Visual Twitter Analytics for Disease Surveillance

COURTNEY D CORLEY, PHD

Pacific Northwest National Laboratory

Richland, WA

@CourtCorley court@pnnl.gov
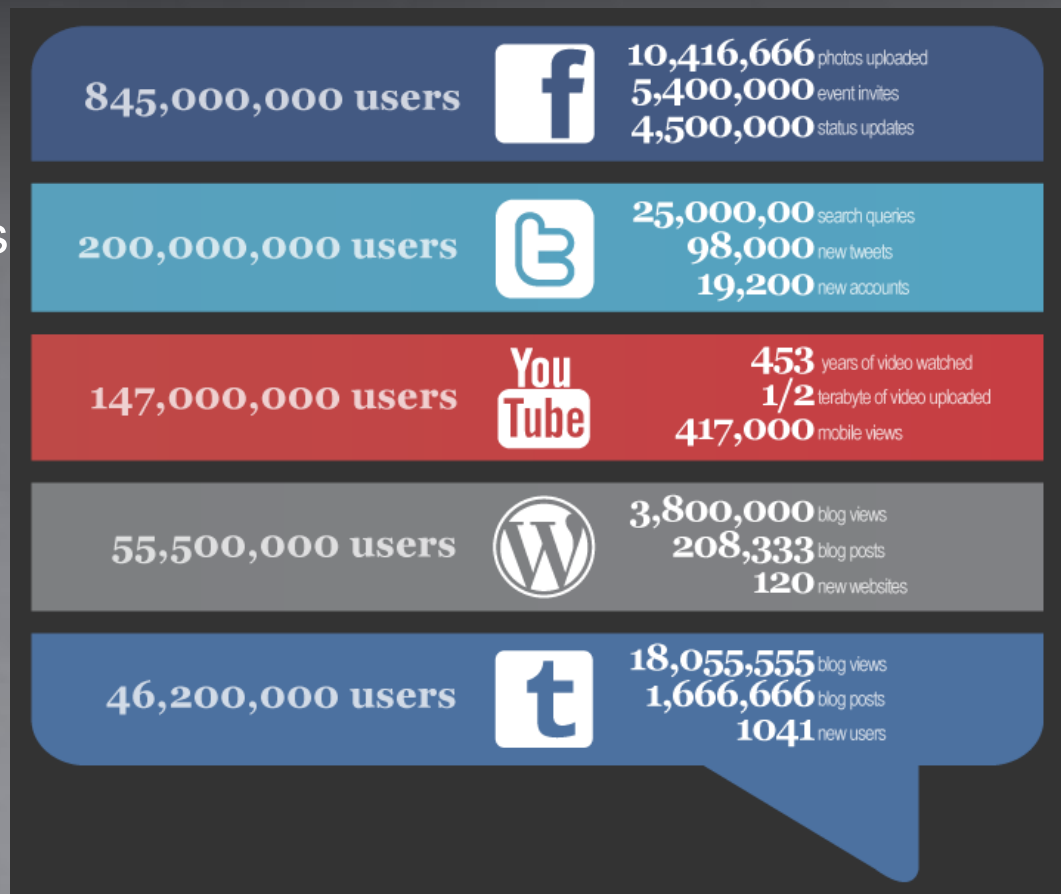
Outline

▶ Introduction

▶ Social Media Characteristics

▶ How to Get Started

▶ Immunization Safety Beliefs

▶ Prototype Visualization

▶ Summary

845,000,000 users — 10,416,666 photos uploaded / 5,400,000 event invites / 4,500,000 status updates

200,000,000 users — 25,000,00 search queries / 98,000 new tweets / 19,200 new accounts

147,000,000 users — 453 years of video watched / 1/2 terabyte of video uploaded / 417,000 mobile views

55,500,000 users — 3,800,000 blog views / 208,333 blog posts / 120 new websites

46,200,000 users — 18,055,555 blog views / 1,666,666 blog posts / 1041 new users

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds

The Pacific Northwest National Laboratory

We deliver solutions

to America's most

intractable problems in

energy, national security

and the environment.

Through the power of our

interdisciplinary teams,

we advance science and

technology to make the

world a better place.

► Proudly operated by Battelle since 1965

► More than 4,900 staff

► State-of-the-Art User-Facilities

► Mission-driven collaborations with government, industry, and universities

Sequim

Seattle

Washington, D.C.

