

Tech Talk

Make Public Perception Data Work for You: Collecting and Analyzing Social Media Data

SEPTEMBER 29, 2022

Annice Kim, PhD, MPH

Robert Chew, MS

Jamie Guillory, PhD, MS, MBA





Webinar Overview

- Social media data overview and what research questions we can answer with these data
- Data science methods to gain insights on public health audience segments
- Recommendations for incorporating social media monitoring into your existing surveillance process

The New York Times

***‘Very Harmful’ Lack of Data
Blunts U.S. Response to Outbreaks***

HealthAffairs

The Fierce Urgency Of Now:
Closing Glaring Gaps In US
Surveillance Data On COVID-19

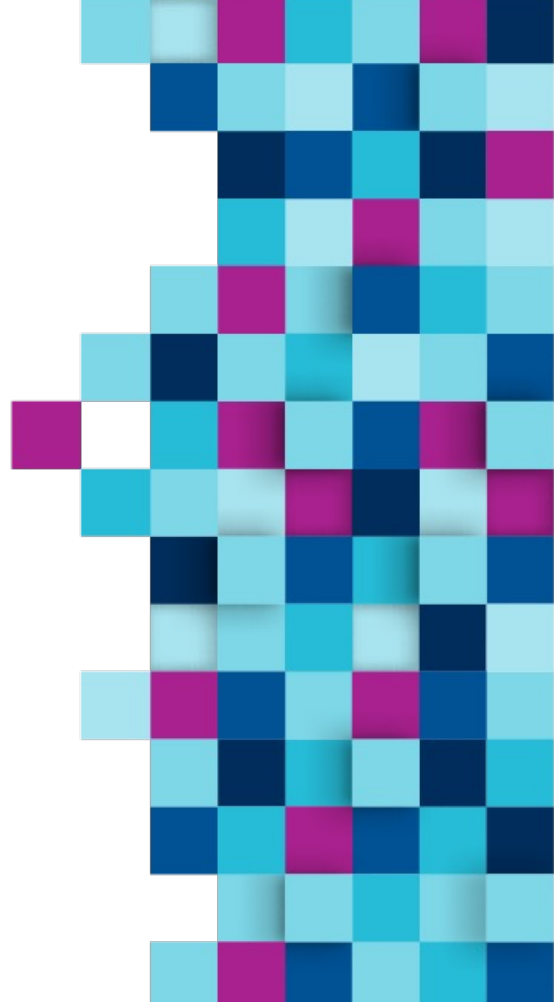




Social Media



Social Media Listening



Social Media Listening

“ *Social listening is the process of collecting data from social platforms and forums on a chosen topic...the collected data is then analyzed to find trends and useful insights.* ”

[Brandwatch](#)





Why Health Agencies Use Social Media Data



Communication



Surveillance



Policies/Regulations



What Health Agencies Want to Know



Communication

- Gauge public response to health education campaigns.
- Identify and address misinformation.
- Identify influencers who will amplify messages and detractors who undermine messages.



What Health Agencies Want to Know



Surveillance

- Monitor emerging products and health issues.
- Identify emerging consumer perceptions and misperceptions about health.
- Monitor emerging consumer health behaviors.



What Health Agencies Want to Know



Policies/Regulations

- Understand public support/opposition for policies.
- Identify potential impact of regulatory policies and unintended consequences.
- Monitor marketplace for emerging products and misleading claims.
- Monitor compliance and enforcement of regulations.



Social Media Listening – Example Research Questions

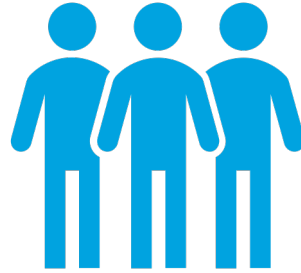
- How are brands marketing their new synthetic nicotine products? What are consumers' perceptions and reactions to these products?
- What are consumers' perceptions and anticipated behavioral responses to the upcoming ban on menthol cigarettes?
- How are consumers using vaping products as a cigarette smoking cessation tool?
- What are North Carolina residents discussing on social media regarding travel and travel safety amid COVID-19?
- What questions do consumers have about staying safe during the COVID-19 pandemic?



