Sport Fans and Online Data Collection: Challenges and Ethics

Brody J. Ruihley
Robin Hardin

Follow this and additional works at: https://trace.tennessee.edu/jasm

Recommended Citation

Available at: https://trace.tennessee.edu/jasm/vol6/iss3/6
Sport Fans and Online Data Collection

Challenges and Ethics

Brody J. Ruihley
Robin Hardin

Abstract

The growth of online communities and social networking has provided opportunities to investigate sport fans from a wide range of perspectives. Motivations to consume online media and engage in interactive web functions are areas providing new and innovative research opportunities. There are several ethical considerations when conducting research in an online environment. This article discusses four major ethical values of honesty, responsibility, justice, and beneficence and how each relates to online data collection. Specifically, these four values will guide the discussion focused on issues of intrusion, interaction, and invitation in online communication contexts. Researchers and administrators must consider fans and other stakeholders’ core moral and ethical values in the data collection process.

Keywords: ethics; values, data collection; Internet; social networking

Brody J. Ruihley is an assistant professor of Sport Administration at the University of Cincinnati.

Robin Hardin is an associate professor in the department of Kinesiology, Recreation, & Sport Studies at the University of Tennessee.

Please send correspondence regarding this article to Robin Hardin, robh@utk.edu
Online Data Collection

Internet and Web usage are on the rise. Adults have access and are using these technologies at an astounding rate. In 2013, Pew Internet and the American Life Project (PEW, Pew Internet and the American Life Project, 2013) found that nearly 85% of American adults are online and using the Internet. This is a sizeable difference from first estimates in 1995 when usage was measured at 14% (PEW, 2013) and again in 1998, measured at 35% (Harris Interactive, 2008). PEW (2011) also reports 83% of Internet users are online to obtain information on a hobby or interest. Additionally, 52% of Internet users seek out news or information about sports (Pew, 2011). With this type of reported usage, collecting fan or consumer data using online communication outlets is quickly becoming a frequent choice for contemporary researchers. This is evident with the prevalence of contemporary research pieces focusing on Twitter, Facebook, and other social media outlets. Journals such as the International Journal of Sport Communication, Sport Marketing Quarterly, Communication and Sport, and Communication and Society are publishing articles based on the social media phenomena. Taking a leading role in recent research are topics such as NCAA Organizational Facebook pages (Wallace, Wilson, & Miloch, 2011), authenticity and engagement with Facebook for professional sport teams (Pronschinske, Groza, & Walker, 2012), and networked fandom through Twitter (Clavio, Burch, & Frederick, 2012). Data collection, for the aforementioned studies and all types of exploration, are able to move away from pencil and paper surveys and towards online questionnaires, online focus groups, and content analyses. This issue of the Journal of Applied Sport Management recognizes the impact technology is having all aspects of sport management from ticketing, to event management, to personal management, to understanding consumer behavior. This manuscript examines issues associated with collecting data in the ever-changing technological environment.

The main reason researchers can begin to move their research activities online is because sport fans have shifted parts of their fandom to online activities. Sport fans are consuming and interacting in online environments. This is occurring within social networks in the form of message boards, Twitter, and Facebook, and in the online activities of viewing or listening to sporting events, gambling, and fantasy sport. Research focusing on online activities of fans is fronted by the work of Hur, Ko, and Valacich (2007) as well as Seo and Green (2008) with research studies focusing on the area of online sport consumption. These two studies represent a beginning in this type of research focusing on a specific type of consumer: the online sport consumer. The purpose of this type of research is to examine the behaviors (Hur et al., 2007) and motivations (Seo & Green, 2008) of online sport fans. Areas of examination include consumer behavior (O’Cass & Carlson, 2010; Pegoraro, O’Reilly, & Giguere, 2009; Williams & Chinn, 2010), fantasy sport motivations (Dwyer & Drayer, 2010; Dwyer & Kim, 2011; Farquhar & Meeds, 2007; Roy & Goss, 2007; Ruohley & Hardin, 2011a; Spinda & Haridakis, 2008), and message board use (Clavio, 2008; Ruohley & Hardin, 2011b).
The move to use the Web and Internet for communication and data collection purposes allows sport management and sport communication researchers a different means in contacting and interacting with potential research subjects. There are many advantages and disadvantages in collecting data through Internet- and web-based programs. One advantage includes having accessibility to fans nationally and globally. No longer is data collection restricted to one local fan base or those with whom you can physically communicate, rather it is now who can be contacted electronically. Other advantages of online data collection involve saving costs (paper and pencils not needed), staff resources (less people needed to physically collect data), and time (less time to collect data and transfer from paper to computer program) (DeVaus, 2002; Lefever, Dal, & Matthiasdottir, 2007). In addition, online data collection assists in organizing information, eliminating entry error, and reducing lost data experienced in traditional in-person gathering.

