

Data Sources for Scholarly Research: Towards a Guide for Novice Researchers

Timothy J. Ellis and Yair Levy
Nova Southeastern University
Graduate School of Computer and Information Sciences
Fort Lauderdale, FL, USA

ellist@nova.edu, levyy@nova.edu

Abstract

One of the biggest challenges the novice researcher faces is determining just where and how to start her or his research. During the research design stage, a novice researcher must take into consideration three key factors: a) literature; b) research-worthy problem; and c) data. While the role of the problem and literature in research has been explored previously, inadequate attention has been given to the centrality of data and access to collecting data in the context of research design. This paper explores data as a vital element of scholarly enquiry by outlining the role of data in research in the informing sciences, identifying some issues with access to data collection, and their impact on the design of a proposed research. This paper explores the categories of data, organized in a 2x2 taxonomy: the Qualitative-Quantitative-Indirect-Direct (Q²ID) Taxonomy of Data Sources. This paper concludes with examples from literature for some research studies and explanations for the types of data used in the context of the proposed Q²ID Taxonomy of Data Sources are provided.

Keywords: Data Sources, Data Categorization, Qualitative vs. Quantitative Data, Categories of Data, Data as a Research Element, Access to Data Collection, Data Source Taxonomy, Types of Data, Data Measures

Introduction

One of the biggest challenges the novice researcher faces is determining just where and how to start her or his research (Zikmund, Babin, Carr, & Griffin, 2010). Since the essence of research is making a contribution to the body of knowledge, *the literature* is certainly an excellent starting point (Levy & Ellis, 2006). One cannot make a contribution to the body of knowledge without being familiar with that body of knowledge first. Research must also be motivated by some reason beyond the obvious ones associated with meeting the requirements for degree completion, tenure, or promotion; *the problem* motivating the research is likewise a viable starting point (Ellis

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Publisher@InformingScience.org to request redistribution permission.

& Levy, 2008). A third, highly pragmatic factor must also be considered is: the data available to support the research or, more precisely, the researcher's *access and ability to collect data* (Leedy & Ormrod, 2010). The most promising research-worthy problem, indisputably supported by the literature cannot lead to scholarly enquiry if the researcher does not have access to the data necessary to conduct that research. All three

factors – *literature*, *research-worthy problem*, and *access to data* – must be taken into consideration by the novice researcher early in the design stage of her or his study.

The role of the research-worthy problem and literature in research have been explored previously (Ellis & Levy, 2008; Levy & Ellis, 2006). This paper explores the third piece, data, as a vital component of scholarly enquiry. The target audience for this paper is the novice researcher, such as doctoral students or junior faculty members. In the balance of this introduction, context will be established by briefly exploring two essential factors: the nature of research in the informing sciences, and data as an element of research. The second section of the paper expands the definition of data by exploring the different categories of research data. The third section explores potential sources for the different categories of data identified in section two by identifying examples from the literature of the different data sources. Finally, a summary and recommendations are provided.

Research in the Informing Sciences

In his seminal work, Cohen (1999) claimed that informing science is a field of inquiry on the process and infrastructure of “providing a client with information in a form, format, and schedule that maximizes its effectiveness” (p. 215). He outlined three interrelated components of the field: a) the client, b) the delivery system, and c) informing environment. Since informing science represents the nexus of technology, processes, and people, research in that discipline is certainly not monolithic in nature (Gill & Bhattacharjee, 2009). Published enquiries have been conducted from a number of different philosophical perspectives on the meaning and meaningfulness of the results of research. The three most universally accepted perspectives are the positivist, the interpretive, and critical research (Kim, 2003). The *positivist epistemology*, for example, being based on the assumption “that physical and social reality is independent of those who observe it, and that observations of this reality, if unbiased, constitute scientific knowledge” (Gall, Gall, & Borg, 2003, p. 14), asserts that objective “truth” can be derived through research. The *interpretive perspective*, on the other hand, is based on the “assumption that access to reality (given or socially constructed) is only through social constructions such as language” (Myers, 1997, p. 241), asserts that reality can be best understood in the context of the meanings assigned by people. A third perspective, *critical research*, “focuses on the oppositions, conflicts and contradictions in contemporary society, and seeks to be emancipatory i.e. it should help to eliminate the causes of alienation and domination” (Myers, 1997, p. 242).