Benefits/Opportunities



Near Real-Time Data
Collection



Massive Volume



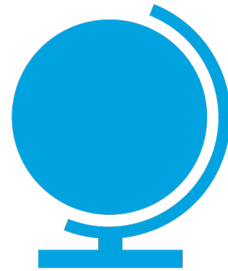
Users' Own Voices



Limitations



Representation



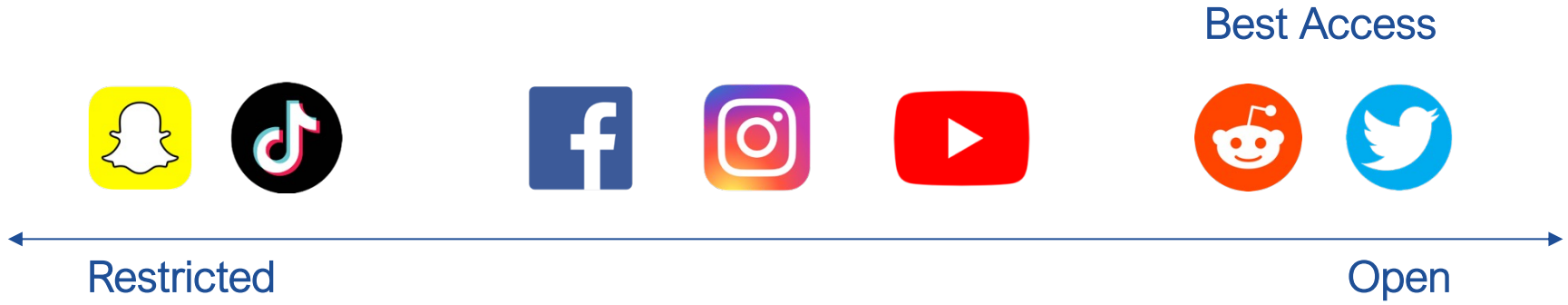
Some Data Limited



Access



Social Media Platforms & Data Access





Social Media Listening Process

Set-up



1

Identify research question, time frame & platforms for analysis.

Listen



2

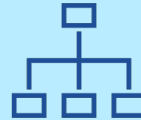
Develop query to identify conversations on social media & forums.



3

Review results & refine query to remove off-topic posts.

Analyze



4

Categorize posts into key themes.



5

Qualitatively code posts & summarize results.

Report



6

Report volume, key themes, emerging topics & illustrative posts.



Key Metrics

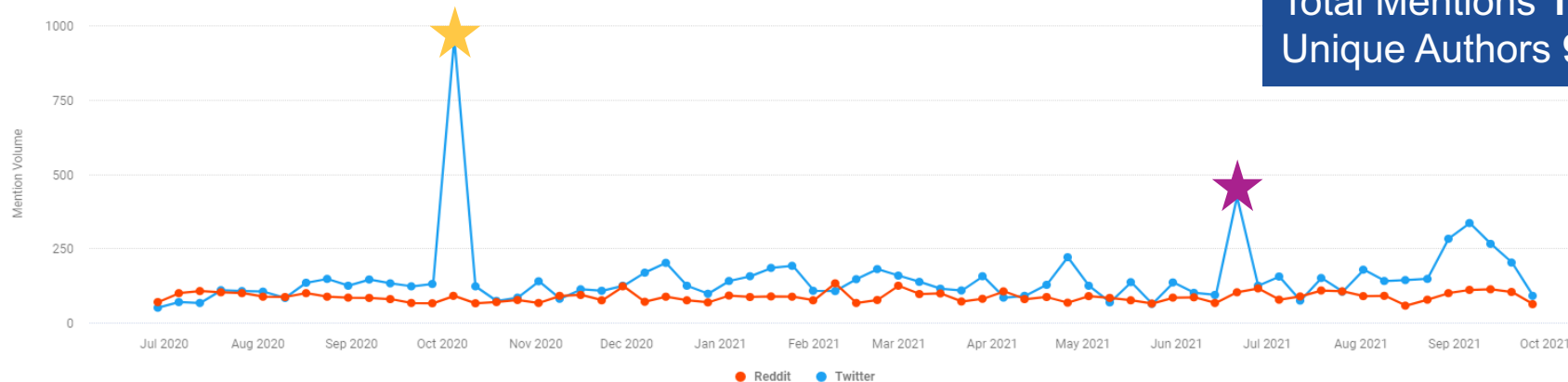
- Volume of conversations about your issue
- Key spikes in volume trends and what is driving these peaks
- Top topics related to your issue
- Emerging keywords/themes
- Illustrative sample posts

Surveillance of Emerging Products

Case Study: Synthetic Nicotine – Conversation Volume



Total Mentions **15.5K**
Unique Authors **9.4K**



October 7, 2020: Twitter volume peaked in early October 2020 when Kenya's health minister banned LYFT synthetic nicotine pouches.



If I owned an e-liquid company I'd certainly be looking at synthetic nicotine right now. But I wouldn't announce it on social media or dare FDA to stop me, as I've seen several do.

Vaping360 @Vaping360 · Sep 17
The Promise and Peril of Synthetic Nicotine as a PMTA Loophole.
By Jim McDonald @whycherrywhy
ow.ly/edwB50G5wGQ

8:11 AM · Sep 17, 2021 · Twitter for iPhone

September 6, 2021: Increase around September 9 PMTA deadline for brands to submit application to stay on the market under FDA regulation.







Surveillance of Emerging Products

Case Study: Synthetic Nicotine – Top Topics



TOPIC NAME	MENTIONS
synthetic nicotine	1957
tag us in your story	259
synthetic e-fuels	122
quit smoking	352
derived from tobacco	149
long cut	211
pure nicotine	99
zyn pouches	1071
tobacco products	218
Zyn nicotine pouches	332

Top phrases centered around the following themes:

-  **Synthetic Nicotine Products/Brands**
-  **Tobacco/Smoking Cessation**
-  **Promotional Posts**
-  **Comparisons with Tobacco**

Surveillance of Emerging Products



Key Takeaway: Consumers discuss using synthetic nicotine pouches and other similar products to quit smoking and/or vaping.

