Addressing data collection challenges in ICT for development projects

Devendra Potnis

University of Tennessee, Knoxville, dpotnis@utk.edu

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

Part of the Library and Information Science Commons

Recommended Citation


This Article is brought to you for free and open access by the School of Information Sciences at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in School of Information Sciences -- Faculty Publications and Other Works by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.
ABSTRACT

This paper equips researchers for addressing a wide range of data collection challenges experienced when interacting with marginalized communities as part of ICT4D projects in developing countries. This secondary research categorizes data collection challenges reported in multiple disciplines, and summarizes the guidance from the past literature to deal with the challenges. The open, axial, and selective coding of data collection challenges reported by the past literature suggests that it is necessary to manage scope, time, cost, quality, human resources, communication, and risks for addressing the data collection challenges. This paper illustrates the ways to manage these seven dimensions using (a) the success stories of data collection in the past, (b) the lessons learned by researchers during data collection as documented by the past literature, and (c) the advice they offer for collection data from marginalized communities in developing countries.

Keywords: Communication Management, Cost Management, Data Collection Challenges, Human Resource Management, ICT4D Projects, Quality Management, Risk Management, Scope Management, Time Management

INTRODUCTION

Problem Statement

A significant number of projects using information and communication technologies for the development of marginalized communities (ICT4D) fail every year to meet their goals (Avgerou & Walsham, 2000; Bailur, 2007; Bhatnagar & Odedra, 1992; Heeks, 2002). For instance, during the fiscal period 2003-10, the World Bank invested around $4.2 billion to support its ICT4D projects that promoted access to and adoption of ICT across all sectors in the developing world. In 2011, the Independent Evaluation Group of the World Bank revealed that 70% of its ICT4D projects failed to achieve their goals (World Bank, 2011).

Studies show that data collection with marginalized communities is one of the main reasons for the failure of ICT4D projects in developing countries (Krishna & Walsham, 2005; Mamba & Isabirye, 2014; Qureshi, 2008). For instance, due to the inability of researchers to address data collection challenges, a majority of ICT4D projects in developing countries fail to sustain...
(i.e. remain continuously operational), scale (i.e. spread, enhanced, scoped, and enlarged heterogeneous networks of technology, people, processes, and institutional contexts), or benefit marginalized communities without undesired outcomes (Heeks, 2002; Krishna & Walsham, 2005; Sahay & Walsham, 2006).

Data collection serves as a medium to understand and interpret various dimensions of the relationship between ICT and marginalized communities. It requires researchers to collect and make sense of the needs and experiences of marginalized communities along with other contextual factors for: (a) designing ICT solutions to address the issues related to the development of marginalized communities (Parikh et al., 2003; Sahay & Walsham, 2006), (b) testing prototypes of ICT solutions with marginalized communities (Heimerl et al., 2010), (c) deploying ICT solutions in marginalized communities (Brewer et al., 2006; McCallum & Papandrea, 2009), or (d) assessing the impact of ICT solutions on the social, economic, and human development of marginalized communities (Mohan et al., 2013; Souter et al., 2005). Since data collection attempts to understand someone else’s experience, researchers’ outsider status from marginalized communities’ point of view may challenge researchers’ effort to understand and interpret the relationship between marginalized communities and ICT solutions (Potnis, 2014).

There is a lot of useful guidance available to address data collection challenges in ICT4D projects but it is scattered across multiple disciplines. As a result, for instance, researchers from technology-related disciplines engaged in ICT4D projects may not benefit from the experience and advice of researchers from non-technology disciplines, and vice-a-versa. It is necessary to systematically organize and summarize the scattered guidance so that researchers from all disciplines could avoid unexpected challenges and manage risks during data collection in ICT4D projects.

**Objective of the Paper**

This paper provides a framework to organize data collection challenges and manage ways to address the challenges reported by the past literature on ICT4D projects. This framework is based on the analysis of a wide range of data collection challenges interpreted and reported using perspectives in communication (e.g., Chib & Harris, 2012; McCallum & Papandrea, 2009), computer science (e.g., Brewer et al., 2006), education (e.g., Lima & Brown, 2007), human–computer interaction (HCI) (e.g., Medhi & Toyama, 2007), information science (e.g., Heeks, 2002; Madon, 1997), information systems (e.g., Krishna & Walsham, 2005; Latifov & Sahay, 2013), public administration (e.g., Heeks & Bhatnagar, 1999; Wang & Chen, 2010), and statistics (e.g., Elahi, 2008). The analysis of the challenges suggests that the inability of the outside researchers to manage scope, time, cost, quality, human resources, communication, and risks during their interaction with marginalized communities creates data collection challenges. This paper summarizes the key advice from multiple disciplines by illustrating ways to manage these seven dimensions for addressing data collection challenges.