Richland
MAIN CAMPUS

Portland, OR

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

The essence of the New Biology is re-integration of the subdisciplines of biology, along with greater integration with the physical and computational sciences, mathematics, and engineering in order to devise new approaches that tackle traditional and systems level questions in new, interdisciplinary, and especially, quantitative ways.
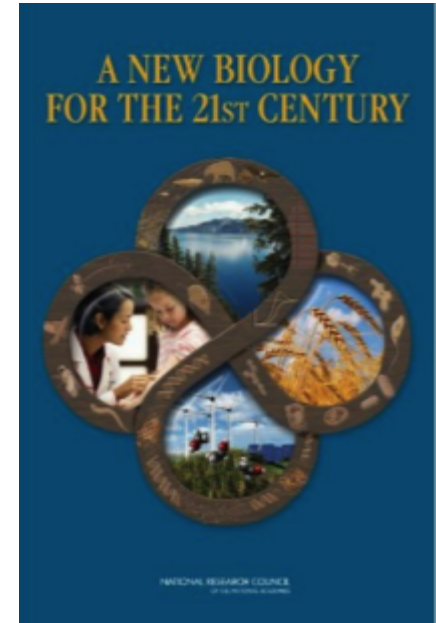
Source: A New Biology for the 21st Century, pg. 17, National Academies Press and the Committee on a New Biology for the 21st Century, 2009.
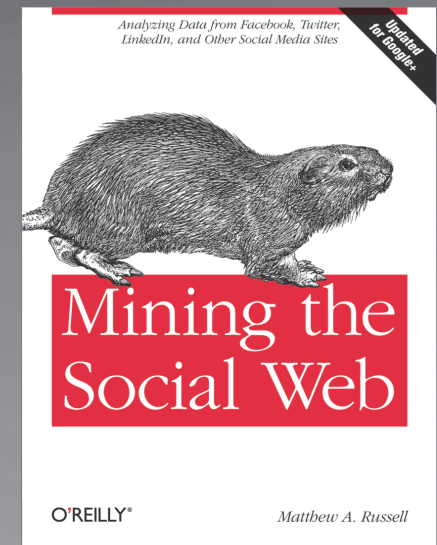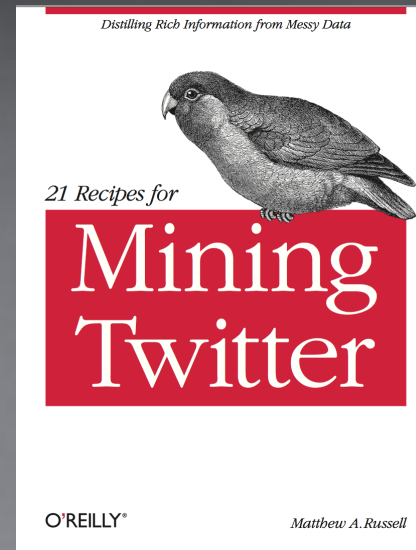
Social Media Characteristics, Premise and Goals

▶ Low cost drives volume up and value down

▶ Noise generated by spam bots

▶ Propagation of signals

   ■ Driven by perceived importance (subjective)

   ■ Amplified by visibility

▶ Different time scales have different meaning

   ■ Temporal aggregation → Minutes, hours, days, months

▶ Variability in usage

   ■ Country, domain, culture impacts use

▶ Social media platforms as global sensor network

   ■ Information propagates into and from social media

▶ Information emitted by individuals

   ■ located in time and space & motivated by interests

   ■ connected through common characteristics

▶ Augment public health analyst with computational methods

   ■ Interaction speed on high volume data with appropriate measurement

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds

How to Get Started with Twitter Data

▶ Use the Twitter Search API
- ■ Superb Learning Resources
  - ● GitHub
  - ● O'Reilly
  - ● http://dev.twitter.com
  - ● limited to 1% random sample

▶ Data Provider
- ■ Expensive
  - ● $0.10 / 1,000 tweets
  - ● ~$8,000 / month for 10% random sample
- ■ Requires Substantial Computational Infrastructure

Distilling Rich Information from Messy Data

21 Recipes for

Mining Twitter

O'REILLY® *Matthew A. Russell*

Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites

Updated for Google+

Mining the Social Web

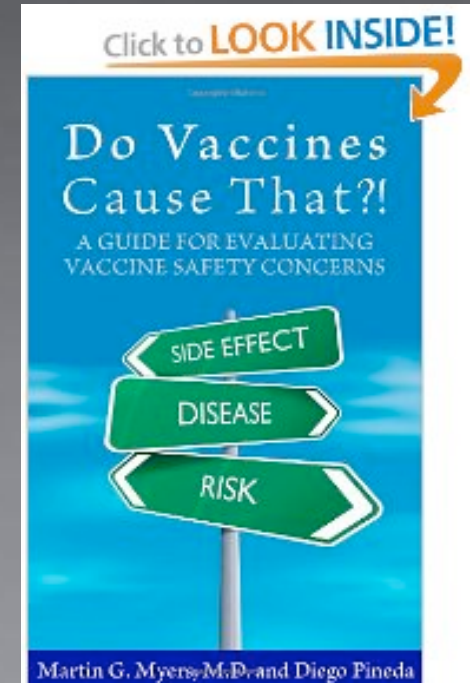O'REILLY® *Matthew A. Russell*

Immunization Safety Beliefs

► Motivation:

  ■ Because search engines only identify topics and list relevant Web sites, the identified information may or may not be reliable. As a consequence, parents with concerns about vaccine safety are as likely to locate misinformation as they are accurate information.

  ■ In social media, they have the opportunity to express their concerns and discuss their findings whether they are informed or misinformed.

► Research Question:

  ■ Can we analyze social media conversations in order to explore whether we can identify emerging concerns about vaccine safety in real time and then develop strategies to address those concerns.

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds

Visual Analytics

Why Visual Analytics

- Identify related threads, intersections of topics
- Investigate backstory
- Quantify convergence of data on a particular topic or event
- Forecast and/or assess impact of events
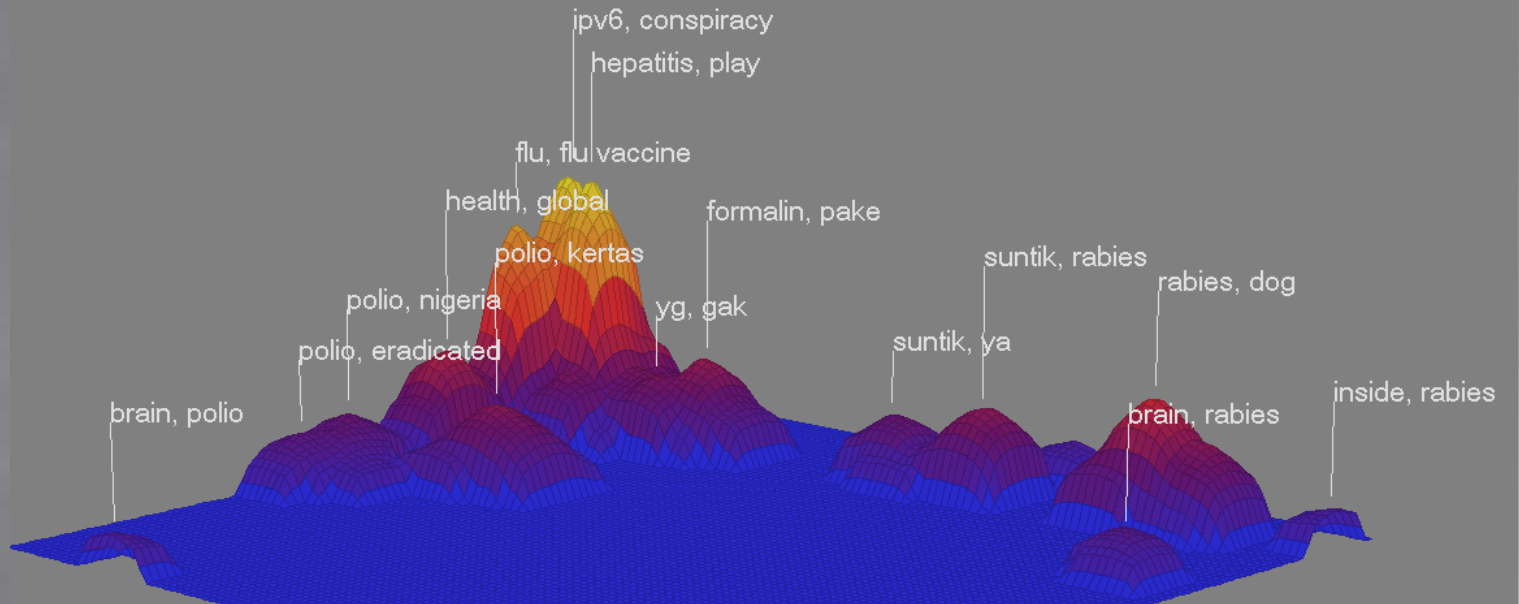
Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds

Terminology (1)

- **Theme**
  - A theme is a set of terms that typically occur within the same documents and a set of documents from which those terms have been extracted as keywords.
- **Term and Key Term**
  - A term (in IN-SPIRE) can be a key term or a major term. A key term is a keyword (which may be a phrase) that has been identified as a representative keyword for some number of documents in the corpus.
  - **On a short text corpus (such as tweets), the set of key terms is identical to the set of major terms.**
  - If the corpus is not comprised of short text documents then another method is applied to select major terms. In that case a major term is an individual word or key term that is highly topical within the set of documents.