- User is asking for recommendations on how to quit smoking, and another user replies “zyn tobacco free nicotine pouches.”
- User attributes success quitting smoking to tobacco free nicotine pouches.
- User says tobacco free nicotine pouches are similar to dip pouches. They also describe them as helpful with managing nicotine withdrawal.



Posted by u/Emergency_Fortune718 3 months ago

How did you quit smoking?



OGButtPirate53 3 months ago

Zyn tobacco free nicotine pouches



pixiemeanow 1 year ago

What really helped me quit was tobacco free nicotine pouches, without them I would've never been able to quit lol



John_from_HR · 5d

They make tobacco free, nicotine pouches that are similar to dip pouches. Helps with the withdrawal side as well. Available in most gas stations

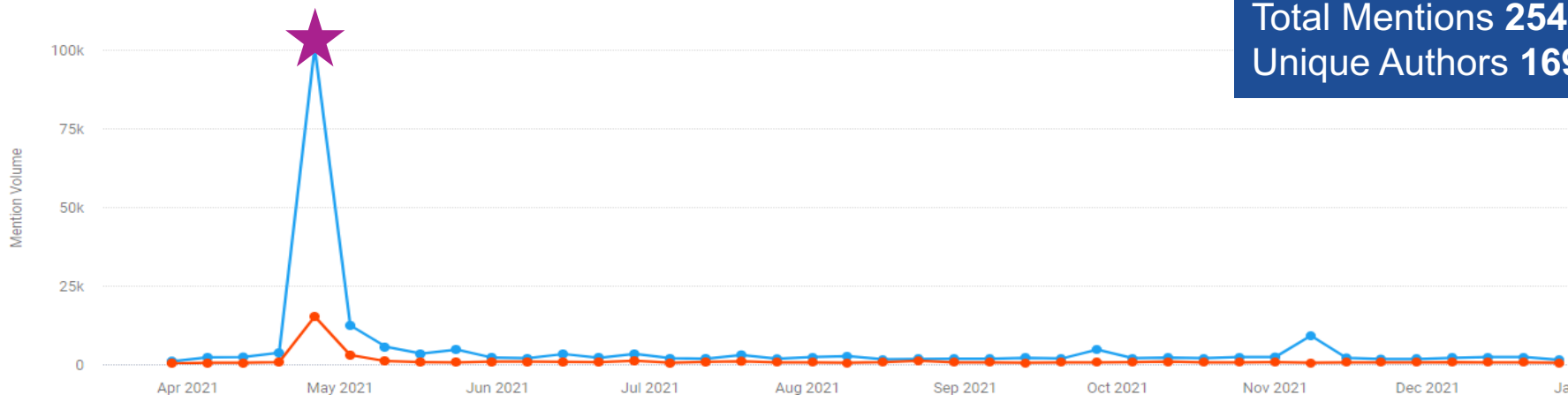


Monitoring Potential Impact of Policy

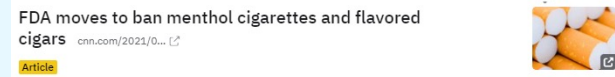
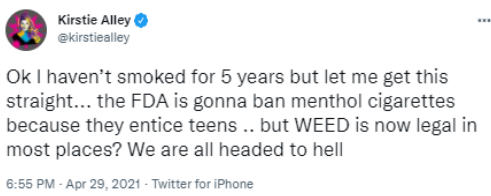


Case Study: Federal Menthol Ban – Conversation Volume

Total Mentions **254K**
Unique Authors **169K**



April 26, 2021: Twitter and Reddit volume peaked because of the FDA announcement to ban menthol cigarettes and flavored cigars. Volume peaked around 115K total mentions.





Monitoring Potential Impact of Policy



Case Study: Federal Menthol Ban – Top Topics

TOPIC NAME	MENTIONS
banning menthols	8627
black people	8421
gonna ban menthol cigarettes	7240
banning menthol cigarettes	6687
FDA is gonna	5763
Biden administration	5721
menthol cigs	4843
ban menthols	4810
pack of Newports	4571
ban on menthol cigarettes	4247

Top phrases centered around the following themes:

-  **Race/Ethnicity**
-  **Menthol Cigarette/Flavored Cigar Ban**
-  **Menthol/Flavored Tobacco**
-  **Government**

Monitoring Potential Impact of Policy



Key Theme: Consumers believe the ban will increase policing among Black people and cause harm in Black communities.

- Consumer suggests the menthol ban undermines Black people's autonomy and will result in criminalizing more Black people.
- User critiques decision to make menthol cigarettes illegal. Suggests punishing tobacco companies or providing universal healthcare would better address the issue.
- User interprets menthol ban as banning a flavor of cigarettes because Black people like it.



The Biden admin seems to think that **black people** are **children to be managed** and coddled because they're **unable to make decisions** for themselves like smoking **menthol cigarettes**. All this will do is **criminalize more black people**.