The next section synthesizes the multidisciplinary scattered guidance available for collecting data from marginalized communities and points out shortcomings in the guidance for addressing data collection challenges in ICT4D projects in developing countries. The Methods section provides rationale for selecting 380 artifacts (past literature) from academic and non-academic sources and the three types of projects (reported by the past literature) considered for data analysis using grounded theory principles. Drawing on the ICT4D projects successful in addressing data collection challenges in the past, lessons learnt by researchers, and their advice for conducting data collection, the next section illustrates a few ways to manage the seven dimensions. The concluding section highlights the contribution of this paper in better equipping researchers for collecting data from marginalized communities in developing countries.
RELATED WORK

A number of studies published by top-tier ICT4D journals, including *IT for Development* (e.g., Krauss, 2013; Krishna & Walsham, 2005; Walsham & Sahay, 2006), *IT and International Development* (e.g., Anokwa et al., 2009; Medhi & Toyama, 2007), *Electronic Journal of Information Systems in Developing Countries* (e.g., Gomez, 2013; Touray et al., 2013), and *International Journal of ICT and Human Development* (e.g., Mathur & Sharma, 2009; Rahman & Ramos, 2013), and books or book chapters (e.g., Chib & Harris, 2012; De, 2012; Vaidya et al., 2013) discuss ICT4D project challenges at great length and offer guidance to overcome the challenges. This literature dedicated to ICT4D projects aims to (a) report “what went wrong?”, (b) increase our understanding of data collection challenges, (c) seek and analyze causality for the challenges, (d) emancipate marginalized communities in developing communities, or (e) educate others for devising better practices to collect data in the future.

Scattered Guidance for Addressing Data Collection Challenges

There is a lot of useful guidance available in the form of precautions to be taken to avoid data collection challenges in ICT4D projects and solutions to overcome the challenges. However, the existing guidance is scattered across multiple disciplines (Dodson et al., 2012). Importantly, due to disciplinary silos, researchers from different disciplines are not aware of, and hence, do not benefit from each other’s experience and advice for addressing data collection challenges in ICT4D projects. For instance, information scientists cluster 43 barriers into eight possible critical factors which affect data collection in ICT4D projects in developing countries (Touray et al., 2012). The factors are: political and leadership, socio-cultural, infrastructural, technical, educational, and skills, economical, security and safety, and legal and regulatory. Whereas Richard Heeks, a development informatics researcher, proposes concept-reality gaps in information, technology, process, objectives and values, staffing and skills, management and structures, and other resources like time, to address data collection challenges in ICT4D projects (Heeks, 2009). On the other hand, management experts engaged in ICT4D projects find committed and knowledgeable leadership to be a key factor for managing data collection challenges in ICT4D projects in developing countries; they advise project leaders and managers to persistently convey their vision to project staff members as part of the effective leadership (Krishna & Walsham, 2005). In addition, they strongly encourage others to learn from the past failures by planning long term for avoiding future failures during data collection (Baumard and Starbuck, 2005).

Shortcomings of the Existing Guidance

Some of the shortcomings of the existing guidance scattered across multiple disciplines are as follows:

A. Around 13% of ICT4D studies published between 2000 and 2010 make recommendations for the effective planning for data collection (Gomez, 2013); however, a majority of those studies do not provide any structured guidance to address different types of data collection challenges. For instance, Brewer et al. (2006) advise researchers to plan hard but remain flexible, collaborate with local partners, and anticipate delays in data collection due to unforeseen circumstances. But what do ICT4D researchers exactly need to do in the field to plan hard but remain flexible? Would a single strategy of planning hard but remaining flexible work for managing funds, ICT infrastructure, human resources, and other intangible
resources during data collection? De (2012) recommends that researchers prepare hard for planning and implementing ICT4D projects sensitively so that they could respond to the data collection challenges in a developing nation context. But how do researchers prepare enough to achieve a great degree of sensitization?

B. A majority of theoretical frameworks proposed for the successful design, development, and implementation of information systems in developing countries (e.g., Krishna & Walsham, 2005; Latifov & Sahay, 2013) are informed by a single project in a specific context in a developing country. The utility of those theoretical frameworks to address data collection challenges is questionable unless those frameworks are validated for distinct ICT4D projects in diverse developing nation contexts.

C. To manage ICT4D projects in developing countries, many researchers propose and apply management information systems (MIS) theories and frameworks developed in the West, which emphasize long-term rigid strategic planning (Baumard & Starbuck, 2005). However, the differences in environment in developing and developed countries in terms of laws and regulations, economic conditions, predictability of the market, level of ICT infrastructure, skilled ICT workforce, etc. demand researchers to be flexible, dynamic, and alert during data collection with marginalized communities (Bingi et al., 2000; Roztocki & Weistroffer, 2009), requiring very different approaches to designing, testing, deploying, or assessing the impact of ICT on the developing world (Roztocki & Weistroffer, 2011). The traditional MIS theories and frameworks developed in the West with comparatively more predictable and stable socioeconomic conditions may not be applied “as is” for addressing data collection challenges in developing countries with unpredictable and unstable socioeconomic conditions.

