

Introduction

Social networks engulf everyday life. They represent a place to share news, ideas and information of all kinds. The connections made among people in these networks, and the resulting information shared, can have a profound effect on the thoughts, attitudes and beliefs of individuals. Moreover, even the flow of information itself, can be a powerful predictor of key business and program outcomes.

Recognizing the power of social networks, opinion researchers have increasingly begun to take advantage of social media to answer critical business questions. In doing so, the research profession has invented new tools and methods to supplement an already impressive array of techniques. The Marketing Research Association (MRA) has developed this guide in order to describe the current landscape of social media research as well as to facilitate and advance further development of the technique. Ultimately, it is the goal of the Association and its members to foster universally accepted and practiced standards and best practices for these and other research methods.

What is Social Media?

There are many definitions of social media but, at its core, social media uses Internet-based technologies that facilitate the creation and exchange of user-generated content. Social media refers to Web sites that permit people to interact with the site and with each other using simple interfaces. At the time of publication, Facebook, qq.com, Twitter and YouTube are among the most popular social media sites.

Social media refers to the information that people share on those sites, including status updates, image and video comments, responses to blogs and forums, and any other individual contributions to the online space. This information reflects naturally occurring conversations among people who may or may not personally know each other.

What is Social Media Research?

Though evolving rapidly, social media research (SMR) is the application of marketing and opinion research methods to social media data for the purposes of conducting research (e.g., usage and attitude studies, social media research tracking studies, custom research, etc.). Similar to other types of marketing research usage and attitude studies, tracking studies, research goals and objectives are developed, methodologies are prepared, and social media data are analyzed quantitatively and/or qualitatively depending on the goals of the project.

SMR is distinct from other forms of marketing research in that it uses social media as its data source as opposed to surveys, focus groups and other data collection modes and techniques. SMR can be a complementary or stand-alone analytical tool for researchers, providing them with a unique opportunity to listen and measure the opinions of potentially vast numbers of people who communicate online, some of whom may not normally or easily be accessible through non-observational forms of research.

About the Authors

MRA is grateful to the following for their contributions to this *Guide to the Top 16 Social Media Research Questions*: Jim Longo, PRC, Itracks, Committee Chair; Janet Savoie, PRC, Online Survey Solution; Annie Pettit, Conversation Strategies; Ray Poynter, The Future Place; Ellie Schwartz; Ed Sugar, PRC, OLC Global; Tamara Barber, Forrester Research; Tamara Kenworthy, PRC, On Point Strategies; Steven Runfeldt, Schwartz Consulting; Benjamin Smith, Spsych Market Analytics; Aaron Hill, PRC, Sawtooth Software; Susan Saurage-Altenloh, PRC; Steffen Hück, HVYE; and Patrick Glaser, MRA.

1 What are the advantages and disadvantages of SMR?

From a capacity standpoint, SMR provides the ability to collect and analyze information from the past as well as in real-time, as it is generated. Moreover, the richness of data available on social media networks is conducive to both qualitative designs (e.g., digital ethnographies) as well as quantitative designs, including numerical aggregation of large quantities of data.

In terms of methodological considerations, SMR utilizes an observational form of data collection. Information is collected from Web sites as posted by individuals who may not be specifically aware of the research role. As such, social media communications are thought to be free of, or less subject to, response biases that occurs in interviewer-administered, and even self-administered, forms of opinion surveys and focus groups. However, social media is inherently a public form of communication, with varying degrees of privacy which may affect some social media user's willingness to reveal information, particularly sensitive or potentially embarrassing personal details.

From an ethical standpoint, SMR has the additional advantage of eliminating the burden that would otherwise be placed on a research participant. Social media users do not participate in "active" data collection (e.g., survey, focus group). They generate data simply by engaging in their natural online communications. However, SMR presents unique ethical considerations of which researchers must be aware (see "Ethical and Legal Issues").