8:47 PM · Apr 29, 2021 · Twitter for iPhone



“Smoking menthols kills Black people so instead of punishing the **Tobacco companies** for pushing this poison, or enacting a **universal healthcare system** that will save lives, we will simply make them illegal 😊”

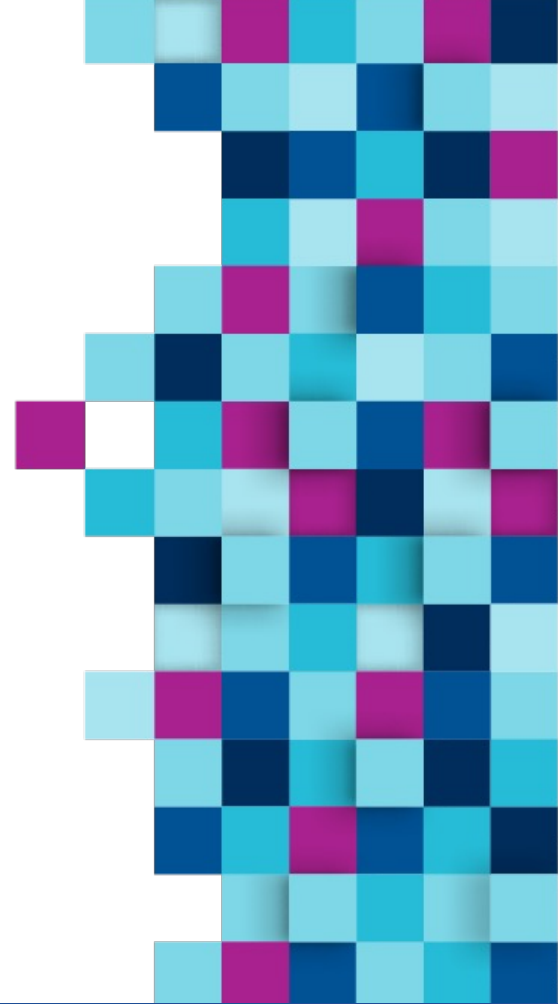
6:25 PM · Apr 29, 2021 · Twitter for iPhone



saying **you're banning menthols** because they **disproportionately impact the African-American community** is a wild way of saying **you're banning a flavor of cigarettes** because **black people like it**

6:25 PM · Apr 28, 2021 · Twitter Web App

Data Science Methods to Gain Insights on Public Health Audiences



Audience Segmentation

- Social media listening platforms can cover many relevant use cases.
- However, additional user information is often important for public health surveillance and evaluation.



Who
is posting on
social media?



Who
is being
exposed
to this
content?

How would you approach this problem?



Traditional Approach: Qualitative (Deductive) Coding

Steps

- Determine categories/codes of interest.
- Take a sample of users.
- Collect and read information from each user's profile, posts, and comments.
- Assign codes to each user, based on best professional judgement.

Strengths

- Conceptually, easy to understand and explain

Limitations

- Manually taxing and challenging to scale/repeat



Alternative Approach: Supervised Machine Learning

Steps

- Create labeled data set, following same steps from deductive coding.
- Split data into training and test sets.
- Develop model to predict categories of interest.
- Evaluate trained model's performance.

Strengths

- Easy to scale to users outside your coded sample and to repeat periodically over time

Limitations

- More challenging to develop and diagnose

Supervised Learning for Audience Segmentation



- Developed classification models to assign users into predicted age groups.
 - [Twitter](#) (Morgan-Lopez et al., 2017)
 - [Reddit](#) (Chew et al., 2021)



- Developed classification models to assign “User Types” to accounts.
 - Marketer, Individual, News Media, Public Health Agencies, Advocates, Bots
 - [Twitter](#) (Kim et al., 2017)

Reddit Age Category Model: Overview

- Manually labeled age groups by searching for user age self-reports



Features

- Derived from metadata and up to last 100 posts and comments

Modeling

- Compared several supervised learning algorithms
 - Logistic Regression, KNN, SVMs, RF, Gradient Boosted Trees
- Conducted feature selection to reduce from ~1500 to ~15 variables.
- Reported 5-fold cross-validated metrics, on training set, as well as test set performance.

Variable group	Metadata used	Example	# variables
Summary Statistics	All	Median post score	189
Subreddit Frequencies	Posts and Comments	Frequency user posted to “Teenagers” subreddit	624
Emoji Frequencies	Comments	Frequency of “🙄” used by user	101
Literary Characteristics	Posts and Comments	Average Flesch Reading Ease score	28
Patterns in Posting	Posts and Comments	Percentage of user’s posts that were videos	42
Term Usage	Comments	TF-IDF score for the term “school”	539
TOTAL			1,523



Reddit Age Category Model: Variable Interpretation



Adolescents (13–20)

- Created account more recently
- More likely to post in popular subreddits
- More likely to post to r/teenager
- More likely to use the term “school”



Adults (21–54)

- Wrote more sentences per comment
- Higher Coleman Liau Index (readability)
- More likely to post in r/news
- More likely to use the term “home”

Chew et al., 2021

Who is posting on social media?

Use Case 1: Cigarette Conversations Among Predicted Youth vs. Adults



Are there differences in themes about cigarettes between youths and adults on Reddit and Twitter?

