Update on Innovations at NORC

Presented at FedCASIC

March 19, 2013

Angela DeBello
Vice President of Strategic Innovation

NORC at the UNIVERSITY of CHICAGO

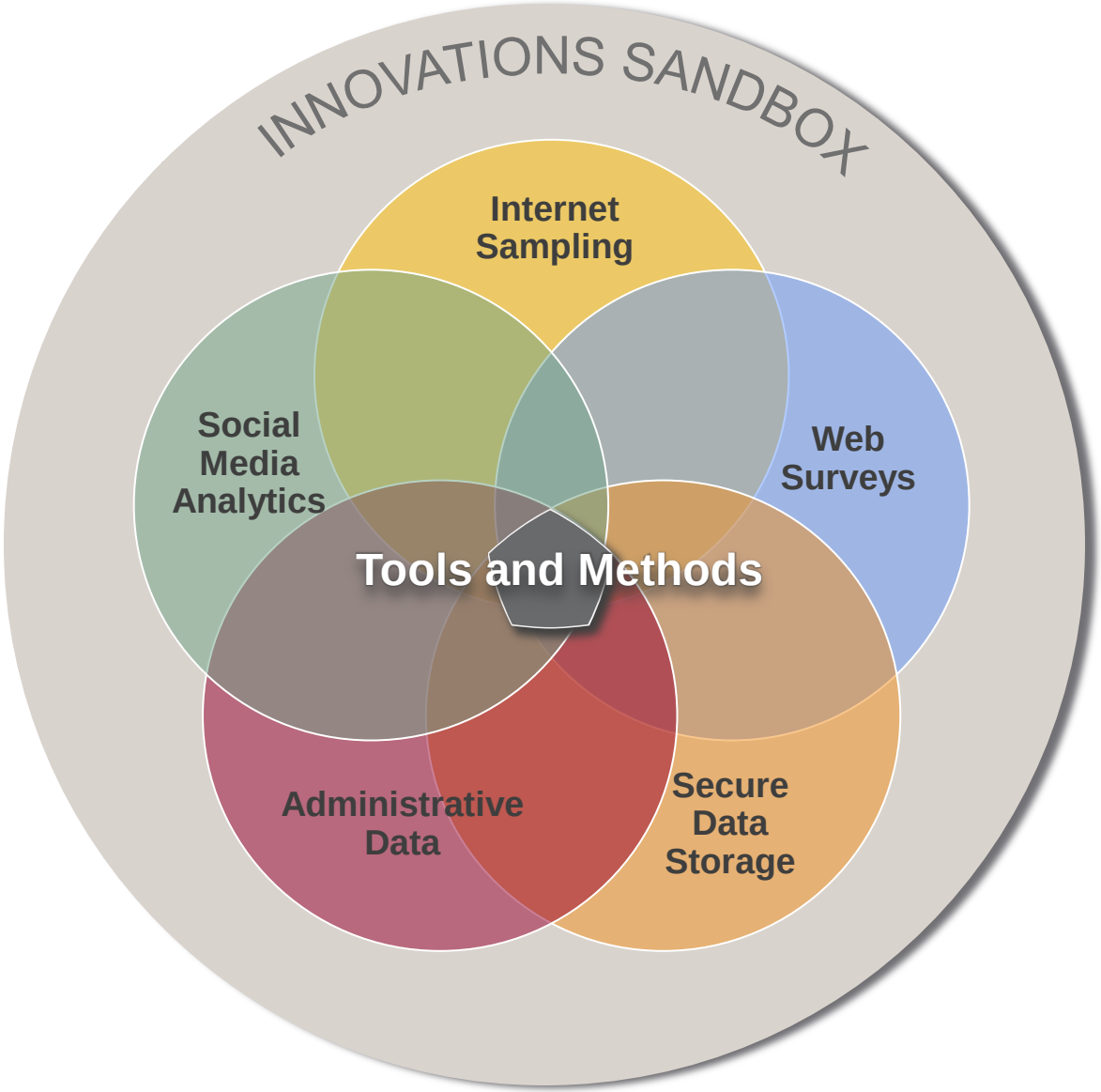
What Innovation Means at NORC

- In a quickly-changing world, NORC closely monitors a wide range of emerging ideas, technologies and trends that are:
  - Disrupting our industry's traditional means of doing work
  - Creating new opportunities to fulfill NORC's mission
- By quickly and routinely identifying, analyzing and acting on these emerging, we can move the industry in new directions

What Innovation Means at NORC

- NORC cultivates a culture toward innovation and adaptability – everyone is an “innovator”
- Our pursuit of innovative approaches is continuous and deeply embedded in the way we work
- First ever ‘advisory council’ made up of staff at all levels to drive innovation and challenge our thinking

NORC's Major Areas in Innovative Data and Analytics Work



Social Media Analytics

- NORC views social media as an extension and compliment to its existing approach—not as a substitute or replacement

- Traditional survey response rates continue to decline as respondents become less and less willing to share information with researchers
  - But every day millions of people post huge quantities of information about themselves on the public web

- This data represents a huge opportunity if we can find a way to make sense of it and take advantage of the it

