN, you do many things to identify and track clusters...among them...

- Produce and upload data
- Log on to PN, download image data, view and compare, assess, loop your epis in, etc...
- This is a **necessary, but time-consuming** process
- Information is not easily accessible
- Limited number of users due to data restrictions on the system

Downloaded PulseNet Info
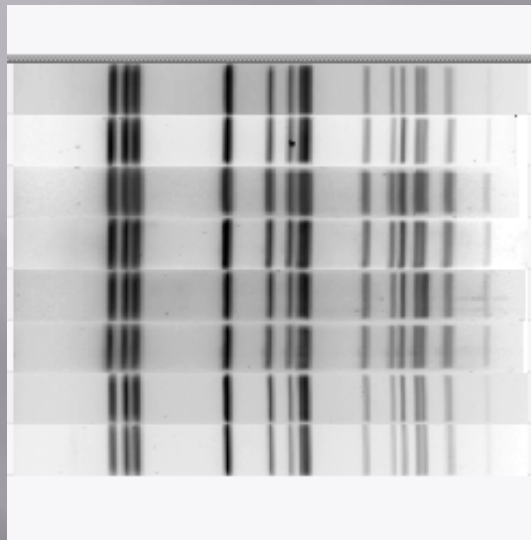
Key	Serotype	SourceCity	SourceCounty	SourceState	SourceCountry	Pati
GA___12C0240277_1	I 4,[5],12:i:-			GA	USA	38
SC___SALM11-0182_1	I 4,[5],12:i:-		ANDERSON	SC	USA	55
SC___SALM12-1521_1	I 4,[5],12:i:-		BEAUFORT	SC	USA	3
SC___SALM12-1562_1	Typhimurium		ORANGEBURG	SC	USA	2
SC___SALM13-0040_1	I 4,[5],12:i:-		SPARTANBURG	SC	USA	0- 10
SC___SALM13-0223_1	Serotype pending		SUMTER	SC	USA	1
VA___R100905980_1	I 4,[5],12:i:-	Norfolk		VA	USA	3
VA___R110516500_1	I 4,5,12:i:-	Roanoke	Roanoke City	VA	USA	55

We need something…

- Better?
- Easier?
- Faster?
- Automated?
- To allow for real time, streamlined analysis of large amounts of data