The advantages appear to be overwhelmingly positive, but it is important to note that there are still disadvantages to online data collection. Concerns of participant trustworthiness, sampling techniques, and technical issues arise in an online format (Lefever, Dal, & Matthiasdottir, 2007). Participant trustworthiness, as discussed further in this article, is an issue in any type of reporting and is only amplified as a result of not having direct interaction with participants (Boyd, 2003; Shapiro, 1987) and the possibility of receiving false responses (Lefever, Dal, & Matthiasdottir, 2007). There is a feeling of disconnect different from in-person interactions. Sampling problems are concerned with the nonrandom nature in which many online participants are invited (Lefever, Dal, & Matthiasdottir, 2007) and the difficulty in calculating response rate (DeVaus, 2002).

The advantages and disadvantages are not comprehensive or universal to all research circumstances. Individual situations, samples, and research questions will guide a researcher in determining if online data collection is the most appropriate avenue to use. With online data collection coming to the forefront in sport management and sport communication research, the purpose of this manuscript is to further clarify some of the challenges of this method, specifically addressing sport fans. While advantages and disadvantages might not be applicable to all research, understanding and acknowledging the ethical components of online data collection should be considered in every case. Therefore, in addition to the challenges, this manuscript highlights ethical values of honesty, responsibility, justice, and beneficence (Lumpkin, Stoll, & Beller, 2003) as each relate to data collection issues of intrusion, interaction, and research invitation.

Intrusion

One issue present in online data collection involves intrusion. Intrusion, for the purposes of this discussion, is newly defined as using message board, blog, or other social media posts as research data without the consent of the participants. Simply taking posts, words, or conversations off of a website and using pseud-
Online Data Collection

Onym代替原始屏幕名会提出一些基于责任、诚实和有益的问题。在许多情况下，用户是付费获取信息并在留言板上发布评论的，他们没有理由相信他们的评论会成为研究的一部分。还有社交媒体平台允许用户批准谁可以访问他们的信息，例如Facebook和Twitter。这些帖子不会是私人的，但是会有一部分控制谁可以访问用户信息和帖子。参与者知道他们将参与一项研究吗？如果人们知道他们的回答将用于学术环境，评论会有所不同吗？如何获得使用网站上的帖子或评论的许可？

指导原则由互联网研究协会制定，特别关注到社会媒体用户通常不清楚帖子是公开的，并且可以被基本上任何人访问（Ess & AoIR Ethics Working Committee, 2002）。社会媒体用户永远不应该假设互联网上的任何内容是安全的或私人的。研究参与者应该自由自愿地参与研究，而这在粉丝帖子在留言板等其他社交媒体网站被纳入研究时并不总是那样（Belmont Report, 1979）。2010年发生的一起事件证明了这种非隐私问题和帖子可能成为主流媒体和研究研究的问题。

Twitter本身是一个公开的论坛，因此对推文进行内容分析类似于对报纸或电视广播进行内容分析。两项研究对Twitter帖子进行了内容分析，以确定运动员正在向球迷发送什么样的信息。Pegoraro (2010)研究了
Twitter posts of the most followed athletes in particular sports during a one-week period. The subjects of the tweets were responding to fan posts and questions as the athletes were simply directly interacting with fans (Pegoraro, 2010). The other subject areas identified as the content of the communication were information about their personal life and business activities. (Pegoraro, 2010). Hambrick, Simmons, Greenhalgh, and Greenwell (2010) conducted a similar content analysis of athletes’ tweets to determine what the content of the messages were and found professional athletes were using Twitter to interact with fans and share information about their personal lives.

Ethical friction arises when non-public fan sites are used for research purposes. The posts at a subscription-based site have a feeling or sense of privacy. Users are entitled to a sense of privacy considering they are paying for content and the ability to post messages and should not have to be concerned if their comments will be used by a third party.