- Identified conversations about smoking.
- Applied age prediction algorithms to 17,707 Twitter and 23,364 Reddit users posting about cigarettes.
 - Guillory et al., 2022
- Conducted SML on final samples:
 - 5,598 youth and 6,347 adults (Twitter)
 - 5,417 youth and 6,131 adults (Reddit)



Adults discuss smoking cessation more than youth on both Reddit and Twitter.



I quit smoking recently and writing is one of my triggers. Days aren't getting better, and the nights are getting longer. What a pathetic excuse of a hole I dug myself in.

10 ↓ Reply Share Report Save Follow



Never start smoking.

The cravings never go away.

7:23 PM - Mar 17, 2021 - Twitter for Android

14 likes



Youth on Reddit discuss health consequences of smoking; Adults do not.



Before i even started. Who wants yellow teeth, bad smell and cancer.. and the pros: that's all. 🤔 be cool, don't smoke 🤔

Who is being exposed to this content?

Use Case 2: Estimated Ages of Juul Twitter Followers



- In the late 2010s, JUUL was the most popular ENDS in US, accounting for nearly 75% of sales.
- Among high school students who had ever used e-cigarettes and had tried 1 of 4 leading brands, JUUL was the top brand used (64.2%).
- Social media and JUUL's advertising strategy of using young adult models and influencers may have influenced JUUL's popularity.
- JUUL claimed it did not target youth and voluntarily initiated plan to address youth access, appeal, and use of JUUL products.

Who is being exposed to this content?

Use Case 2: Estimated Ages of Juul Twitter Followers



Are kids being exposed to social media posts by JUUL?

- Examined 11,861 accounts following JUUL's Twitter account in April 2018.
 - Kim et al., 2019
- Predicted age of JUUL followers based on metadata and sample of tweets.
 - ~80% < 21 years old
 - ~45% < 18 years old

The New York Times

Juul Settles Multistate Youth Vaping Inquiry for \$438.5 Million

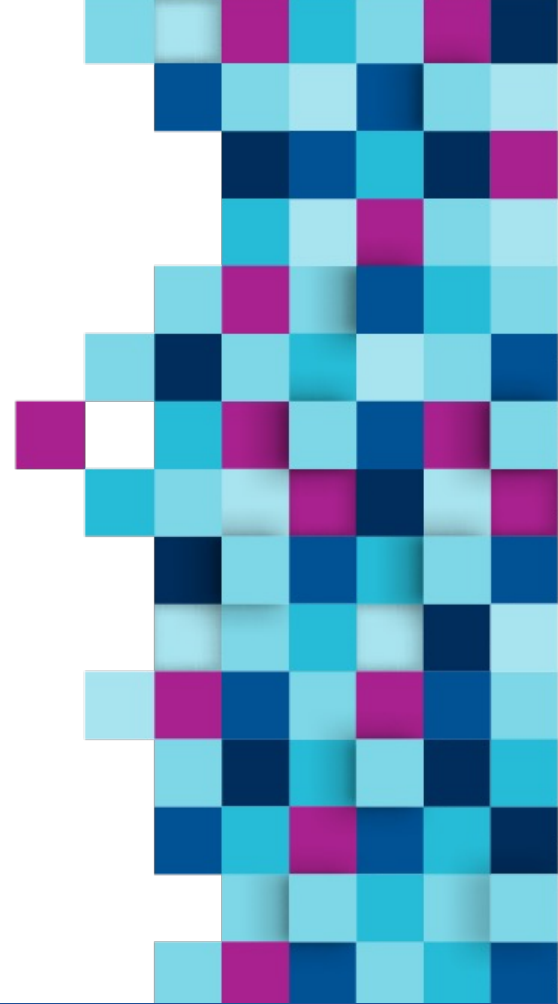
The tentative deal would close an investigation by nearly three dozen states into the company's marketing and sales practices.

Give this article 314



An investigation found that the company had a “porous” age verification system for its products and that 45 percent of its Twitter followers were ages 13 to 17. Bryan Anselm for The New York Times

Recommendations for Incorporating Social Media Monitoring in Your Work





Do you need to monitor social media?

- Do you need continuous surveillance on a health topic?
- Are you monitoring an emerging issue that is rapidly evolving?
- Do you need early indicators to help inform other research or response efforts?
- Is there currently no gold standard method for capturing these data?
- Is your target audience of interest on social media?



Social Media Monitoring is an Iterative Process





Clarify Research Questions:

Know what you can / cannot answer with social media data

- **Very specific questions:**

- “Did our campaign influence vaccine hesitant adults to change their attitudes and get vaccinated?”

- **Instead, ask these types of questions:**

- What are emerging trends in health topic x?
- How is the public responding to policy x?
- What questions do they have about x?