As a result, often-times inexperienced post-graduate students and researchers collecting data from marginalized communities do not have clear answers to critical questions like: “How do I know which expert advice is applicable to what circumstances in developing countries?” and “When do I select a particular lesson-learnt-in-the-past for resolving certain data collection challenges?”

This paper consolidates and organizes the scattered guidance from multiple disciplines, illustrating a systematic and structured approach for addressing a wide range of data collection challenges. It also summarizes main advice from others to better prepare researchers for undertaking ICT4D projects in the future.

METHODS

This study builds on an empirical base of 380 artifacts. The artifacts include journal articles, books, master’s theses and doctoral dissertations, national and international conference proceedings, newspaper articles, and project reports submitted to grantees. The artifacts are authored by scholars from academia, public agencies/institutions (e.g., World Bank, Ford Foundation, and Bill and Melinda Gates Foundation), and private firms (e.g., Yahoo and Microsoft Research). The artifacts can be categorized into empirical study findings, observations from ICT-prototype pilot studies, reviews of ICT designed for a particular marginalized community, and comparisons of ICT-based practices in developing countries.
Rationale for Selecting the Sources

The author commenced the study by compiling peer-reviewed articles published by top-tier ICT4D journals. Proceedings from the conferences like ICT4D, ITU World Telecom Development Conference, Mobile Communications for Development, and conferences held by IFIP, with acceptance rates of 40% or less also enriched the collection. In the second phase, a snowball sampling approach was adopted to access artifacts from a variety of journals (e.g., MIS Quarterly, IT and People, Communications of the ACM, etc.) in the disciplines related to and outside of ICT4D. On multiple occasions the EBSCO, ERIC, ProQuest Dissertations, and Google Scholar databases were used to retrieve academic artifacts other than journal articles. Search engines of public agencies/institutions and private sector firms including Google were instrumental in locating practitioner reports and non-academic artifacts published since 2000.

Data Analysis

Author’s Lens for Data Analysis. The author has a 7-year experience of collecting data from marginalized communities in India. In the past, he actively participated in six ICT4D projects focusing on the (a) barriers to the adoption of mobile phones by female slum-dwellers, (b) adoption and continued usage of mobile phones by women earning less than dollar a day, (c) over-indebtedness of microfinance borrowers, (d) factors influencing the adoption of IBM’s spoken web service by farmers, (e) success and failure of mobile banking initiatives that do not require the ownership of mobile phones by clients in remote rural areas, and (f) factors essential for the catalytic innovation in microfinance, which helps microfinance institutions scale rapidly to serve millions of unbanked poor in developing countries. He and his teammates encountered and addressed a number of data collection challenges at various stages of the above ICT4D projects, which motivated him to undertake this study and shaped the process of analyzing data collection challenges in this paper.

The author analyzed 380 artifacts using grounded theory principles. Grounded theory is a method for conducting qualitative research originally conceptualized by sociologists Barney Glaser and Anselm Strauss (1967). The data analysis for this study is based upon the method as explained by Glaser and Strauss. The open coding, i.e. line-by-line coding, of more than 150 challenges documented by the 380 artifacts led to more than 50 axial codes that were further clustered into 7 select codes. For instance, the author recommends managing the scope of data collection based on two axial codes, namely, setting unrealistic goals for data collection and lack of planning for data collection. The axial codes were derived from open codes such as: lack of time, money, and other context-specific resources, inflexible schedule for data collection, selecting research sites without map or transportation, and uncoordinated data collection at multiple research sites of the same research project. In another instance, the author identified insufficient funds for data collection, abrupt discontinuation of funds during data collection, and sudden increment in the cost of hardware devices and other resources during data collection as some of the open codes which led him to propose the following three axial codes: insufficient operational budget, unreliable sources of funding, and unpredictable operational cost; the axial codes underscores the significance of managing cost during data collection.

Table 1 demonstrates the process of deriving seven clusters of challenges extracted from the ICT4D literature considered for the analysis. The analysis of the challenges suggests that the inability of the researchers to manage scope, time, cost, quality, human resources, communication, and risks related to data collection makes it challenging.
The next section illustrates the ways to manage the seven dimensions drawn from the literature on ICT4D projects successful in addressing data collection challenges in the past, the lessons learnt by researchers, and their advice for collection data from marginalized communities in developing countries.