Web Portal–Urban Legend

2004 — San Diego
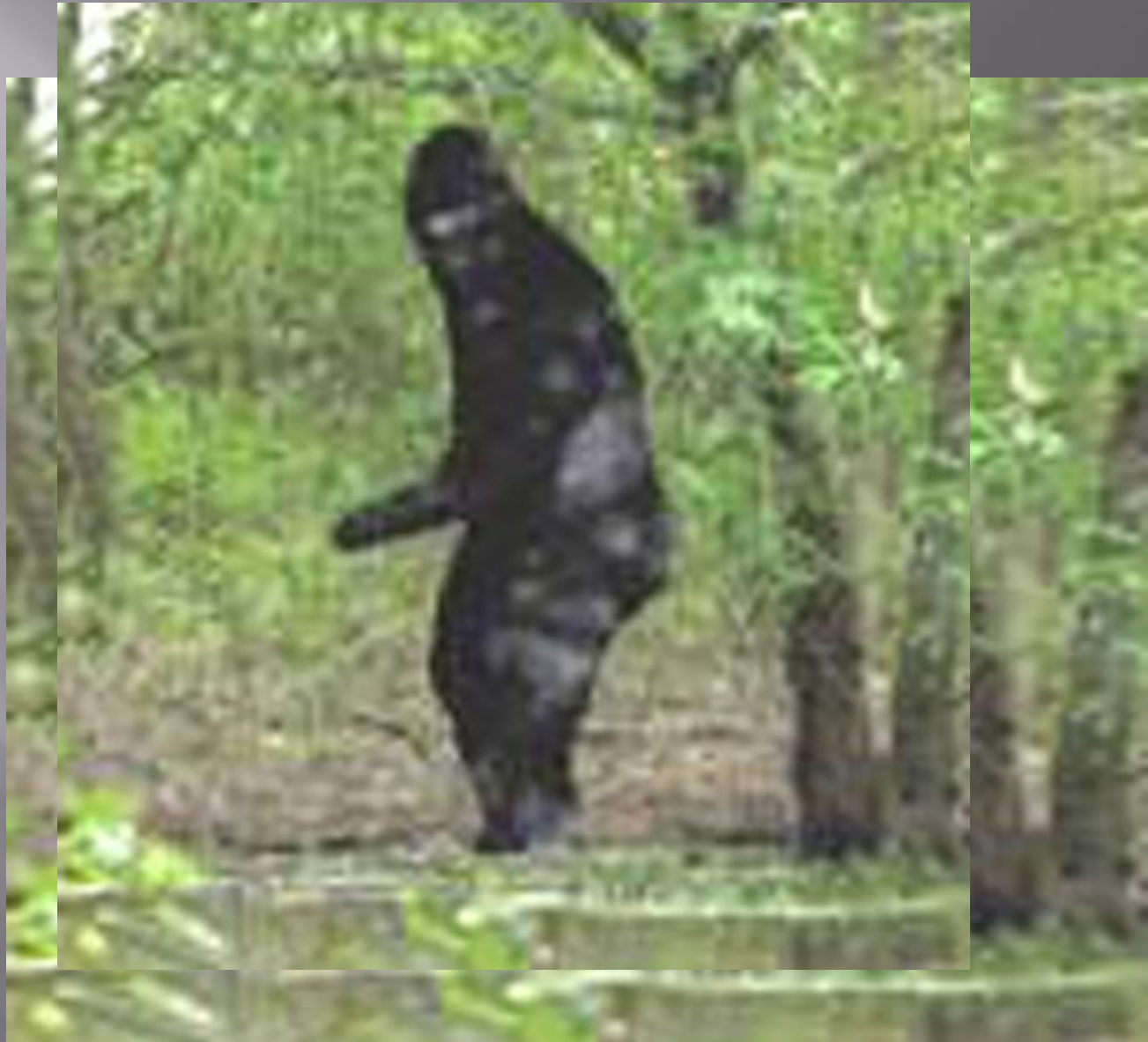
CALIFORNIA
The Golden State

BioNet

2010-Chicago

2011-Long Beach

2009-Utah

The WebPortal? More attainable than Bigfoot? Maybe…

Home

Introduction to T-Cube

T-Cube levera
and interactive
infeasible.

Modern Data Warehouses are often tombs of data

In practice, modern Data Warehouses are often tombs of data. While BI tools improve its visibility, they do little to support real-time trending and detection of emerging patterns across multiple dimensions and streams. It is because typical software response times are too long for large scale interactive investigations. Ordinarily, slow retrieval of data limits Information Systems to simple analyses of relatively small sets of data with a small number of variables considered at one time. Important patterns can therefore be missed or detected too late. There is a practical need for new approaches to mitigate this risk.

T-Cube addresses the risk of missing patterns by caching statistical distributions of data across time and multiple dimensions. It enables comprehensive processing of very large multivariate data at near-interactive speeds. This allows for automated detection of complex patterns and resolves "we don't know what we don't know" dilemma. T-Cube enables processing complex data orders of magnitude faster than commercially available tools. It is a game changer in multiple application domains including monitoring health of equipment (USAF and Navy aircraft) and food safety (USDA, CDC: pathways of transmission of Salmonella to humans). New applications are found each year.

Our interactive web-ba
the hood. It enables use
integrated drill-down an

Web-based software that can quickly perform complex ad hoc queries on complex data sets

Web-Portal, take 6…

- In 2007 Carnegie Mellon began working on T-cube portal for USDA
- CDC joined with USDA to collaborate on a system to combine non-image datasets
- After ironing out some high level wrinkles, CDC and USDA began working on making the data transfer from PN to T-cube housed at USDA a reality
- 2011-Sixteen members of working group have received access retrospective PN data via T-cube