Regardless of the philosophical perspective underlying the research, all scholarly enquiry has three elements: the applicable, scholarly literature; a research-worthy problem; and data (Creswell, 2005; Leedy & Ormrod, 2010). As detailed in Figure 1, the literature serves as the foundation for research. Research worthy problems are identified and supported through the scholarly literature, and the applicability and validity of the data is established through the literature. The research worthy problem motivates and justifies the research by answering the question “Why is the study being conducted” (Ellis & Levy, 2008). The data serves the dual function of limiting and enabling the research (Zikmund et al., 2010). The type of study possible is both indicated and restricted by the data available to the researcher. Data obviously serves a central function in research. However, just what constitutes data is less obvious. The text below will attempt to address that question.

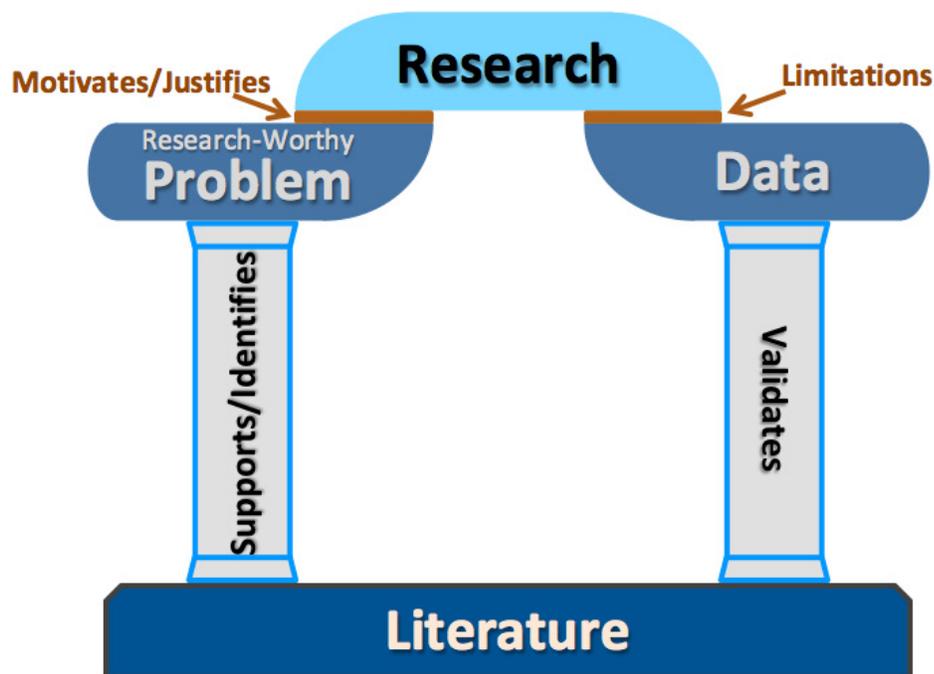


Figure 1: The Role of Data in the Structure of Research

Data as an Element of Research

The manner in which “data” is defined is of critical importance (Mertler & Vannatta, 2010). According to Leedy and Ormrod (2010), “the term *data* is plural (singular is *datum*) and derived from the past participle of the Latin verb *dare*, which means “to give” (p. 94). Too narrow a definition, such as restricting data to only those things that can be measured with numerical precision, can restrict the meaning of research in informing science to such an extent that few, if any, outside academe would find the results of value. On the other hand, too broad a definition, such as including in “data” anything that one purports to observe, could expand the meaning of research to such an extent that any assertion one makes could be categorized “research”. A well-formed definition of data is essential to a meaningful understanding of research.

Although the importance of data to research is well accepted, there is not a universally accepted definition of data. Many resources that detail research methods devote chapters on how to acquire and analyze data without ever describing and delimiting the term (Gall et al., 2003; Richey & Klein, 2007; Sekaran, 2003). When the term has been defined, a number of approaches have been followed. Gay, Mill, and Airaisian (2006) adopted an operational approach to the definition by indicating: “Data are the pieces of information you collect and use to examine your topic, hypotheses, or observations” (p. 122). Leedy and Ormrod (2010), on the other hand, defined data from a more functional approach: “Data are those pieces of information that any particular situation gives to an observer” (p. 94). Zikmund et al. (2010) defined data as “facts or recorded measures of certain phenomena (things or events)” (p. 19). The classical database definition for data, reiterated by Elmasri and Navathe (2011) is data are “known facts that can be recorded and that have implicit meaning” (p. 4), implying a descriptive approach. Although all of these definitions offer insight to the meaning of data in a scholarly enquiry environment, none seem to adequately capture the role of data in research. Thus, in this paper, *data* will be defined as *a purposive collection of perceived facts*. Obviously, that definition needs some refinement. Toward that end, the expression is deconstructed as follows:

Purposive: All that can be observed or otherwise sensed is not necessarily “data”. An observation or other sensation is data only within the context of its use. For example, an observation of ethnicity was “data” in the United States during the 1950s in the context of eliminating certain racial groups from employment opportunities. In the 1970s that same observation was “data” in the United States in the context of providing favored employment consideration to those same racial groups. In scholarly research, the literature is the primary resource for attributing the context necessary to change observations and sensations to data.

Collection: Despite common usage to the contrary, grammatically, the word “data” is plural, not singular. That distinction is not of only grammatical importance, however. A single observation, no matter how purposive and contextualized, does not constitute data. Data must be comprised of a set of related observations. Trying to draw conclusions from a single observation would be analogous to draw a line with only a single known point.

Perceived facts: It is important to remember that data is not equal to facts or, by extension, “truth”. The accuracy of data can be negatively impacted by collection errors such as trying to determine the average height of the residents of a city by measuring the first 20 people you see leaving a grade school building. The accuracy of the data can also be negatively impacted by measurement errors such as using the wrong instrument to measure an observation (i.e. using an IQ test to measure academic achievement) or using the correct instrument incorrectly. Of greater impact than either collection or measurement errors, however, is the fact that data are based on an observation of phenomena, not the phenomena themselves. Data can point toward reality, but should never be confused with reality. A simple example to illustrate this point can be a measure of body weight. Although one might think that the numerical value displayed by the scale is the absolute “truth”, although that value might be close to the “truth”, it can never be regarded as such. Instrument errors (i.e. lack of calibration) and instrument rounding (i.e. a weight that can provide numerical data to the ½ of pound, another to the 1/10 of a pound, another to the 1/100 of a pound) preclude observation of the “truth”.

Access to Data Collection

A research study, as previously indicated, is based on data that the researcher sets forth in order to provide evidence supporting the conclusions of the study (Zikmund et al., 2010). While there are theoretical and anecdotal scholarly literature pieces, research, by definition, is unique in that it is an endeavor that must use data in order to provide evidence to the theoretical (Ellis & Levy, 2008). One of the great challenges for many researchers in the design stage of their study is to secure access to data (Zikmund et al., 2010). In this paper, access to data collection is defined as the ability of the researcher to secure ways to obtain data for the purpose of his or her proposed study. Such access is a requirement that all novice researchers must consider as early in their research design as possible (Ellis & Levy, 2008), and will impact the type of study proposed (Ellis & Levy, 2009). Although in some research institutions or other unique contexts the novice researcher might be given access to data, not all are so fortunate.

Access to a solid source of the data necessary to conduct the proposed research is a major challenge facing the novice researcher. A common example of a case where access to data is not viably solid may include a proposal where the novice researcher wishes to investigate the role of Information Technology (IT) investments in organizations and how they relate to security breach incidents within that organization. The novice researcher’s proposed approach of access to data was by using mail (electronic & regular) to send a survey to all Chief Information Officers (CIOs) of Fortune™500 companies. While the intent of the research might have been interesting, access to the necessary data is not viably solid in this case as the possibility of receiving a meaningful number of participants in such proposed data collection is slim. Certainly, such access to data might have been considered viable if the novice researcher had the personal ability to ask For-

tune™500 companies' CIOs to take part in the research on the basis of being an executive in a company that provides information security (InfoSec) consulting to a large number of the Fortune™500 companies.

Categories of Data

Data can, of course, be categorized in a number of ways. Two dimensions for categorizing data – *proximity* and *precision* – are particularly useful in the context of scholarly research. For the purposes of this paper, *proximity* is defined as the degree of separation between the actual phenomena of interest and the method in which it is observed and measured. *Precision*, again in this paper, refers to the degree to which the value of the data can be objectively represented. Both proximity and precision can be subdivided into two levels: for proximity, direct and indirect measures; for precision, qualitative and quantitative data. Since data have both proximity and precision characteristics, there are four subcategories of data. Figure 2 provides an overview of the 2x2 Qualitative-Quantitative-Indirect-Direct (Q²ID) Taxonomy of Data Sources. The four proposed subcategories include: a) direct measure for qualitative data (DiQual); b) direct measure for quantitative data (DiQuant); c) indirect measure for qualitative data (InQual); and d) indirect measure for quantitative data (InQuant). While these are the four main subcategories of measuring data, a research study is often made more rigorous when the two categories of precision of data, qualitative and quantitative, are both included in a 'mixed methods' study.