Terminology (2)

► **Summary term**
  ■ A term that is highly associated with the current set of selected documents. This calculation is based on the similarity of the term's topical signature with the topical signature for the selected set of documents.

► **Surprising term**
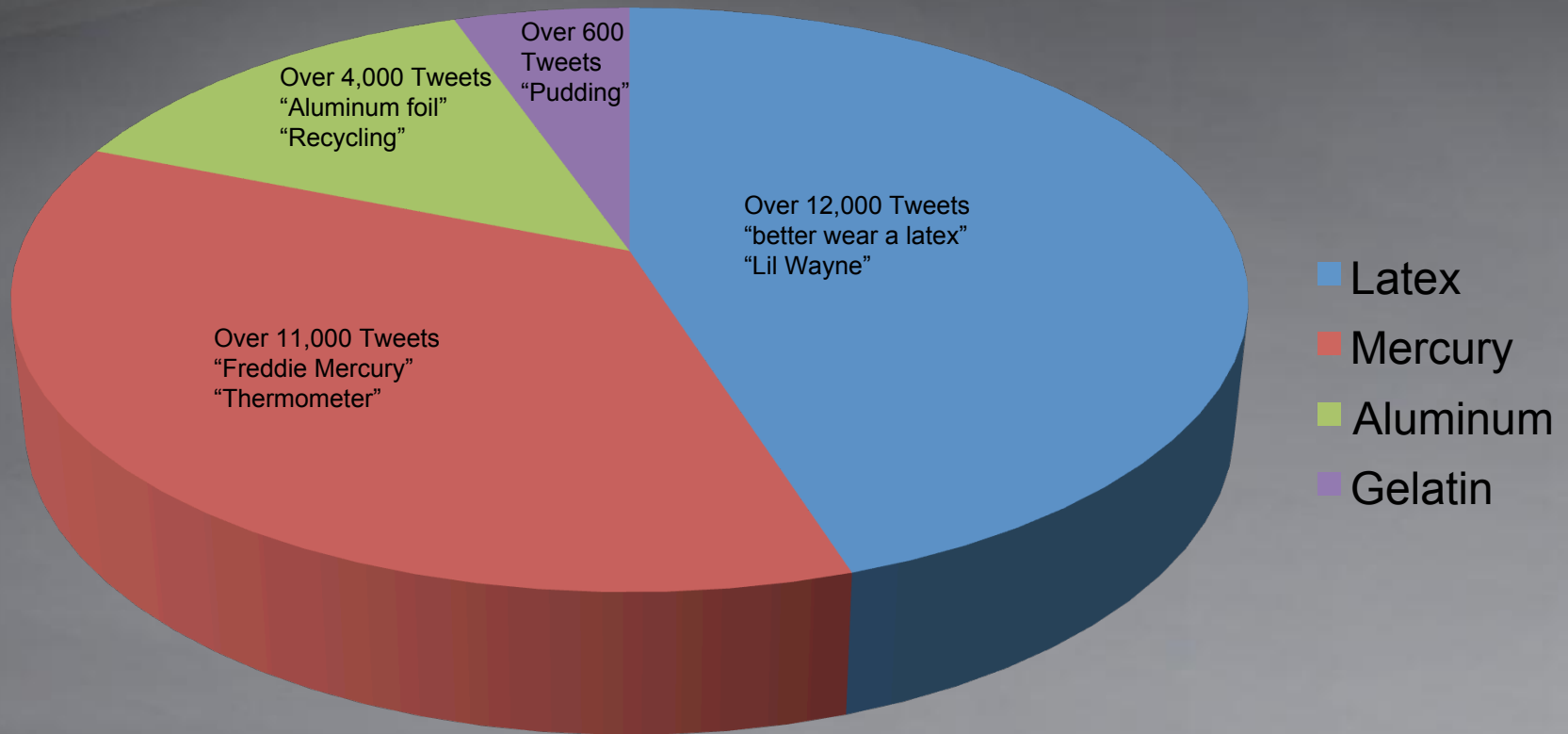 ■ A term for which a high surprise score was calculated.

► **Trending Topic**
  ■ In Twitter parlance, a word, phrase, or topic that is tagged more often than other tags. Trending topics become popular by users retweeting them. Basically these topics help the users understand what is happening in the world.
  ■ In IN-SPIRE parlance, a trending topic displays throughout the timeline and has a tendency to continue on an incline. For example, if you select a top theme and open the Time tool, you will see the selection in green and how the theme is displayed over time.