Proposed Data Flow for Web-Portal

WY_E20120025692   Berta   WY   Stool   USA   3
Cheyenne

Assign CDC Pattern

CDC

Proposed Data Flow for Web-Portal

CDC
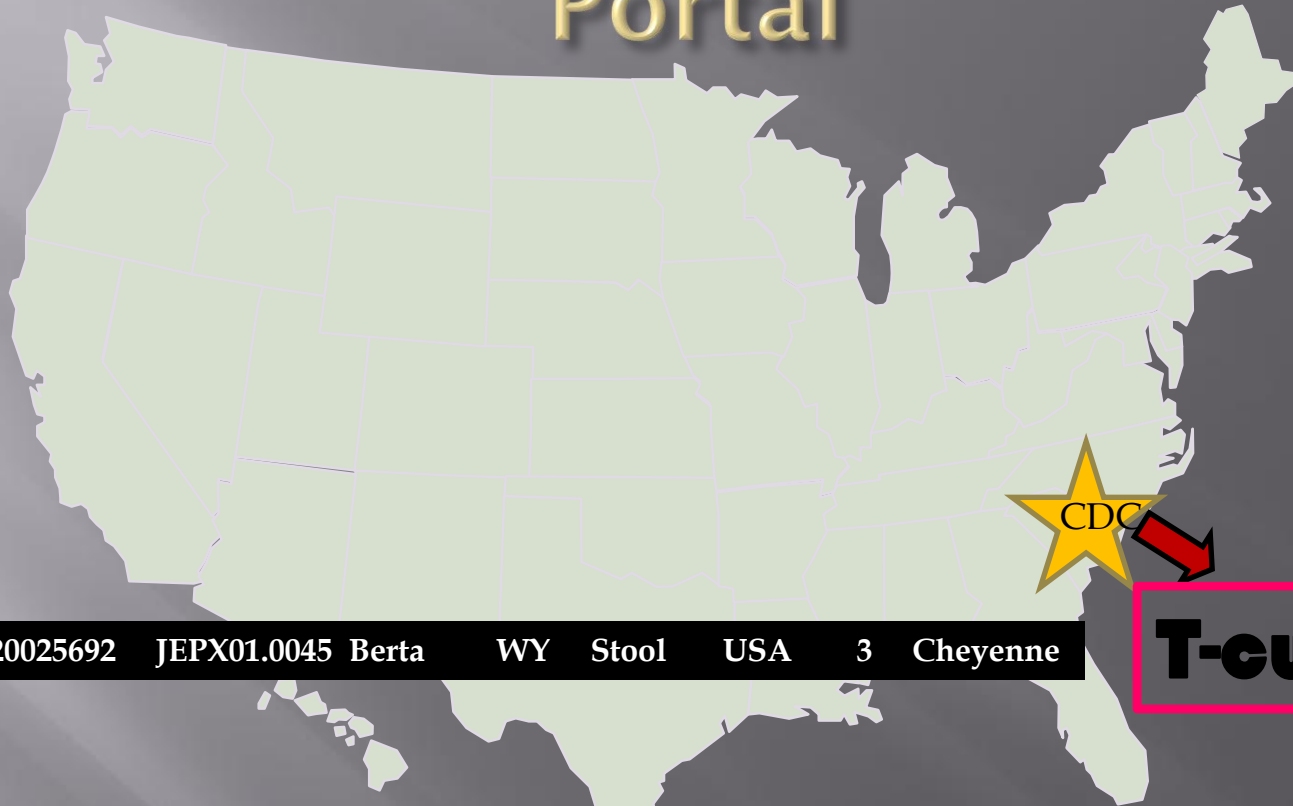
T-cube

WY_E20120025692	JEPX01.0045	Berta	WY	Stool	USA	3	Cheyenne

With the next download of data from PN to T-cube (housed at USDA), the pattern desiganation and specific isolate information goes…the big, bulk image data stays behind.