- **Difficult to answer because:**

- Specific audience
- Behavioral outcomes
- Causal changes

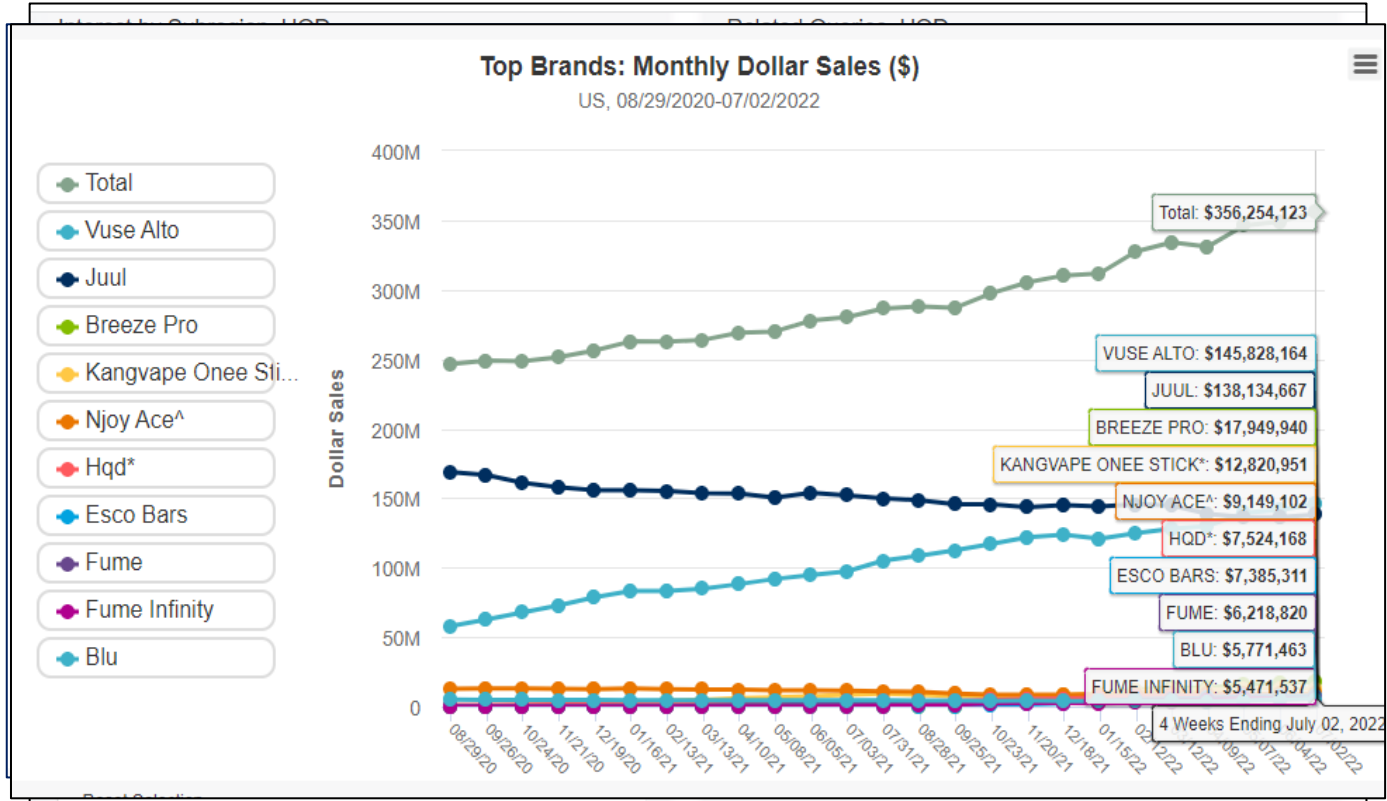
- **Easier to answer because:**

- Open-ended, exploratory
- Identifying emerging issues and trends
- Early indicators/signals that can inform further research and/or triangulation with other data

Triangulate Social Media Insights with other Data Sources

ENDSTRACKER

- Home
- Advertising
- Sales
- Google
- Twitter
- Instagram
- Reddit
- Methodology
- Logout