Another issue that can arise is that a researcher might make a post to solicit responses to an issue or idea without informing the participants the responses will be a part of a research study (Andrew, Pedersen, & McEvoy, 2011). A researcher can go to the message board site of a particular college and make a post about what the fans’ opinions were about the halftime promotion at a particular basketball game and then do the same for other halftime promotions. An analysis of the posts could be made, but when in fact it seems the researcher is actually using an opened-ended survey question to solicit responses. Fan responses to questions can be gathered quickly and in large numbers, but researchers must adhere to the guidelines involving human subjects in research in that informed consent must be granted by the respondents, and the study must be approved by the university’s Institutional Review Board (IRB). This confronts the value of justice when some researchers are appearing to cut corners while others may be taking the proper avenue to collect data (i.e., it is unfair to those conducting research the right way).

Ethical issues that evolve from online communication-based research tend to involve informed consent of the participants in not knowing they may be participating in a research study. Typical procedures in research include providing an informed consent form to sign, or language informing respondents that by completing a questionnaire or interview they give their informed consent simply by participating. However, in message board posts or blog posts, participants have no reason to believe they will be participating in a research study (Department of Health and Human Services, 1991). Thus, informed consent was not sought nor given. This does not mean every message board or social media site should post a disclaimer informing participants that their posts may be used as part of a research study. It does, however, raise issues in using posts in research studies. The posts or comments in many ways could be viewed as responses to open-ended questions in a questionnaire or interview, and there would certainly need to be informed consent and Institutional Review Board (IRB) approval to collect data.
in that manner (Department of Health and Human Services, 1991). There is also concern that the posts or comments used could be from minors and would require approval from the Institutional Review Board (Department of Health and Human Services, 1991). Beneficence comes into play whenever IRB is or should be involved. Utilizing minors or researching adults involves asking questions or using information from their life. The posts would not have to necessarily be from the researcher, either. Another member of the message board could pose a question to seek opinion or input on an issue. The researcher would still be using responses to an open-ended question even though the question was posed by someone else. This could lead to a situation of then determining if this is secondary data analysis, but regardless the information was obtained initially without IRB approval.

A final area of intrusion concerns the fact that this online environment makes protecting privacy even more challenging than in person. Respondents to an online questionnaire have an IP address from the computer used to complete the questionnaire recorded as part of the data collection. This can be used to potentially trace the origin of the where the submission of the responses occurred (Fairfield, 2012). Respondents could also be using a public computer available in a library or other facility and have someone visit the cache to see what websites the person has been visiting or using (Miyazaki, 2008). Respondents must be instructed to ensure they are doing all they can to protect their anonymity as well. There is also the possibility that someone can enter what it is supposed to be a secure environment and see the conversation in an online focus group. All precautions must be taken to provide confidentially and as much anonymity as possible (Semitsu, 2011).

Communication is changing rapidly in the sport landscape, and social media is providing a stage to gain more access, receive more information, and learn more about sport and sport fans. A new area of research has emerged in studying the online sports fan in terms of motivation, consumption, and interactivity. This is an evolving field and is continually changing, offering many areas of academic interest. Researchers have much to learn with this new type of coverage. With that said, it is important to note that just because this is a new area or trend in research, it does not mean all bets are off and rules and guidelines concerning research involving human subjects are no longer applicable. Researchers must still remain responsible and think critically about the manner in which they collect research data. In a similar vein, just because the subject area involves a hobby or activity like sport and is not covering a sensitive area (e.g., cancer, sexual orientation, etc.) doesn’t mean rules become less important (Miyazaki, 2008; Semitsu, 2011). There are still moral and ethical accountability (responsibility) concerns to conduct research in the right way (justice) and not intrude upon someone’s personal life. Researchers should not place an undue burden on the participants, being upfront about collection effort (honesty), and protecting the respondents from harm (be-
neficence). Harm does not only imply physical danger but also psychological and mental stress (Belmont Report, 1979).