T-cube Capabilities For PulseNet Data

- Ad hoc queries
- Map out cases within state, regions or country
- ID trends in patterns, specific areas, etc.
- Rapid detection of clusters while still local or regionally relevant.
- Create tables, line lists, epi reports, etc…
- Many, many others

Web-Portal User Group

- APHL/CDC established group to look at T-cube capabilities and make recommendations for how it could be used for PN info

- Training session w/ CMU in August of 2011

- Upon completion of training session, group agreed to provide feedback to CMU to enhance the system and make it more 'user-friendly'

- Began with training calls for those who missed Pittsburg, and developed into smaller group to address functionality of T-cube

Post-training Assessment of T-cube

- T-cube has huge capabilities
- Very complex for regular usage
- Tremendous potential for local or regional cluster detection
- For PN Central-very sophisticated ways to identify clusters/trends
- Local/regional needs not always the same as national needs

Proposed Streamlined Web Portal Output

- Identified information that would be routinely requested
- States/localities/regions can have reports created daily based on previous day's upload
- Routine reports are run automatically and emailed to specifically identified staff or a shared mailbox
- Ad hoc reports easier to produce

Auton Lab | T-Cube Web Interface
for fast extraction of time series from large datasets

Regular Case Report

This report returns a list of cases matching a query. Where applicable, user may navigate to Cluster Detail Report and/or PN outbreak Code details report.

Cluster Details Report

This report captures details of a single cluster. It presents high level profile information of a cluster, also a line list of cases associated with this cluster.

Cluster Summary Report

This report gives a summary of currently developing clusters, with high level description of each cluster. Through hyperlink, users may navigate to Cluster Details

Cluster Tracking Case Report

This report simplifies the manual tracking process where a user needs to look up a pattern for a cluster, then run a regular case report as the one described in regular case report.

Regular Case Report

This report is generated on   12/31/2012

This report can be run from a template with a list of saved parameters or from "adhoc" query. It can also been subscribed and run automatically on a daily basis and the resulting report will be sent to an email address;

If any of cases are associated with a cluster (i.e. flagging), the Cluster ID field is populated, and shown as a hyperlink, clicking on it will open up a report in the format of "Cluster Details Report" (Template A). This report can be queried by state/region, date range, cluster ID, serotype and pattern. The report is sortable by any of the fields.

The report can be exported to exernal file format such as PDF, CSV etc.

Load Saved Report Template

my_saved_customized... ▼

Adhoc Query Report

| **State** | MA,MN,CA ▼ | **Region** | North East ▼ | **Date** | **From** [   ] 🔳 | **To** [   ] 🔳 | **Or On** [   ] 🔳 |

| **Cluster ID** | X10020 ▼ | **Serotype** | Enteritidis ▼ | **PFGE-Xbal Pattern** | JEGX01.0004 ▼ |

Save This report Template as :   [   ]   **Save Template**

Subscribe this report

E-mail   [   ]   **Subscribe**

Export this report

File Name   [   ]   Browse...

Lab ID	KeyID	Date	SourceCity	Source County	Source State	Patient Age	Serotype	PFGE-Xbal Pattern	PFGE-Binl Pattern	Cluster ID	PN Outbreak Code
sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	*X10020*	*1211PAJPX-1*
sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	NA	NA

Cluster Detail Report

Date of Report   12/14/2012

Cluster ID   X10020 ▼

Serotype   Heidelberg

PFGE Pattern   JPXX01.0022

Time Series and Map

This report captures details of a single cluster. It can be retrieved through cluster ID, or through clicking on a hyperlink of "cluster ID" in "Cluster Summary Report". This report will be updated when data are updated. This report is broken down into sections of narrative, time series chart/map/pivot table and table of case details. Case details can be filtered by state/region and date range. The report can be available as html page, may be converted to PDF file. Time series and map section can be hidden at user's choice.

Significant Characterization

Gender	Percentage
Female	70%
Male	30%

Cluster Description   This cluster starts from 12/1/2012, occurred in states CA, OR, WA. In last 7 days, there are 10 counts of isolates of this pattern, the expected count from historical baseline (*) is 5 counts, and the majority of patients are woman (70%).