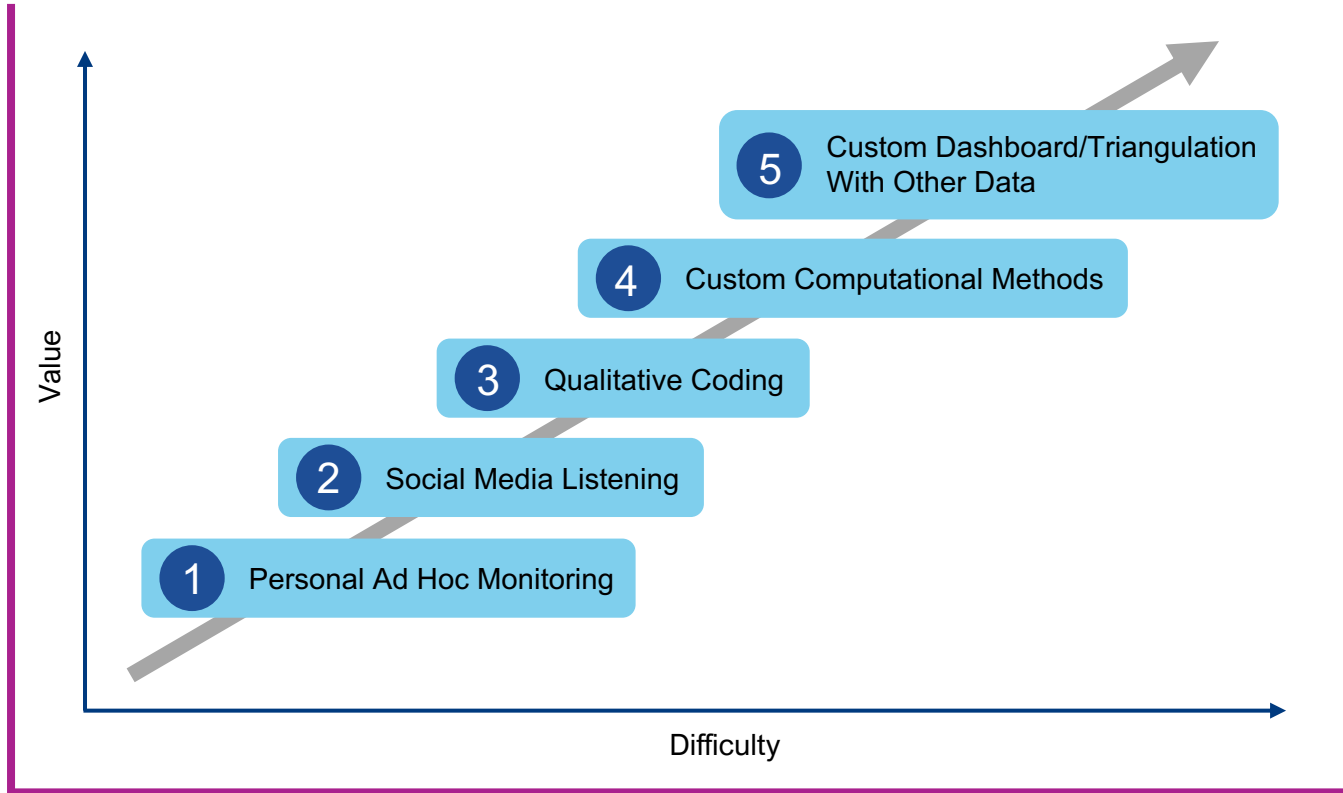


Evolving Social Media Ecosystem – Where is the signal?

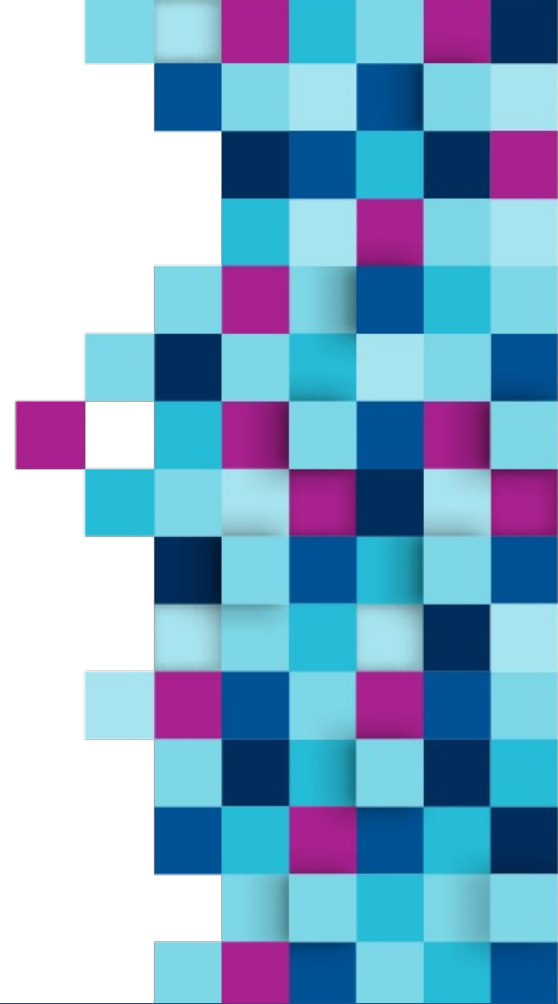
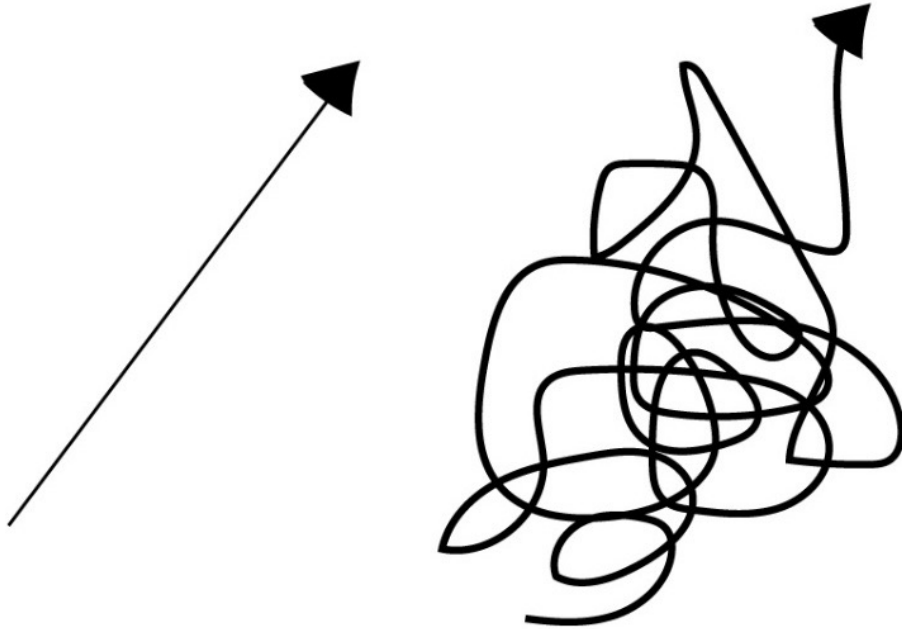
- What data are accessible?
- How will you get these data?
- What is your budget and resources?



