Paper-Digital Workflows in Global Development Organizations

Nicola Dell† **Trevor Perrier**† **Neha Kumar**† **Mitchell Lee**† **Rachel Powers**‡ **Gaetano Borriello**†

†Computer Science and Engineering, University of Washington
{nixdell,tperrier,neha,mitchl2,gaetano}@cs.washington.edu

‡VillageReach
rachel.powers@villagereach.org

ABSTRACT

Global development organizations rely on the essential affordances provided by both paper and digital materials to navigate hurdles posed by poor infrastructure, low connectivity, linguistic differences, and other socioeconomic constraints that render communication and collaboration challenging. This paper examines the collaborative practices around paper-digital workflows within global development organizations operating in low-resource environments. We use a mixed methods approach to gather data from 23 organizations in 16 countries. Our findings show the tensions that arise between the ubiquitousness of paper and the desirability of digitized data, and highlight the challenges associated with transitioning information several times between paper and digital materials. We also reveal design opportunities for new tools to bridge the gap between paper-based and digital information in low-resource settings. Finally, we contribute a nuanced understanding of the cross-cultural and infrastructural challenges that influence the paper-digital lifecycle. Our findings will be useful for researchers and practitioners interested in understanding or participating in the workflows that drive global development.

Author Keywords

Paper; digital; paper forms; data entry; affordances; global development; ODK Scan; ICTD; DEV; HCI4D.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

General Terms

Design; Human Factors; Measurement.

INTRODUCTION

Global development organizations, big and small, have gradually but steadily evolved into an established presence across the world, particularly in low- and middle-income countries. Teams within these organizations are frequently present in multiple countries at once, engendering complex and collaborative work environments that require an ever-present awareness of temporal, spatial, and cultural divides. Individuals with different backgrounds, perspectives and responsibilities work together on the design, deployment, and evaluation of initiatives that aim to examine and/or improve the lives of socioeconomically disadvantaged populations around the world. At the same time, these initiatives may require them to collaborate with various partner, donor, and government organizations to perform large-scale data collection in the communities they target. For the relevance and success of these development initiatives, collecting this data becomes an essential endeavor.

The task of gathering accurate and timely data from people living in low-resource environments is not without its challenges. In regions constrained by poor infrastructure and limited resources, where the availability of technology is limited and digital literacy is slowly rising, data must often be collected using paper - not digital - materials. People trust and are familiar with paper. Paper is tangible, portable and does not require batteries. These material affordances make it easy to use and almost universally accessible. As a result, paper documents have always provided essential support for organizational work processes, particularly in low-resource settings, and will remain an integral and critical component of global development workflows for years to come.

However, although paper has been used to collect data for hundreds of years, the onset of digital data has transformed the organizational workflow. Data in digital formats, particularly 'big' data, is easier to navigate, process, manage, and store than paper-based data, and many of the complex analyses and visualizations that now routinely help people to make sense of collected data are only feasible if the data is in communicable, searchable and mutable digital formats. Our research examines the tensions between the ubiquitousness of paper and the desirability of digitized data as we uncover the collaborative practices surrounding paper-digital workflows in global development organizations. We use a mixed methods approach to study the paper-digital lifecycle from the perspective of the researchers and practitioners that drive development initiatives, organizing our findings in relation to the different stages of the data lifecycle as it is sought, collected, digitized, and analyzed.

The CSCW community has long interested itself in the tension between paper and digital materials [3] [30] [32]. We introduce a novel dimension to this conversation by examining paper and digital artifacts in low-resource settings,

CSCW 2015, March 14–18, 2015, Vancouver, BC, Canada.

ACM 978-1-4503-2922-4/15/03…$15.00.
http://dx.doi.org/10.1145/2675133.2675145

where infrastructural challenges play a significant role. We also contribute to the rich body of prior work that is concerned with understanding organizational workflows and exploring how individual activities are organized to achieve collective ends [31]. Although there has been a growing interest in CSCW research that focuses on disadvantaged communities [21] [29] and geographically distributed teams [18], very little research thus far has examined the workflows of global development organizations. These organizations often operate in low-resource settings where they must navigate hurdles posed by poor infrastructure, low connectivity, cultural differences, and other socioeconomic constraints.