SMR offers researchers a host of benefits, a few of which include:

- ▶ Ease of adjusting research criteria throughout the study
- ▶ Potential cost savings and reduced logistical burden
- ▶ Ease of application across locations
- ▶ Access to hard-to-reach research participants
- ▶ Benchmarking (e.g., reported vs. observed opinions)

Likewise, researchers should be aware of various challenges associated with SMR. For example, researchers who are new to SMR methods will need to familiarize themselves with both the characteristics of social media users as well as specific SM sites in order to properly draw conclusions about research findings. Additional considerations include the need to learn and become proficient with:

- ▶ SM tools and techniques including sentiment and content analysis
- ▶ Indicators of SMR validity and reliability at each stage of the process
- ▶ Relevant types of biases, particularly those arising from unique SMR tools
- ▶ The types of brands and categories that are more likely to be successful carrying out SMR, e.g., due to volume of data or consumer importance

2 What data sources are typically used in SMR?

Millions of Web sites (small and large) currently facilitate the practice of social media research. However, online sites, which currently facilitate social media communications come and go, and change very rapidly. Researchers involved in SMR need to stay abreast of changes in social media communication patterns and trends, including the rise of mobile access, and popular SM vehicles. Current examples of SM Web sites that generate data suitable for SMR include:

- ▶ Social Networking Sites:
 - Facebook: Search, Community Pages, Fan Pages, Groups, Chat, Facebook-based Apps
 - Twitter: Location-based Application, Real-time Search, Advanced Search (search.twitter.com)
 - LinkedIn: Search, Groups, Q&A
- ▶ Social News: e.g., Digg, Reddit, Mashable, Technorati
- ▶ Photo/Video Sharing: e.g., YouTube, Flickr
- ▶ Online Communities: Industry, Topic-related, Branded or Unbranded
- ▶ Blogs: e.g., Blogger, Posterous, Wordpress
- ▶ Forums: Industry or Topic-related
- ▶ Questions and Answers: e.g., Yahoo Answers, LinkedIn Answers, Yedda
- ▶ Commenting: e.g., Disqus, Backtype
- ▶ Traditional News: e.g., CNN, *BusinessWeek*

3 How does SMR interact with other forms of traditional and non-traditional research, including online, offline, in-person, and qualitative and quantitative?

SMR can effectively stand on its own, but may also be integrated with traditional research methods to create a holistic research solution. In fact, SMR may sometimes springboard or support other forms of traditional research. Examples of SMR integration with other research methods include:

- ▶ Observing the flow of conversation in real time, thus prompting the most effective methodology for further research
- ▶ Accessing user supplied media such as photos and video
- ▶ Measuring trending topics for further “traditional” research
- ▶ Assisting in the preparation of discussion guides or surveys
- ▶ Identifying key influencers in an industry or on a topic
- ▶ Reaching a segment of the population that may not otherwise be reachable
- ▶ Comparing community-based insights to natural observational social media insights
- ▶ Establishing trust between researcher and participant, potentially for further recruitment into another form of research
- ▶ Exploring, and discovering “unknowns” via observations

4

How reliable are SMR results?

Validity refers to the degree to which results reflect truth or reality while reliability reflects the degree to which results can be replicated if someone else were to conduct a similar study. Because research suppliers have different methods, standards of quality, and processing rules, research consumers must conduct their own validity and reliability analysis of any potential supplier to ensure the quality of work is sufficient. As with all types of marketing research, the validity and reliability of social media research varies greatly:

- ▶ What is the validity and reliability of the sentiment and/or content analysis processes? If manual coders are used, reliability might be lower. If automated coders are used, validity might be lower.
- ▶ Given that sentiment differs by Web site (e.g., Twitter is more negative while blogs are more positive), what is the range of social media venues that are measured and what percentage of the Internet population do they represent? Do any of the sites overwhelm the data collection strategy in a proportion that does not reflect the Internet space? Does the vendor know how and why to sample and weight data?
- ▶ To what extent is the intended target group reflected by the social media venues being used?
- ▶ Is the intention to measure and generalize to the general Internet population or to a particular segment of the Internet?
- ▶ How is geographic and demographic information being measured in order to assess the validity of generalizing outside of the sample?
- ▶ What timeframe is appropriate for the research objectives? Though small samples may be acceptable for long-term research, shorter time frames must use larger sample sizes.

5

Within businesses and organizations, how will SMR activities be tracked and aggregated, and whose responsibility is it to handle each of those functions?

Social media research may be executed in multiple ways. For example, numerous departments within a single company may be involved in SMR, including internal research departments, and cross-functional teams from marketing, customer relationship management, public relations, public affairs, and other departments. SMR may also be outsourced to vendors who may or may not specialize in research. Regardless, the skill set of the user must be appropriate for the function.

6 What additional knowledge, skills and abilities will a corporate researcher need to learn in order to improve their level of competency with SMR?

