How to Collect Data from Social Media

Erik Bean, Ed.D.
Associate University Research Chair
Center for Leadership Studies and Organizational Research

LauraAnn Migliore-Nolan, Ph.D.
Doctoral Instructor - Doctor of Management in Organizational Leadership Program – DBA
University of Phoenix - College of Doctoral Studies

Thursday, September 24th, 7 to 8 p.m. EDT, Blackboard Collaborate

• This workshop provides an overview about issues and strategies for collecting data from social media, Bean and Migliore share their insights about effective practices for using social media as data collection sources. In addition, challenges and issues related to using social media as data collection source will be discussed.

• Audience: All interested researchers including faculty, staff, students, and alumni.
Agenda

• Dr. Papcharissi, one of the world’s leading social media data collection and analysis experts
• Dr. Papcharissi Seminal Twitter Groundbreaking Research
• Social Media Data Collection: Best Practices
  -- Public Versus Private Social Media Data
  -- Ethics
• Host Dr. Erik Bean, Content Analysis Examples, Software
• Erik Sample Web Published Study (Web Operational Example)
• Recommended Academic Social Media Analysis Software
• Social Media, You, and Big Data
• Dr. LauraAnn Migliore, Data Analysis Expertise
Zizi Papacharissi, PhD

• One of Several Foremost Leading Social Media Experts.

• Department Head and Professor of Communications
  University of Illinois

• Her Co-Authored Seminal Paper on Studying the 2011 Arab Spring Uprise Using Twitter Was a Groundbreaking Example of Social Media & Data Collection

• Published 9 Books, 70 articles, and serves on the editorial board of 15 peer-reviewed journals

• Founder and current editor of the open access journal Social Media and Society, Sage.
“We conducted a high-level quantitative analysis of approximately one million tweets broadcasted through #egypt during the month-long period in 2011 that led to and followed the resignation of Honsi Mubarak to locate prominent actors. We followed up with a qualitative analysis of significant frames that emanated from emergent prominent actors to examine how web publics revised, rearticulated, and reworked popular narratives in this networked environment characterized by affect and ambience” (Meraz & Papacharissi, 2013, p. 2, para 3).
Social Media Data Collection Best Practices

- The Ethics of Online Research Journal
- Chapter 4 Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges
- Twitter is among most accessible seen as public
- Facebook, however, seen as private since users control security levels, for example
- Obtaining private data requires social media company permission and can be challenging to obtain
Dr. Erik Bean

• Quantitative Content Analysis
• Almost Pandemic Proof. No Human Interaction
• No Formal IRB Approval Since No Human Subjects
• But IRB Approval Necessary at Doctoral Student Level
• Primary Sources Include: Websites, Books, Videos, Almost Any Published or Visible Materials.
• When Measuring: Important to Determine Proper Count Instances, Understand Screen Resolution if Measuring Square Inches.
• Highly Dependent on Pre-Research Operationalizing and Well Developed Coding Book of Definitions.
Collecting Web Data Using Software

• An Academy Customer Experience Benchmark Observation. Erik’s 2015 CLSOR Fellowship

• This study conducted in 2015, examined the content of the first CLSER academy research center website to determine the degree of customer experience (CX) best practices built in as well as survey those affiliated with it to determine their degree of CX expectations met.
Study, Purpose, & Design

- Do universities understand and use CX leadership?
- University of Phoenix, nontraditional students.
- Most scrutinized university since 1976.
- Launches new 2014 Website to complement the offerings of its Center for Leadership Studies and Educational Research (CLSER).
- Target: Practitioner doctoral faculty and students to obtain a bona-fide research agenda. Support them in the research process including scholarship and camaraderie among more seasoned and published colleagues.
Study, Purpose, & Design

- **Content analysis** of a University of Phoenix research center website iteration to determine the level of CX theory used as well as the personas of its affiliates—dissertation chairs, students—who use the center to help them prepare research studies for publication.

- **Surveyed** affiliates to see if implicit promises made, were kept, a requirement of purposeful CX design.
Website Software Pros and Cons

- “Corpus Linguistics Content Analysis software was selected as a basic algorithmic tool to parse through CLSER pages to examine the frequency of terms that were operationalized as promises. Such software can parse only those pages on the site that belong to the root CLSER (see the Appendix). While this tool provided the frequency data, like most such algorithmic text analyzers, it cannot readily determine the journalistic prominence of such messages (Budd, 1964).” (Bean, 2017, p. 24).
Website Software Pros and Cons

• “A total of 2,775 different words and numbers also were found within the CLSER site via the software. For example, the word “and” was the most popular at 831 instances. A total of 1,348 words were only denoted once on the site. For this study, the words that can intrinsically motive scholars to more affordably perform such scholarly activities and that equate to a promise of fundamental support (from conception through publication) were operationalized. These included: funding (1x), financial (50x), scholarship (59x), fellowship (25x), stipend (0x), opportunity (3x), opportunities (8x), as well as words that were deemed to encourage prospects to start the process such as: apply (0x), applying (1x), assist (0x), assistance (0x), help (0x), contact (3x), email (2x), e-mail (0x), start (0x), call (41x), and questions (0)” (Bean, 2017, p. 25).
Personas

A fictional customer who represents a real segment of your market.