We contribute an in-depth analysis of the complex organizational workflows that drive global development, focusing on the perspectives and understandings of the researchers and practitioners who design field studies to collect data from target communities. We show that workers within these organizations spend considerable time and effort transitioning data between paper and digital materials, while attempting to preserve the correctness and completeness of the data. In addition, we highlight workers' frustrations with the software tools that they use to collect, communicate and synchronize data between remote offices, and present design opportunities for new tools to better support their workflows. Finally, we reveal the impact of infrastructural, cultural and socioeconomic challenges on the paper-digital lifecycle in the context of global development work, many of which can only be understood, not eliminated. Taken together, our findings will be useful for survey designers, researchers and practitioners interested in understanding and participating in the complex workflows that drive global development.

BACKGROUND AND RELATED WORK

Paper and Digital Materials in Low-Resource Settings

Paper has long played a central role in the collection and communication of information within governments and other organizations. Gupta [14] and Hull [19] note that, by moving through a bureaucracy as documents, physical materials such as health registers, government notices, or written complaints gain status and trigger actions by officials. Veeraraghavan [36] describes how digitizating paper records can reduce corruption. Singh et al. [33] analyze the use of paper in non-governmental organizations (NGOs) and show that workers prefer numeric and multiple-choice surveys since they can fill them more quickly and accurately than text-based surveys.

A variety of technical interventions have been proposed to bridge the gap between paper and digital worlds in low-resource settings. CAM [26] was among the first to propose the use of a camera-enabled mobile phone for assisted data entry from paper surveys. ODK Scan [7], a smartphone-based system that uses computer vision to interpret data from paper, recently showed that transforming field data into machine-readable documents is a challenging task [8]. Shreddr [6] (now Captricity [5]) takes a novel approach to digitization, segmenting paper surveys and assigning them as images for transcription by crowd-sourced workers. Ratan et al. [28] combined physical paper with a digital slate device that provided electronic feedback. Although this performed well in usability studies, the purchase and maintenance of specialized slate devices hindered its scalability and sustainability [25]. The PartoPen [35] uses a digital pen-and-paper solution for monitoring women in labor. The system provides real-time reminders for health workers, but extracting data from the pens and integrating it with patient medical records or other data collection tools remains a challenge. In addition, a smart pen is needed for each worker, rendering the system less scalable for low-cost.

Although many of these interventions aim to bridge the gap between paper and digital materials in low-resource settings, they focus mainly on evaluating the usability of new technologies with end users. Our work examines the collaborative practices surrounding paper-digital workflows at multiple organizations, focusing on key moments at which data is transitioned between paper and digital materials.

Bridging the Gap Between Paper and Digital Materials

According to Harper and Sellen, *"Observation of any organizational setting only serves to confirm that the most pervasive, ubiquitous artifact in support of collaborative work is paper"* [16]. The fact that paper and digital materials are both widely used suggests that each medium provides essential affordances that the other does not [3][32]. Paper is cheap and has inherent material properties that make it easy to use and access in almost any environment. Digital documents are easy to store, search, link and analyze. However, coordinating information across paper and digital materials has proven challenging, and a large body of prior work examines and aims to bridge this paper-digital divide.

Proposed technical solutions focus on personal document management and new ways of integrating the affordances of paper and digital materials. Several solutions use camera- and projector-based systems to enhance paper with digital resources [20] [34] [37] or augment computers with paper affordances [9] [13] [17]. Other systems use digital pen technologies to link physical locations on paper to digital content [15] [23] [27]. Although these studies repeatedly emphasize the importance of paper for a diverse array of work processes, they focus mainly on the management of personal documents for greater individual productivity. By contrast, we consider how *organizations* collect and transition information between paper and digital formats. In this context, Luff et al. [22] examined the relevance of paper in collaborative workflows in an architectural practice, a medical center and a London Underground control room. Harper and Sellen used ethnographic methods to assess the role of paper in an air traffic control room, a police station and the International Monetary Fund [16]. Other studies examine the use of paper and digital media in healthcare [11], particularly with respect to medical records [10] and information flow [4]. These studies emphasize that, despite efforts to replace paper with digital alternatives, paper remains an integral component of any work environment.