SMR may involve several different methods and analytical approaches. As such, corporate researchers may find it most advantageous to learn a wide breadth of relevant techniques while continually honing their skills and knowledge in the areas that are most relevant to their organization. Commonly used techniques include both sentiment analysis and content analysis. Additionally, researchers will need to learn about, and become comfortable with, important explanatory variables beyond traditional “respondent” demographics, such as how different types of Web sites (e.g., blogs, forums, media, etc.) generate and facilitate different types of data (e.g., whether data is more positive versus negative, descriptive versus condensed, etc.).

7 Are the participants aware that their user-generated content is under observation?

Research contributors have demonstrated the occasional tendency to provide sub-optimal information when they are aware that others are studying or observing them. Oftentimes, this is attributable to concerns over the privacy of sensitive information or feelings of being compelled to give a socially-desirable response to a question. In SMR, though it commonly is understood that conversations are generally public and open to viewing by almost anyone, the individual under observation may or may not be aware of the presence of a researcher.

At the same time, participation in the social media space offers varying degrees of privacy. Users may participate for personal and/or professional reasons and they may or not seek relationships with other users. Researchers should be aware of the potential and likelihood for “social observational bias” and the effect it will have on the type, candor and direction of the user’s comments.

ETHICAL AND LEGAL ISSUES

8 How are sources cited in research reports and on research Web portals? Are the citations different based on the source, e.g., Twitter, Blogger, forums?

As in traditional forms of research, it is important to protect the privacy of contributors. As such, without prior express consent, data transmitted from vendor to client should not include direct references or citations to individuals that would reveal their identity.

However, sources may be recorded for validation purposes as well as for potential data quality checks. Any data or reporting intended for transfer to an outside entity should be purged of personally identifiable information (PII) prior to changing-hands. This includes IP addresses, usernames, user id numbers, user photos, e-mail addresses, and other types of commonly available online data.

Where detailed information must be shared for the purposes of data quality or validation, the data should include source citations using the current link of the information (e.g., <http://twitter.com/xxxx/xxxx/>). Notably, links should be expected to expire or become “broken” over-time. Researchers should plan to record any pertinent administrative or relevant source data (e.g., date/time, source identifier, query details, etc.) to be used in validation at the time of data collection.

9

What are the controversies and legal issues regarding the rights of the people whose data is being used?

Social media is a relatively new form of communication and individuals from every stakeholder group, including the public, researchers and governments, are participating in an on-going conversation about the nature of its privacy and ethics. For this reason, it's critical for researchers to understand that they have a responsibility to respect social media user's privacy and that the definition and expectations for social media user's privacy can and will change over time. Some brief areas of consideration are described below.

Privacy: Individuals and their social media privacy expectations should be respected. If an individual has posted information on a public Web site under a public “privacy” setting, they may be considered to have a very low or no expectation of privacy for the information they reveal. Even so, researchers who collect and analyze this information should take care to protect it from becoming identifiable to an individual.

Conversations should not be copied verbatim into reports as those direct quotes can be searched and identities discovered. A small number of relevant conversations can be summarized, without losing their flavor, in reports. Moreover, full quotations can be used with permission.

Interacting with individuals: Clients must never use information collected during or for social media research for the purpose of direct marketing or otherwise influencing the opinions and behaviors of the data subject. Marketing may only occur in places like branded and client communities where contributors would naturally expect those types of conversations to take place.

Combining data from multiple sources where privacy policies differ: In general, the policy provisions that tend to favor the rights and needs of the contributors should be given weight. Best practices call for researchers to respect the coded crawling terms of every Web site they visit. Where Web sites are coded to indicate that crawling is not permitted, those Web sites should not be crawled even if it is technically possible. Researchers must not join Web sites under the pretense of being a member so that they then have access to crawl a Web site that prohibits such crawling otherwise – this condition holds for both automated

and manual crawling. Where researchers do join groups, they must immediately make it explicit that they are there for the purposes of marketing research. Notably, issues concerning access to data sources are paramount to the conduct of social media research and can be expected to be a major focus of the opinion research industry moving forward, both in terms of how to ethically gain access to the widest net of sources as well as appropriate ways to handle and adjust for cases where this is not possible.

