Social Media Data as a Public Data Resource
at APDU 2013 (9/16/2013)

Mark Silverberg      @skram

SOCIAL HEALTH INSIGHTS LLC
About Social Health Insights

• Developers, clinicians, and designers
• Passionate about using IT & data to improve health

About MappyHealth

• Built in response to the HHS/ASPR #NowTrending2012 Challenge which we won
• Utilizes the Twitter streaming API to look at disease & disaster related conversations
• Continued research into how Twitter data could help inform disease and disaster surveillance
• Exploring concepts that could be applied broader than just Twitter and health
Social Media’s Pervasiveness

• “As of May 2013, almost three quarters (72%) of online U.S. adults use social networking sites”
  – up from 67% in late 2012
  – up from 8% when Pew started asking in 2005

• All age groups are seeing growth
Why Twitter as a Data Source?

• 500 million users - “well over 200 million active”
  – 400 million tweets per day
  – Source: Twitter Engineering Blog, 3/21/2013

• All age groups seem to be adopting Twitter at rates similar to other social networking sites

• Twitter users are increasingly mobile (60% use a mobile client according to The Guardian 12/18/2012)

• Twitter has lax privacy defaults compared to FB
You might be surprised what people will talk about

319,079 Tweets collected from Twitter on 02/05/2013 that match terms tracked on MappyHealth.com
Questions and Challenges When Working with Social Data

- What problem do we want to solve?
- How will users of the platform use the information provided? What is implied and what should we explain?
- What sources will be utilized to contribute towards the insight sought?
- How does one visualize the data?
- What semantics will be meaningful and useful?
- What analytical intervals are pertinent?
- Do we need geo-located data or more generalized?
- How do we work around non-geo-located data points?
@WHO: So far, WHO is aware of 14 confirmed #H7N9 cases, with 6 deaths. Awaiting confirmation of 2 more cases. #influenza #flu
Are All Tweets Created Equal?

Total Estimated Impressions = Max Followers (during collected timeframe) x Number of Tweets. For Timeframe from 03/31/2013 6 am to 4/8/2013 12 pm GMT
Total Estimated Impressions = 1,996,125,456 - Data captured using the Twitter streaming. Visual crafted by Social Health Insights LLC – H7N9 Tweets Captured
Same Deal for Disasters
Example Earthquake Activity

![Hourly Volume of Tweets about Natural Disaster](image)

- **Mexico Earthquake**
- **Nepal Earthquake**
Social Data Can Compliment Traditional Data

Flu terms, regional view

With off-line WIP geo-location algorithms applied
Influenza Vaccination – Opportunity?

Data captured using the Twitter streaming API by MappyHealth, a web platform by Social Health Insights
## Ongoing Challenges

<table>
<thead>
<tr>
<th>Cough, Running Nose and Fever. Can you go away? :</th>
<th>(</th>
<th>Vs.</th>
<th>I have a Bieber Fever ♥</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 3% geotagged</td>
<td>Vs.</td>
<td>97 – 99%</td>
<td></td>
</tr>
<tr>
<td>Washington, DC</td>
<td>Vs.</td>
<td>I am on Earth!</td>
<td></td>
</tr>
</tbody>
</table>

- Access limitations to full data streams ( $$$, rate limits, connectivity )
- Character limitation for message
  - Linguistics and slang (can be automated, to an extent)
- Knowing what to look for
Location-driven listening: Knowing where and what people are talking about
We Use Open & Public Data Too

CheckQM
a Social Health Insights venture

Focus on implementation while saving time and money

Taking in complicated HQMF (XML) files from HHS-AHRQ, NIH, and CMS to help organizations understand, collaborate on, and achieve Meaningful Use CQMs

Eldercare.gov has a website and hotline staffed 9-5 but we built a SMS/phone automated attendant that anyone can use anytime
More Open Data “Mashups”

Open source hospital compare using data from several agencies – using “big data” technology

Blog: Visualizing HIPAA Data Breaches

Blog: EHR Vendors being Used for MU