The above studies focus, however, on the work practices of organizations situated in infrastructure-rich settings. Our research focuses on paper-digital workflows within global development organizations that span multiple cultures and

geographies, including low-resource environments, and must frequently struggle with infrastructural hurdles such as poor electricity and Internet connectivity. Through an in-depth consideration of these constraints, we identify factors that could influence the design of future technologies that aim to support global development workflows.

METHODOLOGY

Our study contributes an in-depth understanding of paper-digital workflows within the global development context by examining the work practices of our target organizations, particularly as they relate to the use of paper and the move towards digitization. The methods we use to arrive at this understanding include an online survey, a design probe, and a set of interviews. We describe these in detail below.

Online Survey

Our online survey consisted of 50 questions that sought information regarding the demographic background of the participants and their organizations, their survey design and approvals processes, data collection and entry, and collaborative work practices. Our 48 participants from 23 organizations (of varying sizes) were situated in 16 countries and have been active in a variety of domains, including health, education, agriculture, finance, and logistics. Their responses gave us a preliminary and global perspective on the relevance of paper and the importance of digitization in these domains.

Design Probe

Based on our survey responses, we prototyped a new survey design tool to expand our understanding of the design process within our target organizations. The tool was designed with the single objective of ease of digitization and allows participants to create a variety of question formats that are optimized for machine-readability (see Figure 1). Participants can move and reformat individual questions, and import content from a variety of other sources, including from existing images or Microsoft Word. The tool is freely available online, and paper surveys that are generated using the tool can be automatically digitized using the ODK Scan application [7]. Our goal was to use the tool as a probe [12] to elicit survey designers' views on the process of converting research questions into material paper surveys and the value of optimizing for digitization and machine-readability.

Interviews

Participants

We recruited nine participants from three global development organizations[1]. All participants were female, aged between 27 and 37 years, and selected based on (1) having participated in our online survey; (2) being actively engaged in designing paper-based surveys for global development initiatives; and (3) being willing and able to participate in our study. We interviewed both researchers and practitioners, since both play vital roles in global development workflows but are somewhat differently motivated. Participants who identified themselves as practitioners were usually employed by NGOs primarily concerned with applying existing knowledge and

[1]Organization names have been anonymized for review.

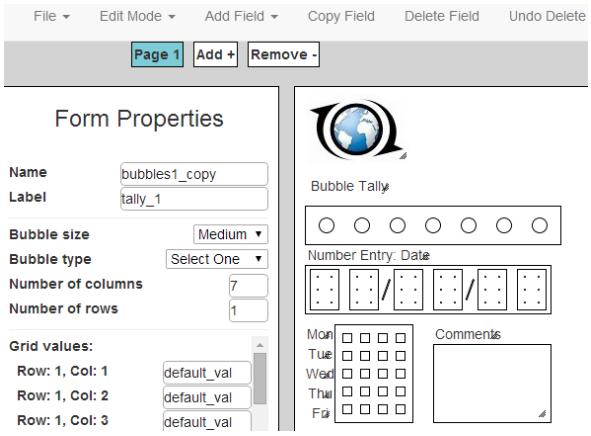

Figure 1. Screenshot showing how our prototype design probe generates a variety of machine-readable question types and incorporates content generated by other tools (such as images).

technologies to improve the quality of life of people in resource-poor settings. By contrast, participants who described themselves as researchers usually worked for academic institutions and were motivated by high-level research questions that they hope will be answered through field studies in low-resource settings. These participants were concerned about ensuring the validity of the study, enrolling well-defined samples of specific target populations and controlling for external variables. Having both kinds of participants afforded us a more nuanced understanding of the workflows involved in global development initiatives.