© 2015 Customer Experience Rules!
Text by Jeofrey Bean
and Illustrations by Steve Hickner
Categorizing Personas

Persona A
Full-time chair with peered publication/presentation experience

Persona B
Part-time chair with peered publication/presentation experience

Persona C
Part-time chair with no peered publication experience
Website Promises Made, Promises Kept?

• **Persona A**
  Full-time chair with peered publication/presentation experience

• **Persona B**
  Part-time chair with peered publication/presentation experience

• **Persona C**
  Part-time chair with no peered publication experience agreed most promises made, were kept

Q5. Degree of Website Promises Met By Persona

[Bar chart showing the degree of website promises met by persona A, B, and C across different types of response: No Response, V. Distort, S. Distort, Neither, Clear, Accurate.]

University of Phoenix®
School of Advanced Studies
Key Findings

1. The results demonstrated that the personas who needed the most publication assistance did agree that promises made were kept.

2. The study documented a default CX Website iteration so that improvements could help move affiliate customers along the CX continuum with more interactions such as interactive webpages they requested.
Recommended Academic Research Software

- Sage Ocean Tips and Tools
- Corpus Software
- Provalis Software
- Nvivo

Access to social media data largest obstacle in any study due to IRB and COR approvals as well as social media companies themselves.
• Nothing replaces thoughtful traditional research, operationalizing data variables, and a coding book of definitions.

• Since the 2016 election the U.S.A. is one of only a few countries that sell our personal data to various agencies across the world.

• Let us never forget Cambridge Analytics, a U.K. company that took Facebook user data without permission.

• Since the 2016 election the U.S.A. is one of only a few countries that sell our personal data to various agencies across the world.

• [https://www.acxiom.com/](https://www.acxiom.com/) [Hacking Your Life, PBS]
• Algorithms in social media applied to networked gatekeeping and crowdsourcing practices can elevate ordinary user prominence to influence a forced perspective – making social issues appear larger or smaller than they really are (Meraz & Papacharissi, 2013). “We found that influential, central figures wielded inordinate influence over a crowd that both promoted and spread elite influence in a contagious manner through cascading waves of retweets and mentions” (p. 21)

• Gatekeeping theory and Social Science Networking theory applied to Twitter platform to support online conversation and enable combined processes of networked framing and gatekeeping for curating social media content – fast and furious capabilities via algorithms

• Can be used for good or bad. “One person’s prudence is another’s provocation” (Migliore & Bean, 2020)
Data Brokering Industry & Ethics

**Data Ethics:** “Outside the regulations of key federal and state laws are opportunities to mine data and strike gold using Big Data with Machine Learning practices by collecting data on purchases and on social media pages and piecing together profiles of consumers” (Migliore & Chinta, 2017, p. 50). “One person’s prudence is another’s provocation” (Migliore & Bean, 2020) – different perspectives of value-added customer experience (Cx).

**Categories used by data brokers for marketing purposes:** *Plus-size Apparel, AfricanAmerican Professional, Biker/Hell’s Angels, Allergy Sufferer, Exercise - Sporty Living, Working Class Mom, Burdened by Debt: Singles, X-tra Needy, Credit Crunched: City Families, Ethnic Second-City Stragglers, Fragile Families, and Small Town Shallow Pockets* (reported by U.S. Federal Trade Commission (FTC) and U.S. Senate Commerce Committee).

**How do consumers feel about being put into one or more of these categories?** – the average consumer does not know! Online data inputs feed data science via social media platforms, retailer records including credit and debit card purchases, or information voluntarily shared online by consumers - surveys, sweepstakes entries, and warranty registrations. Also, generic information such as name, phone number, residence, buying habits, etc. which are collected from courthouse records, mortgages, voter registration information, website cookies, and loyalty card programs.

Algorithms make decisions on employment, credit scores, and insurance rates. He/she who controls the algorithm controls the information.
Public Data Sets to Support Social Media Research

Panel Study of Income Dynamics (PSID) – University of Michigan
https://psidonline.isr.umich.edu/

“Longest running longitudinal household survey in the world.”

“The study began in 1968 with a nationally representative sample of over 18,000 individuals living in 5,000 families in the United States. Information on these individuals and their descendants has been collected continuously, including data covering employment, income, wealth, expenditures, health, marriage, childbearing, child development, philanthropy, education, and numerous other topics. The PSID is directed by faculty at the University of Michigan, and the data are available on this website without cost to researchers and analysts.”

“The PSID was originally created to study the dynamics of income and poverty. Over the past five decades, the study has been expanded to collect information on health, wealth, expenditures, philanthropy, child development, the transition to adulthood, and much more.”

“The PSID has now collected data for 50 years from more than 77,000 individuals. As of 2017, more than 11,000 families are followed, and there are as many as six generations represented within sample families.”
Questions?

Erik Bean, Ed.D.
Associate University Research Chair
Center for Leadership Studies and
Organizational Research
deadline@email.phoenix.edu

LauraAnn Migliore-Nolan, Ph.D.
Doctoral Instructor - Doctor of Management in
Organizational Leadership Program – DBA
University of Phoenix - College of Doctoral Studies
References


