

Secondary Data Analysis: A Method of which the Time Has Come

Melissa P. Johnston, PhD

School of Library and Information Studies, University of Alabama, Tuscaloosa, AL,
USA

Abstract: Technological advances have led to vast amounts of data that has been collected, compiled, and archived, and that is now easily accessible for research. As a result, utilizing existing data for research is becoming more prevalent, and therefore secondary data analysis. While secondary analysis is flexible and can be utilized in several ways, it is also an empirical exercise and a systematic method with procedural and evaluative steps, just as in collecting and evaluating primary data. This paper asserts that secondary data analysis is a viable method to utilize in the process of inquiry when a systematic procedure is followed and presents an illustrative research application utilizing secondary data analysis in library and information science research.

Keywords: secondary data analysis, school librarians, technology integration

1. Introduction

In a time where vast amounts of data are being collected and archived by researchers all over the world, the practicality of utilizing existing data for research is becoming more prevalent (Andrews, Higgins, Andrews, Lalor, 2012; Schutt, 2011; Smith, 2008; Smith et al., 2011). Secondary data analysis is analysis of data that was collected by someone else for another primary purpose. The utilization of this existing data provides a viable option for researchers who may have limited time and resources. Secondary analysis is an empirical exercise that applies the same basic research principles as studies utilizing primary data and has steps to be followed just as any research method. This paper asserts that secondary data analysis is a viable method to utilize in the process of inquiry when a systematic process is followed. This paper contributes to the discussion of secondary data analysis as a research method for library and information science (LIS) and utilizes a study of U.S. school librarians to describe and illustrate the process, benefits, and limitations in conducting an investigation utilizing secondary data analysis method.

Received: 21.4.2013 / Accepted: 12.12.2013
© ISAST

ISSN 2241-1925



2. Defining Secondary Data Analysis

of re- ,5 : 70
p. 11), yet there remains a dearth of literature that specifically addresses the process and challenges of conducting secondary data analysis research (Andrews et al., 2012; Smith, 2008). Hakim (1982) defines secondary analysis

0
conclusions or knowledge additional to, or different from, those presented in the , 25-2

begins with an investigation to learn what is already known and what remains to be learned about a topic through reviewing secondary sources and investigations others have previously conducted in the specified area of interest. Secondary data analysis takes this one step further, including a review of previously collected data in the area of interest. While secondary data analysis is a flexible approach and can be utilized in several ways, it is also an empirical exercise with procedural and evaluative steps, just as there are in collecting and evaluating primary data (Doolan & Froelicher, 2009). Secondary data analysis remains an under-used research technique in many fields, including LIS. Given the increasingly availability of previously collected data to researchers, it is important to further define secondary data analysis as a systematic research method. Yet, few frameworks are available to guide researchers as they conduct secondary data analysis (Andrews et al., 2012; Smith et al., 2011).

3. Process of Secondary Analysis

In conducting research, the area of investigation and the research questions determine the method that the researcher follows. The research method consists of how the researcher collects, analyzes, and interprets the data in the study (Creswell, 2009). Secondary analysis is a systematic method with procedural and evaluative steps, yet there is a lack of literature to define a specific process, therefore this paper proposes a process that begins with the development of the research questions, then the identification of the dataset, and thorough evaluation the dataset. This procedure is illustrated by a LIS research study in which the researcher investigated school librarians as leaders in technology integration.

3.1 Develop the Research Questions

The key to secondary data analysis is to apply theoretical knowledge and conceptual skills to utilize existing data to address the research questions. Hence, the first step in the process is to develop the research questions. The purpose of this study was to investigate the enablers and barriers that school librarians experience enacting a leadership role in technology integration. The research questions that guided this work are: What enablers or supporting factors do accomplished school librarians perceive as enablers in enacting the role of leader in technology integration? What barriers or constraining factors do accomplished school librarians perceive to enacting the role of leader in

technology integration? What is the association between accomplished school librarians involved at a high level in technology integration leadership and the identified enablers/barriers in comparison to the other participants?

3.2 Identifying the Dataset

Most research begins with an investigation to learn what is already known and what remains to be learned about a topic (Creswell, 2009); including related and supporting literature, but one should also consider previously collected data on the topic (Dale, Arbor, & Procter, 1988; Doolan & Froelicher, 2009). Data may already exist that can be utilized in addressing the research questions.