		Precision of the Data Type	
		Qualitative (Non-Numerical Data)	Quantitative (Numerical Data)
Proximity of the Measure	Direct	Historical non-numerical records Open-ended questionnaires Interviews Direct observations DiQual	Historical numerical records Surveys Numerical simulations Test results DiQuant
	Indirect	InQual Open-ended questionnaires Interviews Observations	InQuant Surveys Numerical simulations Test results

Figure 2: The Q²ID Taxonomy of Data Sources

Proximity: Direct vs. Indirect Data

Direct data are derived from direct observations of the phenomena of interest. Indirect data are derived from indirect observations; observations of representations of the phenomena rather than the phenomena itself. Both direct and indirect data are of great value to research in information systems and informing science. Some examples include:

- In human-computer interaction research, the cognitive walkthrough would produce direct data while the heuristic evaluation produce indirect data.
- In end-user computing skills research, logs of actual system use would produce direct data while a survey based on the perceived skills, produces indirect data.

- In education research, direct observation of a student's performance in multiple settings over time would produce direct data while a standardized test produce indirect data.

It is important to note that both direct and indirect data present benefits and challenges (Hair, Black, Babin, & Anderson, 2010). Direct data, being based on direct observation of the phenomena of interest, are often richer and accurately describes the phenomena than indirect data. Direct data often have a much greater degree of internal validity than indirect data. On the other hand, direct data are often much more difficult to collect than indirect data and more subject to the skill of the one collecting the data. Furthermore, the very richness of the data makes direct data more related to a specific instance of phenomena than indirect data. Direct data, as a result, often have a smaller degree of external validity (generalizability) and reliability than indirect data.

To illustrate the strengths and weakness of direct versus indirect data, consider the task of measuring human intelligence. One way to determine just how intelligent someone is would be to observe that individual in a number of different situations over an extended period of time – to gather direct data. Those data would be very rich and representative of the individual's capabilities – they would be internally valid. The data would, however, be very dependent on the interpretation the observer. Since two different observers, or even the same observer at different times, might view the same activity of the subject in entirely different ways, those data might not be reliable. On the other hand, one could use a standardized measure of intelligence such as the Wechsler Intelligence Scale. The questions regarding the “fairness” and capacity of such instruments to actually measure human intelligence are well documented in both the scholarly and popular literature – the internal validity of the data so derived would not be as strong as that derived from extensive close observation over time. Since, however, administration of the instrument is standardized and well documented, it is quite probable that essentially the same results would be derived regardless of who administered the test (providing, of course that the documented procedures are followed). The data derived from the standardized test would likely be more reliable than that derived from observation over time.

The discussion of direct versus indirect data should not, however, be in terms of which are the better data for research. If a researcher has access to both direct and indirect data, the strength of the study would be increased significantly if **both** were included. In many instances, the researcher will have access to only direct data or only indirect data. It is important in those circumstances to recognize the limitations inherent in whichever form of data is available for the research.

Precision: Quantitative vs. Qualitative Data

The term “precision” is perhaps misleading in drawing the distinction between quantitative and qualitative data. As mentioned above, “precision” is used in this context to indicate the degree to which the data can be objectively represented, *not* the accuracy, validity, or usefulness of the data. Simply put, quantitative data are expressed in terms of numbers while qualitative data are expressed in terms of words (Sekaran, 2003). Although quantitative data are considered sometimes to be more precise, they are not necessarily more meaningful than qualitative data, depending on the goals of the proposed study. For example, if one were to ask ‘How big is an aircraft carrier?’, the answer using quantitative data would be “1088 feet” while the answer using qualitative data would be “More than three and a half football fields”; the more useful answer would be dependent on the context.