Data Summary: Search Networks

People	Questions	Materials	Places	Emotions/ Side Effects	Vaccine Types
Nurse	Who	Mercury	School	Worry	Flu
Doctor	What	Aluminum	Store	Concern	HPV
Teacher	When	Thimerosal	Mall	Relief	Pertussis
Principal	Why	RFID	Doctor Office	Safety	Smallpox
Coach	How	Formaldehyde	Restaurant	Risk	Pneumonia
Researcher	Where	Squalene	Hospital	Frustration	MMR
Health Official	Did		Neighborhood	Anxiety	Yellow Fever
Pharmacist	Whom		Daycare	Unvaccinated	CSM
State Officials	Whose			Outrage	Tdap
Mother	Which			Swollen	Malaria
Father	How come			Redness	
Pregnant Mother	Does			Arm	
Parent	How much			Side Effect	
Child	Want			Pain	
Student	Looking			Symptom	
	Find				
	Seeking				
	Searching				
	Qmark				

Types of Questions

"How"	"Is"	"Why"	"Does/Do"	Other
How do you talk about the vaccine with your kid?	Is H1N1 live vaccine safe for pregnant women?	Why don't parents not want their girls to get the HPV vaccine?	Do we want to get a Hepatitis B vaccince for our new boy?	Should parents be concerned about the 12 month mmr vaccine?
How safe is the HPV vaccine?	Is it save to take the swine flu vaccine?	Why did Rick Perry issue an executive order that forced almost every single girl in Texas to receive the deadly Gardasil vaccine?	Does a decision not to vaccinate against HPV put other people at risk?	To vaccinate or not to vaccinate?
How does a baby get whooping cough when its mandatory to take the vaccination in the hospital?	Is the HPV vaccine safe?	Why some parents choose not to vaccinate?	Does mercury in vaccines cause autism in children?	What happens when your allergic to latex?

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds

Pacific Northwest
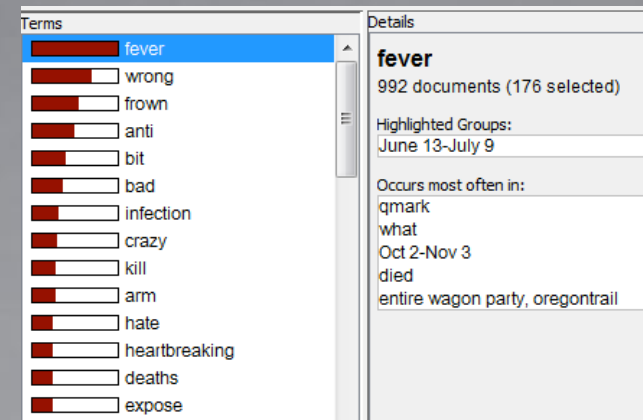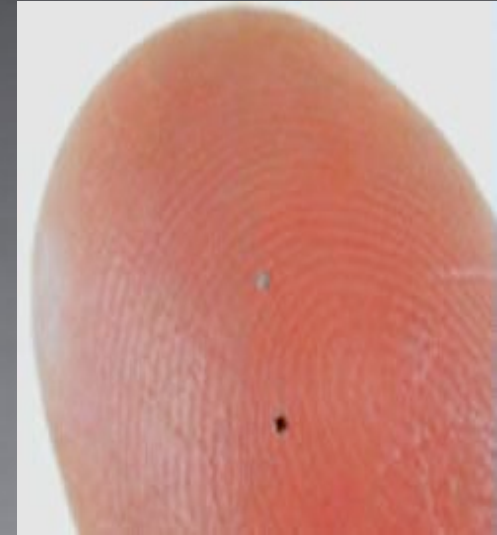NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

"Surprising" Vaccine Safety Tweets

► **Surprising tweets** are found utilizing a novel algorithm that leverages word co-occurrences, temporal trends, and other signatures

► Questions:
  - How do you all feel about the hpv vaccine? i was offered it today at the doctor but i'm unsure... #pleaserespond
  - Michele Bachmann says she 'heard of a woman who got mental retardation from the hpv vaccine.' it was her, right?
  - Cut my finger on a can. do i need a tetanus shot?
  - Will you give your son the hpv vaccine?
  - do you recommend any vaccinations ? /yes, but i think that the schedule needs to individualized and some vax aren't needed

► There is accurate information out there:
  - Why don't we confront those parents who refuse to vaccinate their kids? http://t.co/hiX3NHTr
  - Pregnant women and influenza http://t.co/dhnl6RQZ [CDC Office of Women's Health]

"Surprising" Vaccine Safety Tweets

But there is a tremendous volume of misinformation:

► Swine flu vaccine recipients could be tracked with RFID bracelets using Big Brother medical tech http://t.co/vCRF7s9M

► #Girls who have #seizures ,#vomitting, and no quality of life due to #Gardasil need help truthaboutgardasil.org

► What the vaccine industry won't admit about the dangers of vaccination http://t.co/K1mhmG8l

► Vaccination? "more recent medical school graduates are 'more likely to believe immunizations do more harm than good" http://t.co/n9PcMEAM

► Scientists hyping 'universal' flu shot to perpetuate vaccine scam http://bit.ly/n3xLaH

Terms

	fever
	wrong
	frown
	anti
	bit
	bad
	infection
	crazy
	kill
	arm
	hate
	heartbreaking
	deaths
	expose

Details

fever
992 documents (176 selected)

Highlighted Groups:
June 13-July 9

Occurs most often in:
qmark
what
Oct 2-Nov 3
died
entire wagon party, oregontrail

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds

Surprising Tweets: Global Perspective

► That is why i keep asking the health authority in kuwait to add the chickenpox vaccine to routine vaccination schedule in Kuwait

► Infectious diseases like #measles & #cholera threaten #hornofafrica refugee camps. How ngos & govts can help? http://bit.ly/pRCcfY

► Nsanje children who fled into Mozambique during the anti-measles vaccine a few month ago were vaccinated at gunpoint

► 2 measles cases in Minnesota babies linked to Kenya http://t.co/lY725Dd
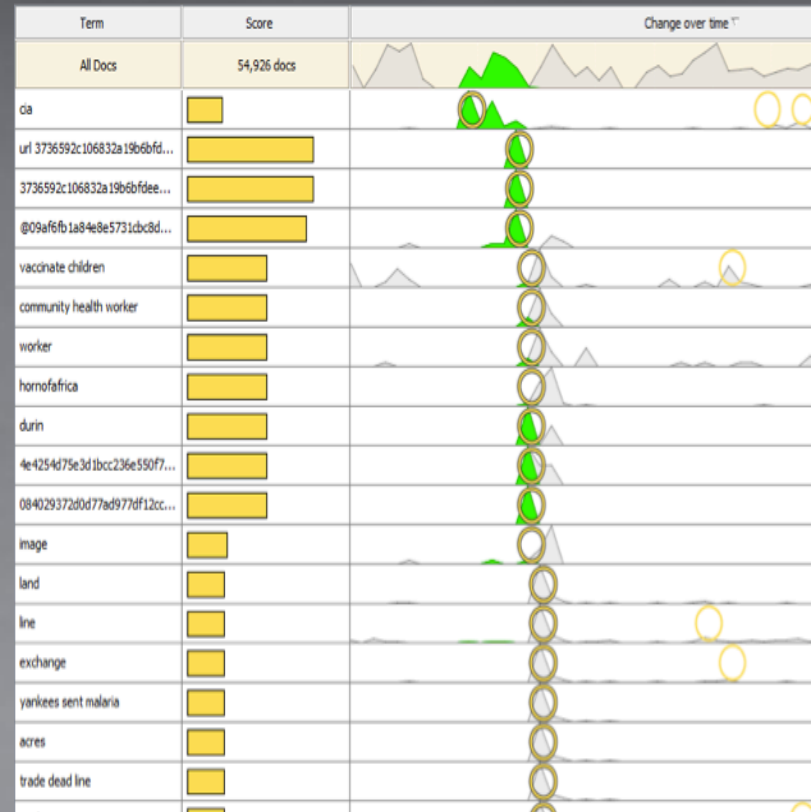
Scalable Reasoning System Overview

▶ Software framework for developing web-based analytic applications

▶ Connects to any data sources

▶ Analyzes and models data to produce visualizations

▶ Enables custom applications to be assembled and tailored for specific uses

Comma Separated Values

Social Media

Patents

Lessons Learned

7:45 AM

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds

Social Media Visualization

Twitter feed showing flu related tweets in the DC area

Term Search

Feature to Visualize

Add Filters

Topics

Feed Selection

New Indicator

Timeline

Raw Tweet

Faceted Browsing

June 28, 2012

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds

Social Media Visualization

Twitter feed showing flu related tweets in the DC area with specific mention selected, showing brushing and linking