Finally, all our participants are survey designers who work with diverse and geographically dispersed teams, but who are themselves primarily based in the US. Our findings therefore document the priorities and perspectives of these survey designers with regard to the paper-digital lifecycle within their organizations. Exploring the perspectives of the other stakeholders will be the subject of future research.

Procedure

Participants were introduced to our prototype design tool by watching a 20-minute sequence of tutorial videos that described the tool's features. After watching the videos, we asked participants to use the tool to create a previously designed paper survey of their choice and observed as they completed this process on their own computers. This phase of the study took approximately 30 minutes and we encouraged participants to articulate their thought processes by requesting them to follow the think-aloud protocol.

We also conducted in-depth 60-minute interviews with the same nine participants. The interviews began with a short design exercise. Participants were provided with a research question and used a paper and pen to design a survey that would collect the data required to answer the question. The research questions were selected at random from the World Health Survey[2] conducted by the World Health Organization. This phase of the interview lasted approximately 15 minutes and allowed us to observe the participants' design process

[2]Available at http://www.who.int/healthinfo/survey/en/

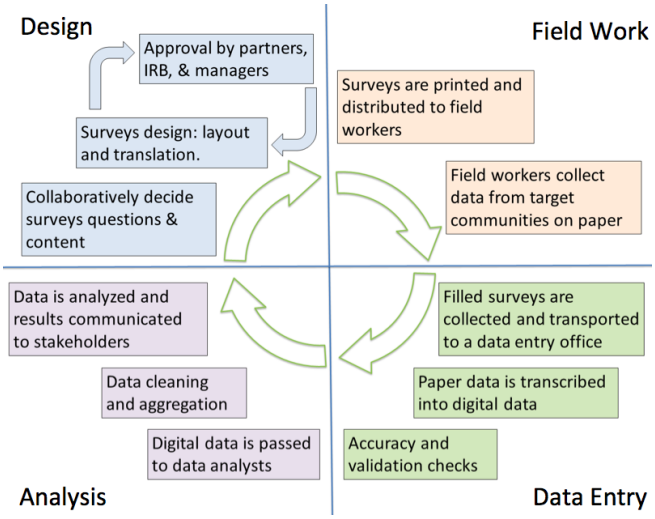

Figure 2. The four phases of the paper-digital lifecycle in the context of global development work: design, field work, data entry, analysis.

and how they prioritized data digitization. The next phase of the interview lasted approximately 45 minutes and consisted of an in-depth discussion of the participants' current work practices, their backgrounds, work responsibilities, current survey design tools and processes, the workflows surrounding the use of paper surveys in the field, and the data entry processes employed by their organization.

Analysis

To analyze our data, we went through our survey results and interview transcripts and organized them using the *"design"*, *"field work"*, *"data entry"*, and *"analysis"* codes to determine which findings were relevant for the distinct phases of the workflows that we examine. For each of these codes, we conducted iterative analyses to ascertain and organize our prominent findings that we present below.

FINDINGS

We study paper-digital workflows from the perspective of researchers and practitioners responsible for driving global development initiatives. The success of these initiatives depends upon effective collaboration across four phases of work between remotely located workers, as depicted in Figure 2. In the design phase, researchers and practitioners use digital tools to create surveys. In the data collection phase, workers in the field convert these surveys to paper and use them to collect data. In the data entry phase, workers transcribe data from paper into digital formats. Finally, in the analysis phase, researchers and practitioners process the digitized data to address research or program objectives.

At a high-level, we identify several layers of issues that affect paper-digital workflows in global development organizations, including (1) issues that relate to the material affordances and limitations of paper and digital media; (2) issues that relate to the low-resource nature of the work environments; and (3) issues that relate to the cross-cultural and geographically distributed nature of global development work. In presenting our findings, we limit our focus to how these issues affect the four phases of the paper-digital lifecycle, although we note that each category is worthy of additional study.