**Interaction**

Online data collection entails many different interactional issues, including discussion format (e.g., blog, live chat, and email), real-time dialogue (asynchronous or synchronous), length of discussion (i.e., amount of time for each discussion), and anonymity. The first three issues are important areas of online research, but they are logistical in nature. Anonymity is an issue that presents the most ethical concerns and is defined as the “inability of others to identify an individual or for others to identify one’s self” (Christopherson, 2007, p. 3044). In many cases, anonymity is given to participants of research studies by either giving a pseudonym or by asking the person not to identify themselves in any way. This protection is similar to those using message boards, blogs, or other social media where a screen name is required. Hayne and Rice (1997) argue for two types of anonymity. First, technical anonymity involves removing any identifying information when material is exchanged. Examples of this type of anonymity involve removing one’s name by either using a pseudonym or not using a signature at all. Other aspects of technical identity may also require pictures, voices, or other identifiable information to be removed or covered up from a technical standpoint. The goal of this type of anonymity is for the content alone to influence the discussion (Hayne & Rice, 1997). Even with technical anonymity, one can be identified by the thoughts, ideas, and social prompts given in the discussion. The second type, social anonymity, involves removing those social prompts and any other aspect of individualization or other identifiable social cues that may allow someone to be identified.

There are many positive outcomes resulting from anonymity in the data collection process. The first highlights the value of beneficence (i.e., not doing harm) by offering the protection of the participants or fans contributing. Many research studies, both online and in person, utilize anonymity to allow the participant the ability to disclose information without fear of repercussion or retribution. Protection is also given when the discussion is about sensitive or private topics. Research involving personal information (e.g., sexual orientation or habits, income, job satisfaction, household structure, marital status, etc.) may be hindered significantly if anonymity is not granted and honored in the data collection process. Another positive outcome of anonymity is allowing those normally reserved or timid the opportunity to voice their opinion in a non-threatening environment. Valacich, Dennis, and Nunamaker (1992) argue that “reduced inhibitions may encourage greater participation of junior or shy group members” (p. 54) and provide a wider variety of opinions.

Anonymity creates ethical dilemmas surrounding responsibility and honesty, with the mere fact that people are not held accountable for their actions or contributions to an online discussion. While use of a pseudonym may allow contribu-
tors the opportunity to come out of their proverbial shells and discuss in an open manner, the reality is many can take advantage of such privacy. Behavior resulting in this privacy might include vulgar comments, a contributor constantly trying to be humorous, disrespect to others’ opinions, or blatant dishonesty. Johnson (1997) contends that integrity is compromised and trust is needed with anonymity. The argument is made that disconnecting the words from the person creates uncertainty and the ability for the contributor to create and socialize using an entirely different persona. Therefore, trust is needed, yet “difficult to develop in an environment in which one cannot be sure of the identities of the people with whom one is communicating” (p. 62). This is a challenge in online data collection. How can trust be built when an in-person meeting hasn’t taken place? There is not any face time to gain perspective on the participant or their contributions.

Friedman, Kahn, and Howe (2000) put it best when discussing the idea of being social in an online environment: “Trust matters” (p. 34). They go on to state, “Common sense tells us that the barriers to trust are least inhibiting when the potential harm is minimal and the good will of the person(s) we trust is genuine” (p. 35). While harm might not include physical harm in an online data collection environment, it can include emotional harm to other participants and harm to the credibility of the researchers if fraudulent information is given. While this type of concern is present in online environments, it would be unwise to assume that all in-person data collection is from trusted and reliable sources. It is important for both types of collection to be proactive in establishing trust by explaining the nature of the study, what it is for, and the voluntary nature that accompanies participation. Establishing trust between the participant(s) and the researcher is also appropriate through introductions and general discussion (based on the type of data collection). Finally, an important area of trust may involve thorough evaluation of the data set for quantitative studies and the responses of qualitative research. Noting the professionalism of the open-ended responses on a questionnaire or the accuracy of a qualitative response can lead to greater or reduced trust in the contribution of the participants.