**WAYS TO MANAGE SEVEN DIMENSIONS**

Considering the variation in contextual factors related to developing nation contexts, it is important to note that a successful practice in one context does not guarantee the same level of success in other contexts. Researchers may need to make appropriate changes to a practice successful in one context when applying it to other contexts. Failures and shortcomings documented in past studies are helpful to understand what not to do or how mistakes could be avoided during data collection. In most of the past cases, researchers were not proactive in avoiding data collection challenges. However, others could learn from their mistakes, failures, and successes for proactively managing the following seven dimensions in the future.
Managing Scope of Data Collection

While planning and defining the scope of data collection, past ICT4D studies advise to take into consideration contextual factors including socio-cultural inequalities and religious beliefs that could possibly threaten the execution of data collection (Braa et al., 2004; Walsham & Sahay, 2006). To study the usage of mobile messages in promoting maternal health, Ramachandran et al. (2010) stayed at a house owned by an upper-caste family in rural India. Due to rigidly structured social hierarchy, lower-caste people in rural India do not mingle with upper-caste families and enter their houses. To have lower-caste members participate in the study, the research team had to move their study from the inside of the house to the front yard. The challenges caused by pre-existing socio-cultural inequalities in developing nation contexts could be avoided if the scope of data collection is designed around contextual factors.

Researchers break down the scope of ambitious ICT4D projects into manageable objectives and deliverables for synchronous data collection at multiple sites. The Technology & Social Change Group at the University of Washington undertook a global study to learn the ways in which trust and perceptions shape uses of ICT at public access venues like libraries, telecentres, and cybercafés in 25 developing countries (Gomez & Gould, 2010). They developed a shared research design to form local research teams and appoint team leaders to conduct surveys, visits of over 500 sites, and interviews with over 25,000 respondents in different types of public access venues in the selected countries. The systematic planning and delegation of tasks and responsibilities to various teams allowed the Group to collect data synchronously and meet the objectives set for the ambitious research project. Souter et al. (2005) undertook a study for assessing the economic impact of telecommunications on livelihoods and poverty reduction in rural communities in India, Mozambique, and Tanzania simultaneously. Identification, recruitment, and training of project staff were three major activities that needed to be coordinated for conducting data collection in the three different countries. To facilitate synchronous data collection at different sites, multiple teams consisting of local gatekeepers were formed.

While defining the scope of data collection, precaution should be taken to avoid forming heterogeneous groups of study participants since it might take away from equal opportunity for participation from the participants. In certain cultures, men tend to dominate everything that is deemed “important” and not allow women to participate effectively. For instance, Masita-Mwangi et al. (2011) observed that in mixed groups, any networking opportunities would likely be taken over by men. In addition, with different cultures come differences in values, as well as stereotypes held by one community against another (Walsham & Sahay, 2010). The differences and stereotypes might make it difficult to bring together groups with distinct beliefs. Cross-cultural issues in data collection could be addressed by developing training material for ICT4D project personnel to educate them for multiple interpretations of verbal and non-verbal communication practices across the developing world (Braa et al., 2004).

Periodic verification of the scope of data collection could help researchers manage the challenges associated with access to research sites and ICT. Poor transportation infrastructure makes it difficult for researchers to reach respondents living in remote rural parts of any developing country. For instance, Vodafone Research Team (Samuel et al., 2005) found that roads to 11 communities and 9 businesses in Tanzanian villages were sealed. Considering the resources at hand, the researchers eliminated potential respondents from the inaccessible villages. In a similar instance, accessing and keeping track of nomadic respondents living in informal settlements posed a unique set of challenges for data collection. While collecting data from urban immigrants in South Africa, Mathee et al. (2010) found that in informal settlements, maps were either not available or considerable change had occurred between the time of map production.
and the time of data collection. The researchers considered using aerial photographs to improve sampling accuracy, but the cost of commissioning dedicated aerial photography was prohibitively expensive. In such unexpected circumstances with the limited resources at hand, they decided to eliminate potential respondents from the inaccessible sites.

Researchers cannot address data collection challenges alone; the involvement and cooperation from local assistants is critical for outside researchers to be able to address the pre-existing context-specific barriers and subsequent data collection challenges.

**Time Management**

Appropriate timing and duration (i.e. not too short or long duration) of data collection is critical in shaping the interaction of researchers with marginalized communities. Aggressive ICT prototype development timelines often limit time spent by the HCI researchers in the field. For instance, the small amount of time that HCI researchers were able to spend in the field did not give them enough opportunities to understand local culture and possible ways in which their system would be adopted by locals in a developing nation (Chetty & Grinter, 2007). Long durations of collecting responses on sensitive personal and social issues could adversely impact data collection. For instance, semi- and low-literate users from Botswana experienced mental fatigue while revealing HIV/AIDS related health information (Sharma-Grover et al., 2009).

To address duration-related challenges, researchers advise others to identify, define, and prioritize data collection activities. They also advocate for estimating and utilizing the type and quantities of resources available to complete the activities (Chetty & Grinter, 2007; Sharma-Grover et al., 2009). Better informed decisions about data collection schedules can be made after gathering the information related to time durations, resource requirements, and schedule constraints.

Fifth grade students from government-run schools in India were expected to participate in an ICT-prototype testing experiment aimed at improving multi-user sharing of existing educational applications. However, the study was conducted during a swine flu scare in 2008. This issue, coupled with widespread absenteeism in government-run Indian schools, led to the absence rates approaching 75% on specific days of data collection (Heimerl et al., 2010). Researchers cannot predict the natural or man-made calamities that might be faced by study participants; hence, to handle such unexpected challenges of the wrong timing for data collection, researchers need to have contingency plans or a plan B for data collection (Heimerl et al., 2010). To address timing-lead data collection challenges, it is advisable to undertake pilot studies for estimating appropriate timeline, duration, and frequency of data collection in ICT4D projects.