June 28, 2012

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds
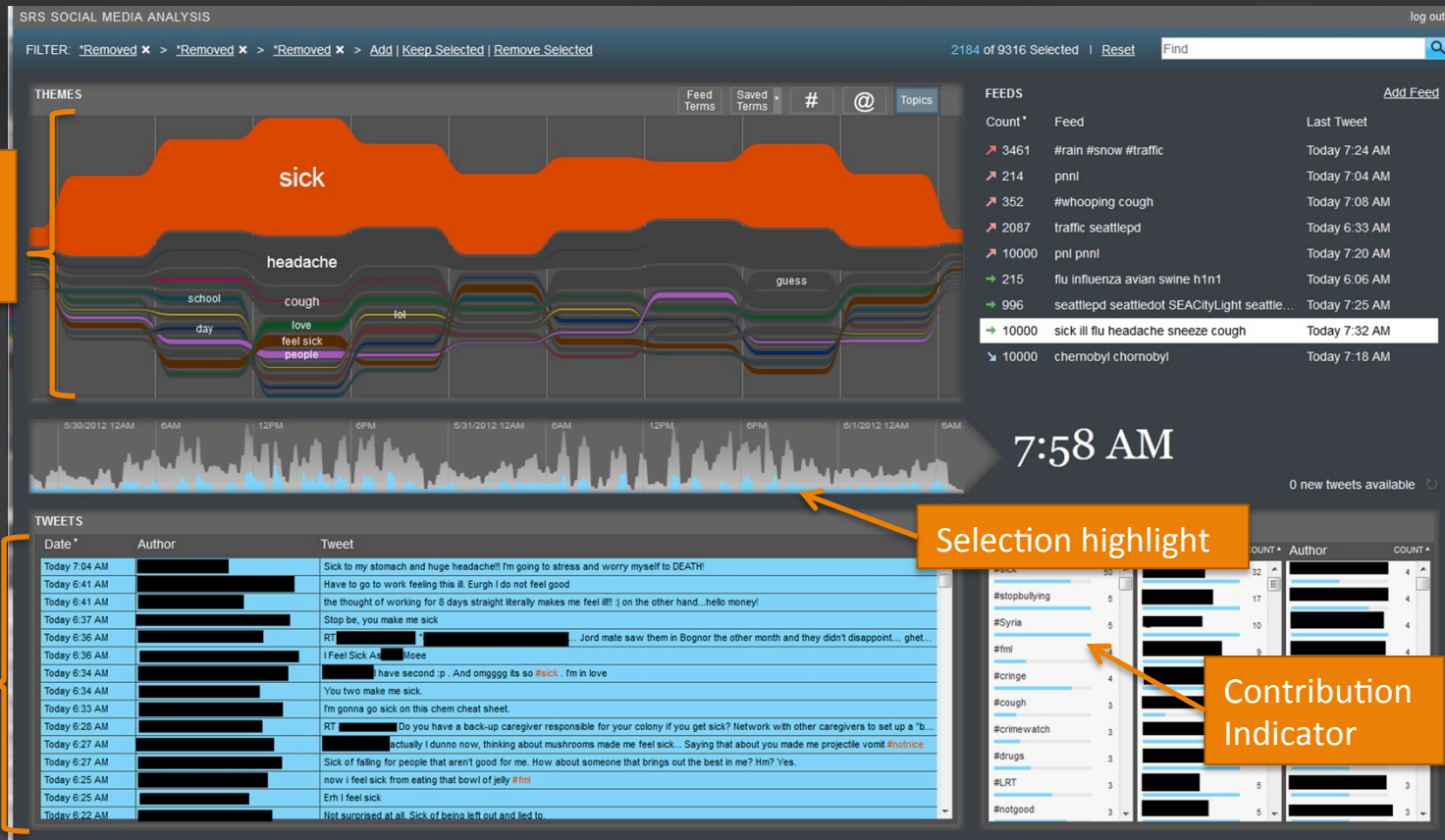
Social Media Visualization