**Online Focus Groups**

A contemporary trend in online data collection is the use of online focus groups. An online focus group can be defined as a qualitative data-gathering tool that utilizes Internet technology to gain the perceptions, experiences, and opinions of participants. Online focus groups contain some advantages and disadvantages when compared to using traditional in-person focus groups. Advantages of using online focus groups include the ideas that the participants can be questioned over a longer period of time, larger numbers of participants can be obtained, and more intense discussion can occur due to the fact it would not be a face-to-face discussion (Creswell, 2007). Anonymity positively benefits research utilizing online focus groups. In a traditional focus group, making comments in front of other people or having to participate with diverse personalities may limit
contributions of shy or apprehensive participants. Online focus groups present a discussion arena where confrontation or opposing viewpoints are controlled and thought through. In person, it is possible to ignore or respect others’ opinions when the intensity escalates. The online format grants participants the opportunity and time to think about and respond in respectful ways. Intensity can arise in this type of format as comments can be made without being in the same room and viewing and feeling the response from the rest of the group. Intensity can be seen as a positive outcome encouraging legitimate opinion as long as it remains constructive, respectful, and professional.

There are disadvantages specific to the use of online focus groups and anonymity. One issue that can arise is the fact that participants can use anonymity as a way to voice a much stronger and negative voice than would be present in an in-person format. While mentioned as an advantage to have anonymity, some use a lack of identity as a way to hide behind comments. If comments reach a point of emotionally harming people, then the moral values of respect, caring, and love have been violated. On the opposite side of too much or too harsh of a response is the idea that people can provide very little response. Without identity, participants can, in a sense, engage in social loafing (Latane, Williams, & Harkins, 1979; Karau & Williams, 1993) without fear of repercussion or opinion of their input. Social loafing is defined as a “reduction in motivation and effort when individual’s work collectively compared with when they work individually or coactively” (Karau & Williams, 1993, p. 681). This type of problem can be an issue with both in-person and online focus groups, but in an online focus group this concern can be amplified with added barriers and privacy.

Invitation

The final issue of online data collection deals with inviting fans to participate in online surveys, questionnaires, or focus groups. Many studies use message boards, blogs, or websites to recruit potential research participants. This presents issues of permission and bias. In many cases, people go to blogs, message boards, or other social media sites as a hobby or a way to pass time, socialize, or for entertainment. It is easy for researchers to avoid the responsible way of collecting data and simply post a message containing a link to a questionnaire in hopes of gathering many research participants off the particular website. Does this contaminate the website, bother the website visitors, or interfere with the site’s objective? Is permission needed to solicit with this type of ambush style? Another question involves the target population of online data collection. If a researcher targets a particular website or message board to collect data, is the data tainted or biased because of the type of fan consuming that website? Is this the same as collecting data with pencil and paper outside of the mall? Certain people shop at malls and certain people go to websites. This should be considered when the invitation and recruitment of fans takes place online. This is a limitation of most all convenience
sampling techniques. The researcher is choosing the sample based upon convenience or access and this does have limitations. The sample may not be representative of the population as a whole but this an acceptable method of gathering data (Andrew, Pedersen, & McEvoy, 2011).

In all research, there is importance placed on sampling and understanding the demographics of the population in which you are recruiting. In staying with the mall example, a researcher should have knowledge on who is going to the mall. The population could range in age, ethnicity, socioeconomic status, and reasons for being at a mall. In addition, knowing the type of mall, stores, and area of the state or country is important when understanding the population. In the same vein, it is important to know that the Internet, while vast, has specific type of users. Age is one of the biggest differences to consider with online users. In 2010 research, PEW (2010) reported that Internet usage steadily decreases as age of the user increases. It is reported that 95% of Millennials (ages 18-33) are online and using the Internet. Other generation usage is comprised of the following: Generation X (ages 34-45) at 86%, Younger Boomers (ages 46-55) at 81%, Older Boomers (ages 56-64) at 76%, Silent Generation (ages 65-73) at 58%, and G.I. Generation (ages 74 and older) at 30%. Additionally, people go online for various reasons. PEW (2011) reports that 71% of adults go online to watch a video, while 65% use social networking sites, 46% send instant messages, 36% play online games, 32% read blogs, and 4% visit virtual worlds. Online consumers are similar in that they have the ability to use the technology, but as far as reasons for consumption, online consumers are as diverse as the general population in terms of demographics and reason for use.

The data provided in the aforementioned paragraphs illustrates how researchers utilizing online data collection methods encounter many of the sampling challenges that confront traditional survey research. The key to any type of research in trying to understand a phenomenon is asking the right people the right questions (Singletary, 1994). This must be considered when using the Web for research and the site used to gather potential respondents. Posting a link to a questionnaire concerning academic reform in college athletics would garner different viewpoints from visitors to the National Association of Academic Advisors for Athletics than visitors to a college football recruiting site. With that, it is important to try to gain a representative sample or specifically target the population the researcher wants to study. It is important to remember that the same issues involving sampling and inviting participants with traditional questionnaire research must also be considered when using online questionnaires.