In the case of this research an in-depth literature review of the areas of interest was conducted examining the previous and current work of experts in the field of school librarianship and technology. Through the literature review other researchers on this topic were identified, as were agencies and research centers that have conducted related studies. Recent research and findings from the top ranked school library preparation programs were identified and reviewed, as were dissertations in the areas of technology, leadership, and school librarians. Finally, local informal networks can also provide valuable information in determining what research is currently being conducted (Magee, Lee, Giuliano, & Munro, 2006). This is especially relevant in the field of school librarianship, which is a very connected community. The researcher had the benefit of an informal network, in that she was a part of the team that worked on background research and survey construction for a then current study by the Partnerships for Advancing Library Media (PALM) Center at Florida State University (FSU).

Original survey research rarely uses all of the data collected and this unused data can provide answers or different perspectives to other questions or issues (Heaton, 2008, Johnston, 2012; Smith, 2008), yet the key to using existing survey data effectively to find meaningful answers is a good fit between the research question and the dataset (Doolan & Froelicher, 2009; Kiecolt & Nathan, 1985; Magee et al, 2006). In this study, the research questions fit well with that of the original study since both studies focused on school librarians

2

investigators, made her aware that data collected from questions addressing enablers and barriers to leadership in technology integration had not been analyzed or reported. Finding that this data would adequately address her research questions and that the primary method of data collection was appropriately suited to her research, the decision was made to utilize existing survey data to find the answers to different research questions than were asked in the original research.

3.3 Evaluating the Dataset

Once a dataset that appears viable in addressing initial requirements discussed above is located, the next step in the process is evaluation of the dataset to ensure the appropriateness for the research topic (Dale et al., 1988; Kiecolt & Nathan, 1985; Smith, 2008; Stewart & Kamins, 1993). The advantage is that the

data already exist in some form and can be evaluated for appropriateness and quality in advance of actual use (Stewart & Kamins, 1993). Stewart and Kamins, 57-

(p. 18). The following evaluative steps should be followed in order to determine the appropriate match of a dataset to a research investigation and ensure congruency, quality of the primary study and the resulting dataset: (a) what was the purpose of this study; (b) who was responsible for collecting the information; (c) what information was actually collected; (d) when was the information collected; (e) how was the information obtained; and (f) how consistent is the information obtained from one source with information available from other sources (Stewart & Kamins, 1993). The researcher was given access to and utilized all documentation on the collection of the data, information found in publication, and consulted the investigators from the primary study in order to complete this evaluation.

3.3.1 What was the purpose of this study? It is important to determine the purpose of the original project that produced the data because this can influence many factors such as the targeted population, the sample selected, the wording of questions on the survey, and the general context of the study (Doolan & Froelicher, 2009; Magee et al., 2006). It is also important to know about the agency or individual(s) that collected the information and the similarities or differences in research goals between those researchers and the researcher contemplating secondary analysis (Boslaugh, 2007; Stewart & Kamins, 1993). The researcher had the benefit of a professional relationship with the researchers who conducted the original study and through maintaining contact was privy to inside information about the data collection process. The original study sought to characterize the technology integration activities of school librarian leaders in

C research goal to improve the

goal. However, in this case secondary data analysis allows the researcher to analyze the original dataset to answer a different question; in this case, what is enabling or deterring these practices.

3.3.2 Who was responsible for collecting the information? In addressing the question of who was responsible for collecting the information the secondary researcher again has the benefit of a relationship with the primary research team.

0

backgrounds and previous research projects finding that that the primary investigators are well-respected academic researchers, and have a reputation for excellence in research integrity.

3.3.3 What information was actually collected? It is vital for the secondary researcher to have access to adequate documentation from the primary research, including protocols and procedures followed in the collection of the data (Clarke & Cossette, 2000; Dale et al., 1988; Smith, 2008; Smith et al., 2011; Stewart & Kamins, 1993). The primary research team kept detailed documentation that

provided evidence of careful and consistent data collection. The documentation, the survey instrument, and published findings were consulted, finding that the survey consists of three sections and collected the following data: 30 demographic questions; 60 Likert scale statements on technology integration practices; and three open-ended questions that asked respondents to discuss barriers, enablers, and other factors that influenced their leadership practices (Everhart, Mardis, & Johnston, 2012).

3.3.4 When was the information collected? In any research the time when the data is collected must be considered (Boslaugh, 2007; Stewart & Kamins, 1993). Survey data may be several years old before it is released and available for use by others. In areas related to technology, as in this research, the time frame of data collection is paramount (Boslaugh, 2007). The researcher looked for data related to school librarians as technology leaders that were no more than three years old. The FSU researchers conducted the primary study and collected data during the spring, summer, and fall of 2009, making this data at that time, the most current dealing with this topic.