Quantitative data, as the name implies, are numerical in nature and, depending on the type of data, can be analyzed with various mathematical and statistical procedures. There are four types of quantitative data. Since each level permits a different range of statistical tests, it is imperative to correctly identify the type of quantitative data available. The Academic Technology Services

department at the University of California Los Angeles provides an overview of the statistical analyses appropriate for each level of quantitative data. That overview is summarized in Table 1 and can be viewed in its entirety at http://www.ats.ucla.edu/stat/mult_pkg/w/hatstat/.

Nominal data, also known as categorical data, classifies the phenomena observed into two or more groupings (Gay, Mills, & Airasian, 2006). Nominal data can be counted, but cannot be further analyzed mathematically or statistically. For example, if one were conducting research in e-commerce and gathered the state of residence for each participant in the study, one could count the number from Massachusetts and from Florida, determine which number was larger, and otherwise describe the data but could make no inferences regarding the significance of those differences. Other examples of nominal data could include gender (i.e. Male/Female), ethnic background (i.e. Caucasian/African-American/Hispanic/Oriental/Other), and job level (i.e. Executive/Management/Administrative/Production/Other).

Ordinal data classifies the instances of the phenomena being observed into rank order. The Fortune™ 500 ranking of a company is an example of ordinal data. Ordinal data can tell the observer that one instance of a phenomenon is in some aspect greater than another instance of that phenomenon, but it cannot tell you in any meaningful sense just how great that difference is. For example, the difference between the number *one* and number *two* companies in the Fortune™ 500 might or might not be equal to the difference between the number *two* and number *three* companies. The data produced from the Likert-type scale is a commonly seen example of ordinal data used in scholarly research (Jamison, 2004). Although numbers are often associated with the various values used on the classical Likert-type scale such as “Strongly Agree”, and “Agree”, Jamison (2004) indicated that “it is ‘illegitimate’ to infer that the intensity of feeling between ‘strongly disagree’ and ‘disagree’ is equivalent to the intensity of feeling between other consecutive categories on the Likert scale” (p. 1217).

Interval data is similar to ordinal data in that the numbers represent meaningful points of comparison. However unlike ordinal data, the difference between values in interval data also is meaningful. As mentioned above, the difference between a value of *one* and *two* is not necessarily the same as the difference between a value of *two* and *three* with ordinal data; with interval data, those differences are assumed equal. An example of interval data in scholarly research would be scores on an intelligence test; the difference between the score of 90 and the score of 100 represents the same value as the difference between the score of 100 and 110.

Ratio data has the same characteristics as interval data, with one important addition; unlike interval data, ratio data has a true zero. In the example of the intelligence test scores mentioned above, there is not a true zero in that a score of zero on the test does not mean that there is a complete absence of intelligence. With ratio data, on the other hand, a score of zero indicates the complete absence of the phenomenon being observed. Examples of ratio data used in scholarly research include counts of the number of click-through on an e-commerce site or number of times a knowledge base has been accessed.

Table 1: Examples of statistical tests by data type

Type of Data	Statistical Tests
Nominal	Chi-Square (χ^2) McNemar test
Ordinal	Ordinal Regression Spearman’s rho Mann-Whitney U Test Wilcoxon T Path Analysis Factor Analysis
Interval and Ratio	t-Test ANOVA Pearson’s r Beta (β) correlation Multiple Regression

Quantitative data can be distinguished along a second dimension: discrete versus continuous data (Leedy & Ormrod, 2010). Discrete data “has a small and finite number of possible values” (p. 261), whereas continuous data “reflects an infinite number of possible values falling along a particular continuum” (p. 261). An example of discrete data would be number of children. Although there is no theoretical limit to that value, there is a practical limit and there certainly is the limitation to whole numbers. An example of continuous data would be amount of time to access a Web site; there would not be either a theoretical or practical limit to the range of values, and fractional values with infinite precision would certainly be possible. All four levels of data – nominal, ordinal, interval, and ratio – can be either continuous or discrete (Creswell, 2005).

Qualitative data are narrative in format and, consequently, inherently subjective in nature. The purpose of qualitative data is to describe, not measure, the phenomenon of interest (Gay et al., 2006; Sekaran, 2003). Unlike quantitative data that has an objective meaning, the meaning of qualitative data cannot be divorced from the context in which it is collected. Included in that context are the researcher as well as his or her background, perspectives, capabilities, and personal biases.