**Contacting Respondents**

One of the first steps in data collection is identifying the population under study and the members of the population. This would be typical of any type of survey research whether the data is gathered electronically or through the tradi-
tional means of mail, telephone, or intercept method. Problems facing researchers in regard to electronic surveys are not necessarily incorrect or outdated mailing addresses or invalid phone numbers, but incorrect or invalid e-mail addresses. But the e-mail address issue provides a different challenge.

A researcher targeting a professional population, such as athletic directors at NCAA Division I member institutions, can easily identify the mailing address of the athletic director and mail the questionnaire. A direct telephone number may not be available for the athletic director, but a researcher can call the general athletic department and hopefully make it through the telephone transfers to the athletic director. So contact can eventually be made or at least the researcher knows a way of contacting each member of the population. E-mails can be different, as the athletic director might have more than one active e-mail account.

There can be an address that is publicly known, but it is a generic address, such as athleticdirector@university.edu, most likely monitored by an administrative person who filters out e-mails not considered a priority. This in and of itself creates friction with the ethical value of honesty if the athletic director never even views an e-mail sent to his or her address. Gaining actual contact may be a challenge. The same can be true for mail surveys, as an administrative person may filter out mail that is not considered a priority. Mail in official university envelopes and letterhead would not likely be considered “junk mail” and would make it through the filter to the athletic director. Traditional mail is also not as easy to discard, as it is to delete an e-mail.

Another method of contacting respondents or fans is through posts on message boards inviting users to participate in a research study. This also brings into the equation the issue of intrusion. Participants are viewing and reading message boards for a variety of reasons including socialization and information gathering. Topics on message boards range from entertainment, sports, and agriculture to personal health, religion, and academics. Reasons that sport fans use message boards include the need to express their opinions (Woo, An, & Cho, 2008), communicate with rival fans (End, Eaton, Campbell, Kretschmar, Mueller, & Dietz-Uhler, 2003), bask in reflected glory (End, 2001), or communicate with fellow fantasy sport owners (Ruihley, 2010; Ruihley & Hardin, 2011b). Obviously, being asked to participate in a research study either through an interview or completion of a questionnaire is not a reason fans frequent message boards. With that, a post about academic research can intrude on one’s consumption. This type of data collection can lead to questioning if the researcher is acting responsible with unsolicited invitations and messages.

Attention must be given to whether or not the message board owner, moderator, or operator even wants his or her consumers being solicited. The message board is where the owner’s business occurs, and having posts asking for participants in research intrudes into this business and may distract a consumer from the initial visited site. A traditional business would probably not want someone to en-
Online Data Collection

ter the premises and distribute research recruitment flyers to customers. The best-case scenario in this matter is to take responsibility by familiarizing oneself with the terms and agreements of the message board, as well as contacting the message board owner and asking for permission to post and seek out cooperation in the study. If fortunate, the owner can then post the message and make it a permanent post at the top of the message board and keep it from dropping as more threads are posted. Regardless, the issue of intrusion remains in that it can be viewed similar to that of telemarketers and unwanted telephone calls (Reihl, 2007).

Conclusion

The preceding three areas inquire about collecting fan data in an online environment. The areas of intrusion, interaction, and inviting are examined, and attention is called to the ethical and research implications concerning the values of honesty, responsibility, justice, and beneficence. While practical application suggestions were given in each of the aforementioned sections, the overall theme suggests that it is important for the researcher to realize that the same guidelines set forth for traditional research must still be considered when conducting research in an online environment. IRB approval must be obtained and guidelines set forth for research involving human subjects, including privacy and anonymity, is still applicable whether researchers are dealing with cancer patients, drug users, sport fans, or general population. Online or in person, the fact remains, researchers still must put the other person first with the core moral and ethical values. Similar to other discussions regarding sport ethics, the NCAA, and compliance, the spirit of the rules must be taken into consideration. Just because some IRB offices do not explicitly discuss Twitter, Facebook, or other social media outlets, the spirit of the rules for human subject research should still be considered.

References