3.3.5 What methodology was employed in obtaining the data?

of secondary data cannot be evaluated without knowledge of the methodology

secondary researcher must be knowledgeable of the primary method as well. In evaluating the existing data, issues with the survey method including the survey instrument had to be considered. The primary research team developed the *School Librarian Technology Leadership Survey* (PALM, 2009), because no instrumentation existed in this area. One disadvantage of utilizing secondary data is that secondary researchers often have to settle for the original measurement tool and therefore have to evaluate and make a judgment call on the instrumentation (Clarke & Cossette, 2000). Even though the researcher was involved in constructing the survey instrument in the primary research, the literature review and documentation of this process were consulted to address validity and reliability issues (Magee et al., 2006). Finally, in evaluating how the original data was collected, the researcher examined how issues such as sampling, response rates, missing responses, and bias were handled in the original research (Kiecolt & Nathan, 1985).

3.3.6 Management of the primary data. It is mandatory for the secondary analyst researcher to obtain all documentation of the processes and protocols followed by the primary researchers, including the questionnaire, all coding materials, and any publications that are related to the data (Boslaugh, 2007; Clarke & Cossette, 2000; Stewart & Kamins, 1993). Finally, it is paramount that the secondary researcher has access to the raw dataset in order to perform new analyses and to consider and account for all of the aforementioned possible concerns (Boslaugh, 2007; Stewart & Kamins, 1993). The researcher was granted permission to access to the raw dataset and all supporting documentation.

3.3.7 How consistent is the information obtained from one source with information available from other sources? It is beneficial to have multiple sources to bolster confidence in findings, whether it is that two or more sources arrive at the same conclusion for comparison or that they do not, providing an option for contrast. In the case of this research, there was no other similar data located on this very specialized topic, so this analysis was impossible.

4. Discussion: Strengths and Limitations of Secondary Analysis for LIS Research

The major advantages associated with secondary analysis are the cost-effectiveness and convenience it provides (Dale et al., 1988; Glaser, 1962; Smith, 2008). Since someone else has already collected the data, the researcher does not have to devote financial resources to the collection of data. When good secondary data is available, researchers can gain access to and utilize high quality larger datasets, such as those collected by funded studies or agencies that involve larger samples and contain substantial breadth. The larger samples are more representative of the target population and allow for greater validity and more generalizable findings (Smith, 2008; Smith et al., 2011). Access to this type of data presents opportunities for all researchers, even the novice or unfunded researcher, therefore equalizing opportunities and building capacity for empirical research (Hakim, 1982) in LIS research.

The use of existing data sets can accelerate the pace of research because some of the most time consuming steps of a typical research project, such as measurement development and data collection are eliminated (Doolan & Froelicher, 2009). In LIS research areas, such as information and technology that are constantly changing, utilizing existing data allows projects to be completed and findings to be produced much faster, and therefore the development and contribution of new knowledge occurs in a timely manner before they are considered dated by the field. Additionally, in the area of information policy, utilizing existing data can allow the researcher to answer important time-sensitive policy related questions quicker (Magee et al., 2006).

Secondary data analysis provides many opportunities for furthering LIS research through replication, re-analysis and re-interpretation of existing research. It provides researchers with opportunities to engage in work to test new ideas, theories, frameworks, and models of research design.

Yet there are unique methodological considerations when utilizing existing data to investigate new research questions and generate new knowledge. The most recognized limitation to the secondary data analysis method approach is that the data were collected for some other purpose (Boslaugh, 2007, p. 4). Since the data were not collected to answer the specific information that the researcher would like to have may not have been collected; or data may not have been collected in the geographic region of interest, in the years the researcher would have chosen, or on the specific population that is the

focus of interest (Boslaugh, 2007; Doolan & Froelicher, 2009).

In this particular project the researcher avoided some common pitfalls often associated with secondary analysis by participating in the primary research design plan and then ensuring a match between her research questions and the existing data through the previously described process. Yet, a significant limitation of this research was that the school identifiers collected in the primary study were not available to the researcher due to confidentiality reasons. The school identifiers connect to the participants, therefore school identifiers were removed from the dataset, in order to ensure all participants remain anonymous in accordance with the original consent agreement. Therefore, subjects cannot be contacted for follow-up questions and additional data cannot be collected. While this lack of opportunity for follow-up or the collection of additional data from the participants has proven to be a limitation in furthering this research, it is important that secondary data analysis abide by the consent conditions of the original study (Heaton, 2008).