Since qualitative data are not numeric, statistical tests are of no use in interpretation. The basic process of the analysis of qualitative data is one of organizing and categorizing, identifying patterns and synthesizing to create a narrative that describes the phenomenon of interest. Specific processes for analyzing and interpreting qualitative extend beyond the scope of this article; there are a number of texts that detail processes for working with qualitative data (Gay et al., 2006; Miles & Huberman, 1984). Computerized tools to preform qualitative data analysis, such as Atlas.ti[®] (<http://www.atlasti.com/>), MAXQDA[®] (<http://www.maxqda.com/>), or Ethnograph[®] (<http://www.qualisresearch.com/>), are available, but their use lies beyond the scope of this article.

Examples of Data Sources from Literature

The scholarly literature includes numerous examples of research that utilized the various subcategories of data sources discussed previously. In this section, a few selected studies are provided to illustrate how data sources shaped the research study, the goals of the research, and how the specific subcategory of data source (or mixture of data sources) enabled the researchers to achieve those research goals. Please note these studies are provided to illustrate the data subcategories in the Q²ID Taxonomy of Data Sources described above and should not be considered as the model studies in each subcategory.

Quantitative

Example of direct-quantitative (DiQant) measure

One common approach to collecting direct-quantitative data is with the use of quantitative surveys on direct measures. One of many examples for this direct-quantitative data collection is documented in the study by Gafni and Geri (2010) who measured the role of mandatory versus voluntary tasks and gender differences in task procrastination. As part of their study, they collected the actual submission data extracted from the submission system including the dates of submission and proximity to the deadline. Their directly measured data was based on two groups: a) participants who voluntarily submitted their files, while not being required; versus b) participants who were required to submit their files to the system. Their findings indicate that in general, when the task is non-required, participants tend to procrastinate significantly more than when it's required. They found no gender differences on procrastination in their data.

Example of indirect-quantitative (InQuant) measure

Some studies that collected quantitative data may mix direct and indirect measures under a single study. As such, although Koh, Prybutok, Ryan, and Wu (2010) assessed directly via survey from 330 end-users' perceived *information quality*, *information satisfaction*, *attitude*, *intention to use a system*, some of the other measures used were measured indirectly on a quantitative data. Specifically, they measured the constructs of *performance expectancy* (i.e. the degree that a system can help the user to accomplish tasks), social influence (i.e. the degree of other people's influence to use the system), and net benefit (i.e. the degree that the system provides positive benefit to the organization).

Qualitative

Examples of direct-qualitative (DiQual) measure

Some phenomena related to a viable research problem identified may not lend itself to measures that are numerical in nature (i.e. quantitative). As such, the use of qualitative measures can provide insight. An example of direct-qualitative measure is documented by Gefen, Ragowsky, Licker, and Stern (2011), who interviewed Chief Information Officers (CIOs) about issues that bother them and how they address these issues. A roundtable approach was taken to perform discussions on the issue of outsourcing challenges and how the CIOs address these challenges. The discussions were video recorded and transcribed. While the issues discussed among the CIOs were on the nature of outsourcing, specific data were collected by the researchers on the penetration of outsourced IT services the participating organizations had, compared to the in-house IT services. Moreover, they collected data on the challenges encountered as a result of outsourcing and ways to address such challenges were discussed. The changes that occurred over the past few years as a result of such challenges were also captured from the CIOs, while some trends were observed on the reduced outsourcing in the recent years of IT services (Gefen et al., 2011).

Another study that used direct-qualitative measures is the research done by Jones and Alony (2011), who investigated knowledge sharing in the Australian film industry. Specifically, they used a grounded theory (GT) approach to seek the key factors of knowledge sharing in such a context. The primary data collection method was long structured interviews, ranging from one to two hours. The data resulting from each interview was immediately coded and transcribed. While in this example, the key factors influencing knowledge sharing might be difficult to quantify, a qualitative approach by asking participants to explicitly verbalize their reasons for sharing knowledge may lend itself better.