**Managing Costs of Data Collection**

The costs of one-time investments in ICT, recurring maintenance costs (e.g., electric bills, etc.) for devices supporting ICT, and ongoing operational expenses for recruiting skilled personnel could be three areas of financial investments in data collection phase of ICT4D projects. For projects offering Internet connectivity to rural Zambians, the high cost of the VSAT connections for Internet services, bandwidth, and network management emerged as a major financial barrier (Mathee et al., 2010). The exorbitant cost for access to the Internet constrained monetary planning for a qualitative study which examined how professionals in Nairobi, Kenya, use ICT in their everyday lives (Wyche et al., 2010). Large sampling requirements and longitudinal designs often incur high financial costs (Duncombe, 2011). Limited funding assigned for studying rural farmers in China prevented Wang and Chen (2010) from conducting large surveys with farmers and farm-workers in rural China.
Affordability to conduct research at a particular venue is one of the criteria for data collection. Hence, Clark and Gomez (2011) targeted libraries, tele-centers, cybercafés etc. – strategic places – to locate and interview marginalized populations accessing computers for free of charge.

Researchers can manage their scarce funds available for ICT4D projects by (a) estimating the costs of the resources needed to complete data collection activities, and (b) aggregating the estimated costs of individual activities to establish a cost baseline (Mathee et al., 2010; Wang & Chen, 2010; Wyche et al., 2010). This approach is likely to control changes to the budget for data collection. Careful planning of field activities for data collection can save monetary resources and help researchers address challenges arising from limited funds available for completing data collection. For instance, in case of low budget, researchers should not plan for undertaking massive quantitative surveys but conduct qualitative interviews or semi-structured surveys with quick turnarounds, producing presentable results acceptable to the ICT4D community; they might not able to produce valid statistical inferences but being able to capture the impact of ICT from the perspective of marginalized communities is always better than having small quantitative samples with inconclusive or limited causation (Duncombe, 2011).

Quality Management

Data quality is valuable to ICT4D projects since it may shape the ability of researchers to make sense of the expectations, needs, and experiences of marginalized communities for using ICT solutions. Researchers advise managing decisions, activities, and external circumstances that deteriorate data quality. In a study observing parental perspectives on computers in rural India, unannounced visits by researchers at the respondents’ homes were considered intrusive and unethical, straining the interactions between the researchers and the respondents (Pal et al., 2007). Tense interactions with participants could deteriorate the quality of information shared by study participants, hence, researchers should avoid any such decision and action, jeopardizing their relation with marginalized communities.

Researchers periodically monitor data collection results to decide whether the planned processes need to be adjusted for collecting the highest quality data. Sometimes it becomes difficult for respondents to answer certain questions in the presence of supervisors, spouses, or teachers if the responses are collected at workplaces, homes, or schools. For instance, lack of enough privacy provided to Ghanaian HIV/AIDS patients adversely affected the quality of information shared by them in front of their families (Paik et al., 2009). Kuriyan et al. (2008) warn researchers of not approaching marginalized communities as a way to get “the true story,” but to understand their perspective for interpreting issues and possible solutions. This approach suggests researchers to be empathetic with respondents and the context in which they respond to ICT and make use of ICT while pursuing their daily information needs.

The mere act of observing a group could influence the group’s reaction to the observation, especially if the group is aware that they have been “chosen,” which is also known as the Hawthorne effect. Paik et al. (2009) made a different arrangement to collect data from respondents in the absence of family members. Lima and Brown (2007) also observed that children and teens attending schools and colleges in Brazil were not comfortable participating in a study in the presence of their teachers. They changed their original data collection plan to make students comfortable sharing their stories in the absence of teachers. Thus researchers timely attuned data collection procedures to eliminate the causes of deteriorated quality of responses.

Datasets related to market prices for commodities, ICT adoption by marginalized communities, and small business owners are useful to jumpstart data collection in unfamiliar contexts. However, lack of systematically collected historical data is a common problem when studying
informal markets and unorganized communities. For instance, Abraham (2006) could not locate historical data on fish prices since few formal records were kept and many people hid or distorted any written records they did keep. In such circumstances, researchers develop data quality standards and eliminate sources and data that do not meet the standards. To understand the impact of subsidies on the sustainability of tele-centers in Kyrgyzstan, it was essential to get hold of datasets on the low-income citizens receiving Internet coupons (Best et al., 2009). However, incomplete lists with missing or incorrect contact information of beneficiaries forced the researchers to adopt convenience sampling which did not represent the entire population.