A second major disadvantage of using secondary data is that the secondary researcher did not participate in the data collection process and does not know exactly how it was conducted. Therefore, the secondary researcher does not know how well it was done and if the data are affected by problems such as low response rate or respondent misunderstanding of specific survey questions. Hence the researcher has to find this information through other means such as documentation of the data collection procedures, technical reports, and publications (Boslaugh, 2007; Dale et al., 1988; Kiecolt & Nathan, 1985). In this research the researcher was at a disadvantage because she did not participate in the execution of the data collection process. In order to address these issues the researcher utilized documentation from the original study, information from published findings, and consultations with the original primary researchers and statistician. Ensuring a match between the research question and the existing data and following a process, as proposed, for careful reflective examination and critical evaluation of the data, can avoid most limitations of secondary data analysis.

5. Conclusion

Secondary data analysis offers methodological benefits and can contribute to LIS research through generating new knowledge (Heaton, 2008, Johnston, 2012; Smith, 2008). The overall goal of this method is the same as that of others, to contribute to scientific knowledge through offering an alternate perspective; it only differs in its reliance on existing data. LIS researchers should take advantage of the high quality data that are available and consider the potential value in gaining knowledge and giving insight into a broad range of LIS issues through utilizing secondary data analysis method.

Yet, successful secondary analysis of data requires a systematic process that acknowledges challenges of utilizing existing data and addresses the distinct characteristics of secondary analysis. The process proposed from this

application in LIS research provides a systematic process that includes steps to undertake to avoid possible limitations. In a time where the large amounts of data being collected, compiled, and archived by researchers all over the world are now more easily accessible, the time has definitely come for secondary data analysis as a viable method for LIS research.

References

- Andrews, L., Higgins, A., Andrews, M. W., & Lalor, J. G. (2012). Classic grounded theory to analyse secondary data: Reality and reflections. *The Grounded Theory Review, 11*(1), 12-26.
- Boslaugh, S. (2007). *Secondary analysis for public health: A practical guide*. New York, NY: Cambridge. doi: 10.1017/CBO9780511618802
- Clarke, S. P., & Cossette, S. (2000). Secondary analysis: Theoretical, methodological and practical considerations. *Canadian Journal of Nursing Research, 32*(3), 109-129.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.
- Dale, A., Arbor, S., & Procter, M. (1988). *Doing secondary analysis*. London, UK: Unwin Hyman.
- Doolan, D. M., & Froelicher, E. S. (2009). Using an existing data set to answer new research questions: A methodological review. *Research and Theory for Nursing Practice: An International Journal, 23*(3), 203-215. doi:10.1891/1541-6577.23.3.203
- Everhart, N., Mardis, M., & Johnston, M. P. (2012). National Board Certified school technology integration: Results of a national survey. *School Library Media Research, 14*.
- Glaser, B. G. (1963). Retreading research materials: The use of secondary analysis by the independent researcher. *The American Behavioural Scientist, 6*(10), 11-14.
- Hakim, C. (1982). *Secondary analysis in social research: A guide to data sources and method examples*. London, UK: George Allen & Unwin.
- Heaton, J. (2008). Secondary analysis of qualitative data: An overview. *Historical Social Research, 33*(3), 33-45.
- Johnston, M. P. (2012). School librarians as technology integration leaders: Enablers and barriers to leadership enactment. *School Library Research, 15*(1).
- Kiecolt, K. J., & Nathan, L. E. (1985). Secondary analysis of survey data. *Sage University Paper Series on Quantitative Applications in the Social Sciences, 53*.
- Magge, T., Lee, S. M., Giuliano, K. K., & Munro, B. (2006). Generating new knowledge from existing data: The use of large data sets for nursing research. *Nursing Research, 55*(2), S50-S56.
- Partnerships Advancing Library Media (PALM) Center, Florida State University (2009). *School library media specialist and technology integration survey*. Unpublished
- Smith, E. (2008). *Using secondary data in educational and social research*. New York, NY: McGraw-Hill Education.
- Smith, A. K., Ayanian, J. Z., Covinsky, K. E., Landon, B. E., McCarthy, E. P., Wee, C. C., & Steinman, M. A. (2011). Conducting high-value secondary dataset analysis: An introductory guide and resources. *Journal of General Internal Medicine, 28*(8), 920-929. doi:10.1007/s11606-010-1621-5
- Stewart, D. W., & Kamins, M. A. (1993). *Secondary research: Information sources and methods*. Newbury Park, CA: Sage.