Examples of indirect-qualitative (InQual) measure

Research problems may pose issues that lead researchers into a great challenge of measuring constructs related to certain phenomena directly. This difficulty is especially true in the event that the researcher wishes to collect data related to controversial issues such as misuse, misconduct, violations of organizational, national, or international rules, to name a few. In some instances researchers will be forced to indirectly measure data related to the phenomena based on perceptions and interpretations of individuals about the phenomena that can help achieve sufficient discoveries. One example for the use of such indirect-qualitative measure includes the investigation of moral issues related to digital piracy. Bhal and Leekha (2008) conducted a study on such issue using "open-ended, in-depth interviews" (p. 638) with 47 employees of six organizations. All participating companies were in the software development (development, customization, sales, etc.) industry. Bhal and Leekha (2008) solicited participants' feedback on the moral logic behind engagement in digital piracy, followed by interviews seeking to assess participants' interpretations if digital piracy is ethical or unethical. They finished the interviews with more narrowly focused

questions about reasons participants were considering digital piracy as ethical or unethical. After the interviews all information gathered was coded and a second round of similar interviews were conducted with 38 of the participants. This round was used to validate the researchers' findings (Bhal & Leekha, 2008). Out of the 38 validated interviews (see additional information about the process to ensure validation in qualitative interviews in Bhal and Leekha (2008)), 17 reported finding digital piracy as ethical and 21 as unethical. While measuring ethical and moral logic appears to be challenging, Bhal and Leekha (2008) were using indirect-qualitative measures to provide insight into such investigation.

Another example is based on the work of Hunter, Evans, and Price (2011), who conducted a study to assess corporate intellectual assets and their value to the organization. They conducted interviews with a group of 10 senior managers that provided comments and their interpretations about such measures. As a direct measure of intellectual assets and their value to the organization is nearly impossible to measure, such long interviews with the managers and their interpretations provide the next best indirect measure of these constructs. While researchers' bias is a significant concern for any type of research, in qualitative research, a stronger emphasis should be given to ensure such bias is controlled to the best possible way, especially when the number of participants is small.

Conclusion and Discussions

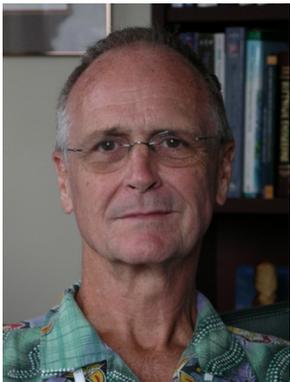
This paper presented a discussion about the role of data as a key element of research in any field of scholarly inquiry. The primary audience for this paper includes doctoral students, junior faculty members, or other novice researchers who wish to better understand the centrality of data during the embarkation of a research study. The attempt of this paper is to help novice researchers and doctoral students to better understand data, challenges with access to collecting it, and its different categories. In this paper, a definition of the term data is provided with additional elaboration of the components of such definition. Access to data collection is then discussed, while emphasizing the challenges associated with such important piece early in the design of any research study. Following, a discussion was provided about the 2x2 Q²ID Taxonomy of Data Sources, followed by few selected examples from informing science literature were provided.

References

- Bhal, K. T., & Leekha, N. D. (2008). Exploring cognitive moral logics using grounded theory: The case of software piracy. *Journal of Business Ethics*, 81(4), 635–646. doi: 10.1007/s10551-007-9537-7
- Cohen, E. (1999). Reconceptualizing information systems as a field of the transdiscipline informing science: From ugly duckling to swan. *Journal of Computing and Information Technology*, 7(3), 213-219.
- Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Upper Saddle River, NJ: Pearson.
- Ellis, T. J., & Levy, Y. (2008). A framework of problem-based research: A guide for novice researchers on the development of a research-worthy problem. *Informing Science Journal*, 11, 17-33. Retrieved from <http://www.inform.nu/Articles/Vol11/ISJv11p017-033Ellis486.pdf>
- Ellis, T. J., & Levy, Y. (2009). Towards a guide for novice researchers on research methodology: Review and proposed methods. *Issues in Informing Science and Information Technology*, 6, 323-337. Retrieved from <http://iisit.org/Vol6/IISITv6p323-337Ellis663.pdf>
- Gafni, R., & Geri, N. (2010). Time management: Procrastination tendency in individual and collaborative tasks. *Interdisciplinary Journal of Information, Knowledge, and Management*, 5, 115-125. Retrieved from <http://www.ijikm.org/Volume5/IJIKMv5p115-125Gafni448.pdf>

- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Boston: Allyn & Bacon.
- Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational research: Competencies for analysis and applications* (8th ed.). Upper Saddle River, NJ: Pearson.
- Gefen, D., Ragowsky, A., Licker, P., & Stern, M. (2011). The changing role of the CIO in the world of outsourcing: Lessons learned from a cio roundtable. *Communications of the Association for Information Systems*, 28(1), 233-242.
- Gill, G. T., & Bhattacharjee, A. (2009). Whom are we informing? Issues and recommendations for mis research from an informing sciences perspective. *MIS Quarterly*, 33(2), 217-235.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hunter, M. G., Evans, N., & Price, J. (2011). Internal intellectual assets: A management interpretation. *Journal of Information, Information Technology, and Organizations*, 6, 1-13.
- Jamison, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38, 1217-1218.
- Jones, M., & Alony, I. (2011). Guiding the use of grounded theory in doctoral studies – An example from the Australian film industry. *International Journal of Doctoral Studies*, 6, 95-114. Retrieved from <http://ijds.org/Volume6/IJDSv6p095-114Jones322.pdf>
- Kim, S. (2003). Research paradigms in organizational learning and performance: Competing modes of inquiry. *Information Technology, Learning, and Performance Journal*, 21(1), 9-18.
- Koh, C. E., Prybutok, V. R., Ryan, S. D., & Wu, Y. A. (2010). A model for mandatory use of software technologies: An integrative approach by applying multiple levels of abstraction of informing science. *Informing Science Journal*, 13, 177 - 203. Retrieved from <http://www.inform.nu/Articles/Vol13/ISJv13p177-203Koh561.pdf>
- Leedy, P. D., & Ormrod, J. E. (2010). *Practical research: Planning and design* (9th ed.). Upper Saddle River, NJ: Prentice Hall.
- Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science Journal*, 9, 181-212. Retrieved from <http://www.inform.nu/Articles/Vol9/V9p181-212Levy99.pdf>
- Mertler, C. A., & Vannatta, R. A. (2010). *Advanced and multivariate statistical methods: Practical application and interpretation* (4th ed.). Los Angeles, CA: Pyrczak Publishing.
- Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage.
- Myers, M. D. (1997). Qualitative research in information systems. *MIS Quarterly*, 21(2), 241 - 242.
- Richey, R. C., & Klein, J. D. (2007). *Design and development research*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Sekaran, U. (2003). *Research methods for business* (4th ed.). Hoboken, NJ: John Wiley & Sons.
- Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2010). *Business research methods* (8th ed.). Mason, HO: Cengage Learning.

Biographies



Dr. Timothy Ellis obtained a B.S. degree in History from Bradley University, an M.A. in Rehabilitation Counseling from Southern Illinois University, a C.A.G.S. in Rehabilitation Administration from Northeastern University, and a Ph.D. in Computing Technology in Education from Nova Southeastern University. He joined NSU as Assistant Professor in 1999 and currently teaches computer technology courses at both the Masters and Ph.D. level in the School of Computer and Information Sciences. Prior to joining NSU, he was on the faculty at Fisher College in the Computer Technology department and, prior to that, was a Systems Engineer for Tandy Business Products. His research interests include: multimedia, distance education, and adult learning. He has published in several technical and educational journals including Catalyst, Journal of Instructional Delivery Systems, and Journal of Instructional Multimedia and Hypermedia. His email address is ellist@nova.edu. His main website is located at <http://www.scis.nova.edu/~ellist/>



Dr. Yair Levy is an associate professor at the Graduate School of Computer and Information Sciences at Nova Southeastern University. During the mid to late 1990s, he assisted NASA to develop e-learning systems. He earned his Bachelor's degree in Aerospace Engineering from the Technion (Israel Institute of Technology). He received his MBA with MIS concentration and Ph.D. in Management Information Systems from Florida International University. His current research interests include cognitive value of IS, of online learning systems, effectiveness of IS, and cognitive aspects of IS. Dr. Levy is the author of the book "Assessing the Value of e-Learning systems." His research publications appear in the IS journals, conference proceedings, invited book chapters, and encyclopedias. Additionally, he chaired and co-chaired multiple sessions/tracks in recognized conferences. Between 2006 and 2011, Dr. Levy served as the founding Editor-in-Chief of the International Journal of Doctoral Studies (IJDS). Additionally, he is serving as an associate editor for the International Journal of Web-based Learning and Teaching Technologies (IJWLTT). Moreover, he is serving as a member of editorial review or advisory board of several refereed journals. Additionally, Dr. Levy has been serving as a referee research reviewer for numerous national and international scientific journals, conference proceedings, as well as MIS and Information Security textbooks. He is also a frequent speaker at national and international meetings on MIS and online learning topics. To find out more about Dr. Levy, please visit his site: <http://scis.nova.edu/~levyy/>