State   CA / PA / OR / MI   **Region**   West / Mountain / North Central / South Central   **Date Range**   From [ ]   To [ ]

Lab ID	KeyID	Date	SourceCity	Source County	Source State	Patient Age	Serotype	PFGE-Xbal Pattern	PFGE-BInI Pattern	PN Outbreak Code
sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	*1211PAJPX-1*
sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text

Cluster Summary Report

This report is generated on 12/31/2012

State | CA / PA / OR / MI

Region | West / Mountain / North Central / South Central

Date Range **From** [ ] **To** [ ]

Cluster ID | X10020

Serotype | Enteritidis

PFGE-Xbal Pattern | JEGX01.0004

Cluster ID	New/Update	Serotype	PFGE Pattern	Starting Date	Ending Date	Number of cases	Baseline Counts	States	Significance Score	Link to ARS	Link to similar historical cluster
X10020	New	Heidelberg	JPXX01.0022	12/1/2012	ongoing	30	10	PA	0.99	Yes	X01023 X33023
X10022	Update	Dublin	JDXX01.0005	11/24/2012	ongoing	4	0	CA,WA	0.7	No	N/A

* In case of a multiple pattern cluster, multiple patterns will be reported in this column; when filtering by pattern, as long as the requested pattern is on the pattern list to be displayed, the cluster will be shown too.

** In order to provide information of this kind, we need to have a historical archive of clusters. Additionally, similarity score(s) need(s) to be defined, that could be used to retrieve K clusters with the closest resemblance to the current cluster.

*** Baseline counts depend on the kind of massive screening algorithm in use. If disjunctive query algorithm is enabled, the baseline counts can be a subset of the states/regions; however, if the screening is state by state, the baseline counts are the same state counts; if the screening is region by region, the baseline counts are the regional counts.

Cluster Tracking Case Report

This report is generated on   12/31/2012

Given a cluster ID, which cannot be empty, this report returns all the cases that share the same pattern as those in that particular cluster. What is different from the regular case report is that it also report the cases which are not necesarily part of the current cluster however share the same pattern name with the current cluster.

The report simplifies the manual tracking process where a user needs to look up a pattern for a cluster, then run a regular case report as the one described in Regular Case Report.

The report can be further filtered by State/Region/Date Range. Like in other reports, if a case is associated with a cluster, the Cluster ID field will be populated, and shown as a hyperlink. Clicking on the hyperlink will open up a report in the format of "Cluster Details Report"

What is the cluster you want to track ? (this field can't be left blank)

Cluster ID   X10020

Those are serotype(s)/pattern(s) inovled in this cluster:

Enteritidis/JEGX01.0001

Case Report on above listed Pattern(s):

Lab ID	KeyID	Date	SourceCity	Source County	Source State	Patient Age	Serotype	PFGE-XbaI Pattern	PFGE-BlnI Pattern	Cluster ID	PN Outbreak Code
sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	*X10020*	*1211PAJPX-1*
sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	sample text	NA	NA

Report Filters

CA	West	
PA	Mountain	
OR	North Central	
State MI	**Region** South Central	**Date Range** **From** _____ **To** _____

Future of the Web-Portal

- T-cube is scheduled to go live on September 30th

- No, we aren't kidding!

- Over the course of the next year, users will continue to be authenticated and granted access by USDA

- Data will be transferred from PulseNet to T-cube in regular intervals, with the goal to be a nightly download from CDC to USDA

- T-cube automated reports previewed today…addition to T-cube to be determined after review by USDA

Limitations

- Need accurate info in PulseNet
- Delay in pattern naming will delay info sent to T-cube
- Portal can be a cluster detection tool-does not replace rapid, thorough case interviews
- Currently only for Salmonella

Web-Portal Workgroup

- Peter Gerner-Smidt, CDC
- John Besser, CDC
- Kristy Kubota, APHL
- Carol Sandt, PA DPH
- Brian Sauders, NYAG
- Laura Kornstein, NYC
- Beth McGlinchy, CDC
- Steve Dietrich, MI
- Thai-An Nguyen, CDC
- John Copeland, CDC
- Brenda Brown, CDC
- Kelley Hise, CDC
- Tracy Stiles, MA
- Dave Boxrud, MN
- Deb Sheehan, APHL/CDC
- ***Artur Dubrawski
- ***Karen Chen
- Jeshua Pringle
- Dianna Bopp
- Denise Toney