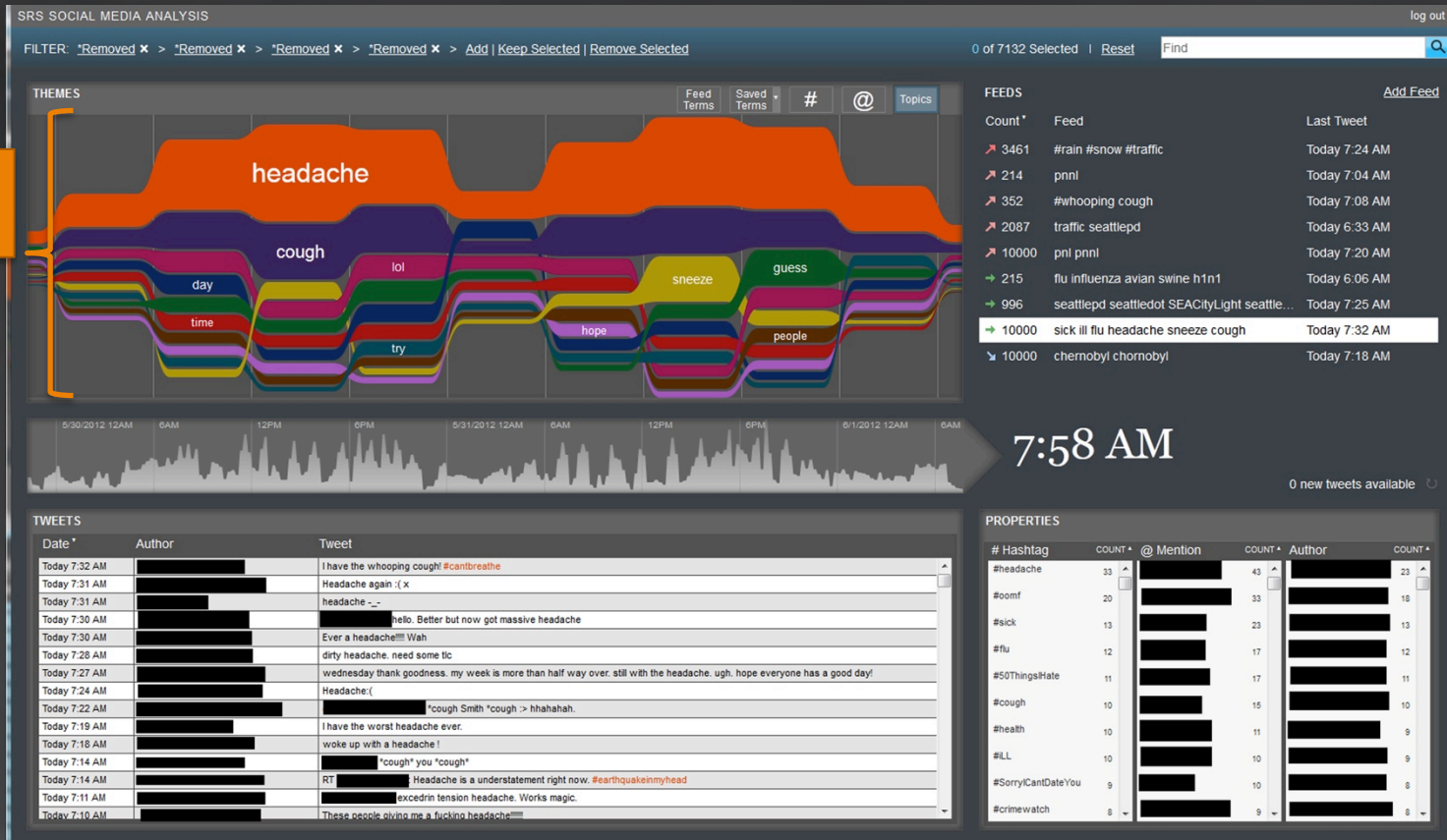Twitter feed showing flu related tweets in the DC area with specific mention selected, showing brushing and linking

"Sick" Removed

June 28, 2012

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds

Focus Visualization

1. Edit
"Saved Terms"

Edit Terms

Terms (#hashtags, keywords and @mentions)

influenza, flu, adenovirus, h1n1, h3n2, h5n1, ah1n1, swine, bird, avian

Separate terms with a space. For example, "icy roads" #traffic @NASA accident

Authors

Separate authors with a space.

Load   Cancel

THEMES

Feed Terms | Saved Terms | # | @ | Topics

flu

influenza

avian

bird

swine

2. See how all data maps to focus terms

Made Possible By

- Courtney D Corley – Senior Research Scientist, Health Security, PNNL
- Stuart J Rose – Senior Research Engineer, Visual Analytics, PNNL (not shown)
- Martin G Myers – Professor Emeritus of Pediatrics and Preventive Medicine at the University of Texas Medical Branch in Galveston Texas.
- Shy Brown – Biological Sciences PhD Student, Washington State University
- Oriana Love – Senior User Experience Scientist, Visual Analytics, PNNL (not shown)

► Funded by:
- Pacific Northwest National Laboratory, Laboratory Directed Research and Development
- Immunizations 4 Public Health
- Infectious Diseases Society of America
- Department of Homeland Security

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Information is the fundamental currency of the New Biology.

    - Committee on a New Biology for the 21st Century

Courtney D Corley, PhD
National Security Directorate
Pacific Northwest National Laboratory
court@pnnl.gov ◆ @CourtCorley

http://signatures.pnnl.gov/

Panel: Twitter for Disease Surveillance
International Society for Disease Surveillance
#tweet4ds