SM RESEARCH PROCESSES & PROVIDERS

10 What is the level of expertise and industry qualifications of social media researchers and/or SMR companies?

Anyone selecting a social media research vendor must be aware that the technique is relatively new. They must be careful to select a research partner with the appropriate level of expertise and skill in the practice of SMR. Some relevant questions to ask include:

- ▶ Is the company primarily an IT or social media company that expanded into research, or a research company that expanded into social media? While IT and social media companies may have expertise in social media, crawling and data collection techniques, research companies have expertise in data analysis techniques.
- ▶ Does the company focus on research exclusively or do they maintain other functions as well? For example, companies that conduct SMR may specialize in buzz monitoring, customer relationship management, public relations, research, or some other social media function.
- ▶ Does the company specialize in qualitative methods, quantitative methods, or a combination of both?
- ▶ Is the provider aware of traditional research practices such as sampling and weighting and, if so, how and when do they apply those practices?
- ▶ For the practice of ethics and standards of quality, does the provider classify themselves as a researcher or as some other profession?

11 What are the standard data and/or research outputs?

Since SMR is relatively new, industry standards for outputs have not yet been developed. It is important to understand the vendor's policies and capacities for standard and custom reporting. Relevant questions include:

- ▶ Does the company offer a full-service model of data collection, analysis and presentation or do they offer a self-service tool such as a portal?
- ▶ In cases where the vendor offers full-service reporting and presentation, what substantive outputs may be expected? What technical explanation and reporting may be expected (e.g., a technical appendix)?

- ▶ Are the SMR analyses incorporated with traditional types of marketing research and does the company have expertise doing so?
- ▶ Does the provider offer standardized or customized tools?
- ▶ How often are outputs updated and/or delivered?

12

What is the process for gathering data?

Like other forms of opinion research, a wide variety of approaches exist for the implementation of SMR. It is important to understand the company policies undertaken.

Relevant questions include:

- ▶ Does the company gather its own data or is a data collection vendor used?
- ▶ How many Web sites are crawled and how are those Web sites selected?
- ▶ Does the company seek out permission-based relationships with the sites they crawl?
- ▶ Does the company honor the electronic privacy notifications of individual Web sites?

13

What data quality processes are implemented in each stage of the SMR?

What quality and validation protocols have been adopted and implemented to safeguard the quality of the research at each stage of the process? Are there validation processes in place for initial data collection, scoring and coding, etc.? Does the organization collect and retain information at the initial stages for validation purposes while removing/anonymizing data for reporting purposes?

14

Does the company provide sentiment scoring?

Sentiment scoring is a process of assigning a positive or negative emotion to a conversation. Some vendors may provide strictly positive or negative emotions, while others may assign a continuum ranging from positive to neutral, to negative. If the vendor provides sentiment scoring, is the process an internal proprietary method, a third party purchased product, or some combination of the two? How is the sentiment scored (e.g., dictionary, bayesian, manually)?

15

If sentiment scoring is provided, what is the process for validating results?

Simple and commonly-used systems of sentiment validation may prove to be inadequate. More rigorous approaches should be used, specifically blinded methods. For example:

For automated systems, researchers should receive a list of uncoded conversations and then code them manually. The manual codes should then be matched back and compared to the automated codes to derive a percentage match (i.e., validation coefficient).

For manual systems, two unique raters should independently code conversations. A validation coefficient may be derived from a comparison of the two outputs.

The above processes are two relatively simple examples of validation systems. More complicated calculations are available, but their use should be weighed according to the capacity of stakeholders to understand the meaning and method of the technique.

Language constantly changes and evolves due to new and lapsed slang, terminology, and speech patterns. As such, simple systems of sentiment validation may prove to be inadequate. When conducting SMR, rigorous and constantly monitored approaches to sentiment analysis are most appropriate.

16

What, if any, methods are used for determining the geography associated with the data?

Demographic and geographic information can often be an important and meaningful element for research and validation purposes. When considering SMR, what geographic information is available and how precise is the information (e.g., city or town, region, country, unknown)? What types of demographic data are available (e.g., age, gender, income, education)?

Researchers must take care to specify the methodology and sample size associated with the information. Inferred methods (based on Web site sources or language) may be associated with large sample sizes but have low validity. On the other hand, precise information is currently only available for an extremely tiny percentage of conversations and therefore often has insufficient generalizability.

Conclusion

The “Top 16 Questions” presented in this guide represent the core matters of importance to the research field with respect to social media research. They include issues of reliability, execution, interaction with other kinds of research, ethics and legal compliance, data quality, process, and outputs.

Importantly, the 16 questions in this document do not stand as the only ones the opinion research profession needs to address, nor do they take the place of standards of practice. Instead, they provide a starting point for experts and professionals to debate and discuss development toward this goal. As in any profession, a reasonable consensus should be reached in order to validly define and represent an industry standard of best practice. It is the goal of the Marketing Research Association that this document be widely distributed and contribute as such.