Without active participation and training of local assistants it is almost impossible for outside researchers to overcome language barriers, data entry errors, and the quality of data affected by the Hawthorne effect. Research methods should be designed in such a way that they respect the privacy of marginalized communities; it is necessary for researchers to learn local norms of privacy, etiquette, and decent behavior while working with marginalized communities, improving the likelihood of success in data collection.

**HR Management**

Local assistants, a key human resource for data collection, often play a multipurpose role in the data collection phase of ICT4D projects. For instance, they might be responsible for arranging transportation to research sites in remote rural parts of developing countries, introducing outside researchers to the gatekeepers of local communities, translating context and communications, conducting interviews with marginalized communities, addressing technology issues, helping researchers scale information systems, or catering to unexpected logistic requirements of data collection (Sahay & Walsham, 2006).

Due to their critical role in data collection, it is important that local assistants speak the same dialect and not just the same language as that of respondents. Researchers studying the speech-based access to health information by low-literate users in Pakistan assumed that local assistants would be able to speak the same language as that of the low-literate users. The researchers did not realize that health workers who could not speak Urdu would be of no use in collecting responses from the Sindhi speaking user population (Sherwani et al., 2007). The assumption made by outside researchers regarding local assistants delayed the data collection process. Hence, it is important to involve local assistants, right from the beginning, i.e. planning phase of data collection. A study examining the effect of mobile phones on maternal and infant healthcare made sure to engage local assistants who could speak the same language and dialect as the midwives in Aceh Besar, Indonesia even before the execution of data collection (Chib et al., 2008).

The shortage of local experts on software development, maintenance, and operation is a more long-term complicated problem than the difficulties associated with hardware purchases (Ewusi-Mensah, 2012). Many times locally recruited technical assistants fail to operate technology platforms or programming languages selected for software to be tested in the field. Sometimes local assistants flounder when asked to fix ICT. Also, the lack of basic computing knowledge by local assistants and the lack of expertise in network administration mean that each time the wireless network fails it stays down until a network administrator fixes it. To manage these issues, researchers plan ahead of time for recruiting assistants with specific ICT skills required to collect data during ICT4D projects (Blom et al., 2005; Hewett et al., 2004). However, developing technology skills and expertise among local assistants requires a significant amount of training and education, which is generally out of the scope of ICT4D projects. Direct recruitment of undergraduate and graduate students in Western universities and residents of developing countries...
living in the West, who have the requisite technology expertise and skills (Ewusi-Mensah, 2012), is a possible solution to the shortage of ICT skills required for data collection.

For addressing HR related challenges, researchers (a) identify and document roles, responsibilities, and reporting relationships for staff members, (b) improve the interaction of staff members to improve their performance in the field, and (c) track staff performance by providing feedback and coordinating changes to enhance the overall performance. Without cooperation of local assistants it is almost impossible for outside researchers to streamline their process of recruiting and retaining the local talent necessary for completing data collection.

Managing Communication

Non-verbal communication between researchers and marginalized communities is equally or perhaps more important than verbal communication in data collection. Socioeconomic and demographic differences between respondents and researchers reflected through researchers’ attire, perfume, eyewear, etc. reinforce their outsider status, and could eventually distort the way respondents perceive the researchers. For instance, women from indigenous communities in remote, rural parts of Australia shied away from researchers due to their feeling of “shame” for talking to outsiders (McCallum & Papandrea, 2009). Trust in ICT and trust in personnel involved in data collection also shapes the involvement of marginalized communities in data collection. For instance, in a study with housemaids from urban India, perceived mistrust for technology led respondents to be hesitant towards touching the technology or self-eliminating from the study (Medhi & Toyama, 2007). Sometimes researchers are being told what their respondents think the researchers want to hear, instead of being told the “real truth.” In many cultures – and often exacerbated by perceived power distances between researchers and locals – it is considered a courtesy to tell “guests” what they want to hear; it is rude to inform guests about real issues or problems. The locals are trying to please the researchers and want to support them in reaching their perceived research objectives.

Building the bond of trust with study participants is a major step toward addressing communication-related challenges for researchers. For instance, researchers can overcome the communication challenges by determining the information and communication needs of the marginalized communities with the help of local assistants, and making needed information available to them in a timely manner. Sometimes when study participants do not trust the ability of ICT, it becomes necessary for researchers to convince the participants of the utility and benefits of using ICT for their betterment. If convinced, the marginalized communities are likely to participate actively in data collection exercises. Without active participation of local assistants it is almost impossible for outside researchers to gain the trust of marginalized communities, which in return, would help them portray a positive image to seek maximum possible cooperation from marginalized communities.

Managing Inherent Risks

When researchers test ICT prototypes with marginalized communities in developing countries, there are some obvious risks associated with the experiments.