Social Media Analysis Maturity



expectation **reality**



Need for Paradigm Shift



What we value in science

- Rigor
- Validity
- Control, experimentation, manipulation
- Generalizability

What we need

- Adaptive, nimble, lean research
- Timeliness
- End user mindset
- Faster dissemination process



Balance Tradeoffs

Signal Detection
Timeliness
Scalability Frequency

Rigor
Deep Dive
Customization

Inform communication
Inform other data collection
Inform prevention development

Publication
Evaluate outcome
Enforce policies/regulations



Outputs

Knowledge of social media data is essential for public health

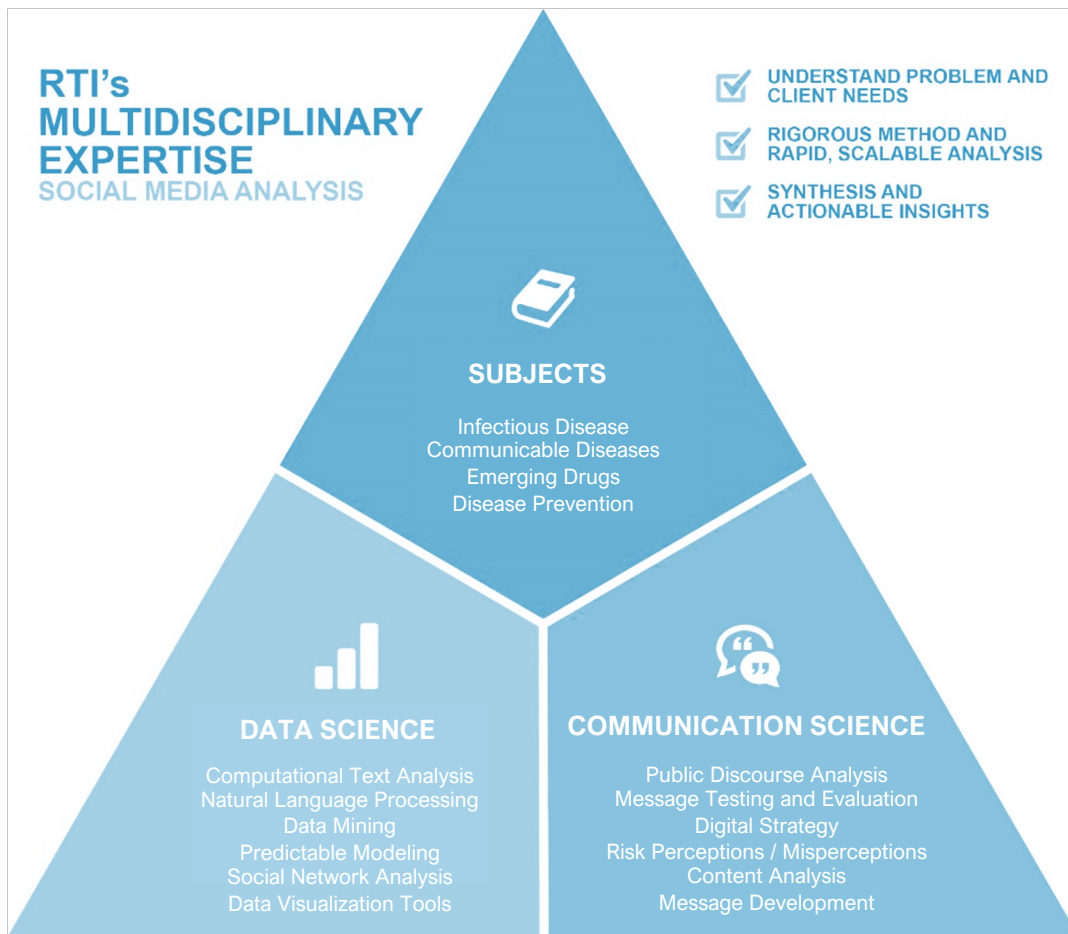


Continually upskill
and bring others
along on the journey





Work with Us!



Tech Talk

Thank you

Annice Kim | akim@rti.org

Robert Chew | rchew@rti.org

Jamie Guillory | jamieguillory.contractor@rti.org