Social Media Analytics

- This field has been largely driven by computer scientists
- Unclear if valid inferences can be drawn from social media analyses
  - All efforts to gauge accuracy of inference have relied on an external “gold standard” (typically a survey)
  - There is nothing akin to probability theory to allow us to gauge the accuracy of social media inferences without a gold standard
- Beyond inference, social media can still be useful for a variety of other purposes.
  - Qualitative explorations of the range of opinions (“how are people talking about issues?”)
  - “organic” focus-groups
  - Maintaining contact with longitudinal respondents

Selected Social Media Projects at NORC

- Examining Residential Rental Prices Through Craigslist
  - Seeks to identify determinants of rental rates in different housing markets using data taken from Craigslist
- Using Twitter Data to Understand the Chicago Teacher Union Strike
  - Tracks the sentiment of Twitter users related to the CTU strike and attempts to correlate them with specific news events
- Can Weighting Twitter Data Improve Inference?
  - Seeks to extract latent demographic characteristics of Twitter users and use those characteristics to draw more accurate inference from Twitter data
- Social Media "Data Kitchen"
  - Creates a standardized framework and toolset for extracting, storing, and analyzing social media data

Refining Web-Based Surveys for the Next Generation

- Web-based surveying is not new - our focus is on the next wave of “challenges” to this mode
- Are Response Patterns to an Address-Based Sample, Web-Only Survey Spatially Clustered?
  - Uses spatial regression models to assess how community-level variables affect response rates to a web-only survey based on a random, address-based sample
- Google Single-Item Surveys vs. National Standards
  - In this experiment, 1,500 responses from a Google 1-item survey for a commonly used question addressing cell phone usage is compared to several national level, probability surveys to assess the 1) generalizability of the Google results as well as 2) what weighting strategies can be applied with these data

Refining Web-Based Surveys for the Next Generation

- Google Single-Item Surveys vs. Cognitive Interviews
  - Collected 4000 responses from four Google 1-item surveys for questions formats we had previously tested in cognitive interviews. Will now be compare these results to those from the cognitive interviews to assess the comparability and usefulness of the Google results

- Three Major Web/Mixed Mode Experiments for the National Immunization Survey (NIS)
  - Web based respondent instrument, text message follow-up, and provider technology-use assessment

Refining Web-Based Surveys for the Next Generation

- Minimize Fraudulent Manipulation and Multiple Responses to Initially-Anonymous Web Surveys
  - Need “layered checks” too discourage users from repeating the instrument multiple times and from abusing incentives

    - **Entry Path:** First, we used different URLs to judge the source path respondents took to reach the survey
    - **IP Checking:** Next, we determined the IP address used by the potential respondent and looked to see if that was associated with previous access or completes
    - **Cookie Checking:** We set or checked a cookie with project ID and expiration date that was validated against the survey. Users considered to be returning before the expiration date were thanked for previous responses
    - **Email Checking:** Address needed both for valid return and for incentive delivery. It is requested at breakoff or at completion. Checking happens just after the completion question and before the “thank you” screen. The entered email address is compared to all email addresses previously received. Matches are treated as errors and the respondent will not receive an incentive

Understanding Internet Samples

- Compares four probability and non-probability based internet sampling approaches to assess the data quality and efficiency obtained via each sampling strategy
  - Address Based Sample (ABS) national sample
  - Email Blast List sampled from InfoUSA database
  - Facebook (opt-in/self-selection)
  - Google (opt-in/self-selection)
- Web-only questionnaire
- Testing various denominations of post-completion Amazon gift cards

Secure Data Storage and Access

- The **NORC Data Enclave a**llows the sharing, among a closed community of researchers, of datasets that are too sensitive to share broadly.
  - archives, curates, and indexes the data; provides researchers remote and onsite secure access to the data; and protects confidential information
- The NORC Data Enclave team
  - manages an active outreach program to inform the national research community of the data and to foster the use of the data in research leading to conference presentations and journal publications
  - conducts an extensive education and training program to ensure appropriate use and disclosure of the data, including its confidentiality.

Harnessing Administrative Data for Research

- Data accumulated/collected in the course of programmatic activities ("non-research") can have enormous contribution to research
  - Often thought of as a "secondary" data source
  - Cleaning, processing/transforming and preparation are essential in order to assure a useable analytic data file
  - Can cover entire populations/good for low-incidence groups
  - Can help provide accurate measures of undesirable self-reported items
  - Often lacks documentation, assessment of data quality
  - Is often "big data" that requires special means of manipulation, techniques, etc

Harnessing Administrative Data for Research

- CMS Linkable Medicare Data Entrepreneurs' Synthetic Public Use Files (DE-SynPUF)
  - First Fully Synthetic Set of Medicare Claims Data
- Basic Stand Alone (BSA) Medicare Claims Public Use Files (PUFs)
  - Testing of de-identification for researcher use
- National Children's Study (NCS)
  - Creating a comprehensive list of all data for possible linking in the future as part of 21 year study
- National Immunization Survey (NIS) Provider Record Check
  - State of the art system for combining survey data with medical record data on immunizations

And so much more…

Thank you!

www.norc.org