A. Fear of, unfamiliarity with, or lack of efficacy for using ICT: A team of researchers testing a multimedia application for information dissemination in marginalized communities found that fear and inhibitions of respondents for using the application kept the communities away from it (Chu et al., 2009). The nervous state of respondents for using ICT discouraged respondents from participating in the project that tested the impact of touchtone vs. speech
recognition on HIV health information access (Sharma-Grover et al., 2009). In a similar case, the negative perceptions of marginalized communities towards newly introduced ICT and their unfamiliarity with text-based technology discouraged them from participating in a study assessing the vulnerability of marginalized communities after surviving disasters in rural Asia (Chib & Komathi, 2009). Researchers should never blame marginalized communities for their inability to operate ICT prototypes. Instead they should empathetically analyze the fear and any other psychological barriers that lead to the inability of study participants to use ICT, and should encourage the participants with the help of various incentives, including financial compensation.

B. Lack of prior knowledge of technological solutions to be tested or implemented discourages or disqualifies people from participating in data collection. For instance, in a research study conducted with tutors for a blind population in India, Zambia, and Qatar, study participants were required to have knowledge of advanced Windows features for installing specific software in the study (Kalra et al., 2007). Due to a lack of ICT knowledge, the participants could not operate the software, underutilizing resources invested in the study. To address such challenges, researchers could use the purposive sampling technique where marginalized community members meeting specific study requirements are invited to be part of data collection.

C. Research teams using multiple devices and pre-installed software on the devices often encounter frequent hard-disk failures, software infections, and reinstallation issues at research sites. For instance, software malfunctioning was a serious technical problem faced by a project evaluating the use of handheld devices within primary education in Malawi (Hollow & Masperi, 2009). Timely replacement or repair of technical components with the help of local assistants is the most frequently advised solution to this problem (Brewer et al., 2006; Veeraraghavan et al., 2009).

D. Unreliable technology infrastructure in developing countries creates a number of communication barriers. Installing power generators for uninterrupted and stable power supplies to run computers and other equipment is a temporary but popular solution in ICT4D projects (Madon, 1997).

Sensitive, embarrassing, or controversial research topics could jeopardize data collection. The social stigma associated with victims of HIV/AIDS often makes it difficult for multiple data collection teams to reach out to the victims for data collection. For instance, Angolan patients were hesitant and reluctant to give personal information regarding their HIV/AIDS and sexual behavior (Cheng et al., 2008). Questions related to alcohol and drug consumption, contraceptive practices, pregnancies, induced abortions, and illegitimate child births cause a similar setback for data collection. Assuring the privacy of study participants during data collection and guaranteeing confidentiality of data collected are common strategies for collecting data on sensitive, embarrassing, or controversial research topics (Cheng et al., 2008; Paik et al., 2009).

Finally, the issue of safety vs. the ability to operate on research sites, which is often faced during data collection. The safety of participants and research teams is paramount but each environment comes with its own set of risks. Mathee et al. (2010) experienced verbal abuse, racial slurs, and physical assault (or threats of it) while conducting interviews in South Africa. The risk of offensive behavior and physical assault was elevated over weekends when the consumption of alcohol and substance abuse were at a peak. Moreover, conflicts with local organizations could also result in high rates of crime, challenging the safety of research teams. It is always advisable to avoid collecting data in risky sociocultural environments.
Researchers should identify and determine which risks might affect their data collection and its staff members, and document the risk characteristics. Prioritizing risks based on the probability of their occurrence and impact is another strategy implemented during data collection. It is important to analyze the effect of identified risks on overall data collection objectives, developing options and actions to reduce threats to data collection objectives, tracking and monitoring identified risks, and recognizing new risks and evaluating their effect on data collection.

**DISCUSSION AND CONCLUSION**

**Contribution of the Paper**

Successful stories, failures, and shortcomings in data collection documented in the past ICT4D studies suggest the strong correlation between the management of these seven dimensions and the ability to address data collection challenges. For instance, the researchers who broke down the scope of ambitious data collection conducted at multiple sites in different developing countries were successful in collecting data synchronously with the help of teams of local assistants (Gomez & Gould, 2010; Samuel et al., 2005). In contrast, the researchers who could not manage one or more dimensions experienced data collection challenges. For example, researchers who made unannounced visits at study participants’ homes were considered intrusive and unethical, straining researchers’ interactions with the study participants (Pal et al., 2007) and jeopardizing data quality. These examples and similar others illustrated in this paper support the author’s claim regarding managing the seven dimensions to address data collection challenges.

Some of the key contributions made by this paper are as follows:

A. This paper informs researchers by providing an argument in terms of how to think about and address data collection challenges. Key areas of interest for researchers covered in the paper include challenges in eliciting participation, gathering data, and working with participants in developing countries. Exploring possible ways of managing the seven dimensions associated with data collection provides a greater learning experience and “eye-opener” for researchers, which is another key contribution of this paper.

B. The overall guidance to manage the seven dimensions is based on the past ICT4D studies. The studies included both failures and successful case examples. In the future, researchers could learn from the mistakes, failures, and success stories reported by the past studies to adopt a proactive approach for addressing data collection challenges. Considering the differences in contextual factors in developing countries, this paper asks researchers to make appropriate modifications in the ways to manage the seven dimensions as illustrated by the past studies successful in addressing data collection challenges.

C. Top-tier ICT4D journals rarely offers any structured guidance to collect data from marginalized communities in developing countries. This paper fills the gap by proposing a set of seven dimensions—scope, time, cost, quality, human resources, communication, and risks—necessary to be managed for addressing the challenges related to data collection in ICT4D projects; the generalizability of these dimensions would be of immense value to ICT4D researchers in the future.
Limitation

The challenges discussed in this paper are reported by academic researchers, who may have very different motivation and resources to address the challenges than that of practitioners. Hence, practitioners may not find the study findings useful or applicable. Also, this paper does not present a comprehensive list of challenges, but it proposes ways to address a large number of data collection challenges rarely documented and analyzed in a single artifact. To a certain extent, this paper addresses the shortcomings in the existing guidance discussed in section 2.

Future Direction: Openly Engage with the Project Management Literature

Thompson and Walsham (2010) express the need to conduct *strategic* research for enhancing theoretical and practical contributions made by the ICT4D community to benefit marginalized communities. In particular, the duo advocates to “engage openly with other research disciplines (p. 112)” by adopting an outward-looking focus to collaborate with researchers across a range of disciplines, allowing for a shift in the ways in which ICT4D research is planned and conducted. The seven dimensions identified in this paper for addressing data collection challenges are similar to the principles of project management. The project management principles (PMP) guide to manage scope, time, cost, human resources, quality, communication, and risks associated with projects. Project management, which is rooted in management science literature, is a scientifically designed approach for managing a variety of projects (PMBOK, 2003).

It is important to note that a majority of the existing guidance for addressing data collection challenges is similar or related to the project management literature. For instance:

A. The lessons learnt and the guidance offered by the experienced researchers (e.g., Braa et al., 2004; Heeks & Bhatnagar, 1999; Heeks, 2009; Krishna & Walsham, 2005; Sahay & Walsham, 2006) for managing data collection efforts in ICT4D projects (see sections 4.1 through 4.7) overlap with the PMP.

B. The success stories (e.g., Best et al., 2009; Brewer et al., 2006; Chib et al., 2008; Gomez & Gould, 2010; Heeks, 2002; Krishna & Walsham, 2005; Mathee et al., 2010; Medhi & Toyama, 2007; Souter et al., 2005) do not mention or directly refer to PMP but follow a set of ideas and activities similar to the PMP.

C. The theoretical frameworks and concepts proposed for the successful design, development, and implementation of information systems in developing countries (e.g., Krishna & Walsham, 2005; Latifov & Sahay, 2013; Roztocki & Weistroffer, 2011), including Heeks’ (2003) seven dimensions of design-reality gap, share significant commonalities with the PMP.

It is time for the ICT4D community to engage openly with the project management literature. For instance, Mama and Isabriye (2014) and Pade (2006) propose applying PMP for addressing data collection challenges.

However, there are fundamental differences in some of the goals of PMP and ICT4D projects. For instance, businesses aim to achieve operational efficiency and economic rationality (i.e. optimizing costs and maximizing profits) using PMP, whereas scaling, sustainability, and benefiting marginalized communities without undesired outcomes are typically the goals of ICT4D projects. Cost optimization and efficiency could help ICT4D projects achieve their goals of scaling and sustaining project activities; however, maximizing profits and similar other profit-
driven goals of PMP could drift ICT4D projects away from their social mission of benefitting marginalized communities without undesired outcomes. In addition, due to the differences in the business environment in developing and developed countries (Roztocki & Weistroffer, 2011), PMP grounded in the West cannot be applied “as is” in the developing world. Hence, ICT4D researchers would need to be careful in applying PMP incepted in the developed world for collecting data as part of the ICT4D projects in the developing world. Potnis (2014) illustrates the ways to customize PMP for addressing gender-related challenges in collecting data from female mobile phone users earning less than a dollar a day in rural India.

The ICT4D community could also learn from the applications of PMP for managing IT projects in developing countries. For instance, the project management literature advises managers to (a) realize, assess, and adapt to local realities, cultural values, organizational norms, and work practices (Nguyen, 2007), (b) mitigate political and social demands and cultural hurdles before and during implementation of projects (Muriithi & Crawford, 2003), (c) revitalize the sluggish and ineffective management practices in organizations (Stuckenbruck & Zomorrodian, 1987), and (d) acquire and retain skilled IT professionals in developing countries (Mia & Ramage, 2011).

The striking similarities between the guidance offered by the ICT4D literature and the ICT4D project management literature call for breaking academic silos and collaborating with each other to address data collection challenges in ICT4D projects aiming to achieve social, economic, and human development of marginalized communities across the developing world.

ACKNOWLEDGMENT

Several informal discussions with researchers and practitioners in the US, India, China, Saudi Arabia, Mexico, Ghana, and Ethiopia in the past enriched this paper.

REFERENCES


Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.