1. Design: Creating Paper Materials with Digital Tools

Our analysis reveals that survey design is a highly collaborative process involving multiple stakeholders, although effective collaboration is challenging for geographically dispersed teams that may speak different languages. Survey designers must also manage a number of trade-offs regarding the quantity of data collected from target communities, the layout and complexity of surveys, and the amount of work required to complete the survey. To make them easy to fill, surveys are typically highly structured documents that designers struggle to create using currently available software tools. However, the structured nature of the surveys means that, for the most part, they are already optimized for machine-readability, despite not currently being digitized by a machine.

1a: Collaborative survey design is challenging for geographically dispersed teams that may speak different languages

The process that determines the content to be included on new surveys is collaborative and involves multiple stakeholders, often remotely located, *e.g.*, office staff in the US, field staff in target countries, Ministry officials, and funding agencies. In many cases, time differences between stakeholders in different countries introduce additional complexity to the collaboration, with one of our participants describing how it took her *"four months to go from the idea to the survey."* Potential delays and constraints for meeting times must be factored in, as must holidays. Feedback from staff in target countries takes at least a day, while feedback from Ministry officials can often take a few weeks. In addition, obtaining the necessary approvals for survey designs can take anywhere from a few days to many months.

Moreover, surveys designed by English-speaking researchers and practitioners must be translated into local languages if they are to be understood by all interested parties. However, communicating the goals of a survey across languages can be challenging and potentially lossy, particularly if the translator is not a native speaker of both. This could introduce subtle differences between the meanings intended by the designers, and the meanings on the translated survey. In addition, as the survey design evolves, these differences could become compounded by iterative design, editing and translating. Furthermore, the translation can affect the appearance and layout of content on the survey, which may complicate the work processes of both the designers and the translators:

> *"For most English-Portuguese translations, the stuff that takes three words in English takes seven in Portuguese. So we really have to think about how we fit it all in."*

1b: Survey designers manage several trade-offs regarding the content, layout and complexity of surveys

Survey designers face a number of challenges deciding the content, layout and complexity of new surveys. Many participants described how they strive to design surveys that collect only the data necessary to answer specific research questions and nothing more, explaining how trying to collect too much data would both complicate their analyses and increase the likelihood of error. However,

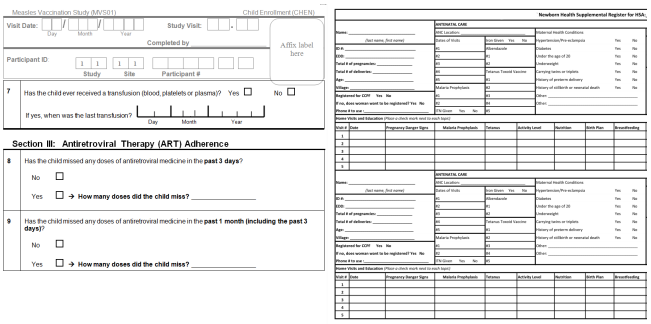

Figure 3. Two surveys that highlight how some of our participants optimize for clarity (left) and others optimize for density (right). Note that we do not intend for readers to read the questions on these surveys, rather, we are illustrating their high-level layout characteristics.

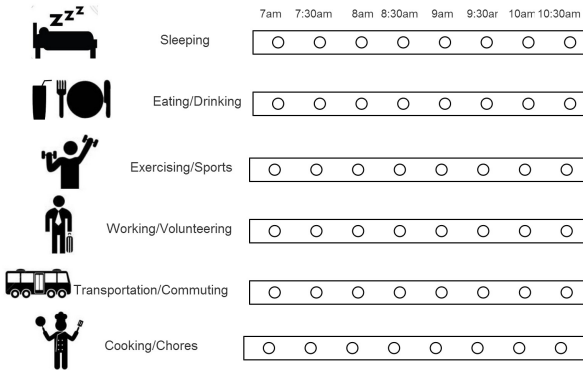

Figure 4. Part of a survey created for use by rural farmers in Tanzania.

since collecting data from target communities in low-resource settings requires organizations to invest large amounts of time and effort, it often also makes sense to collect additional data that could be used to answer future research questions:

> *"It's always a push-pull of collecting enough information to be useful in the future, even for questions we haven't thought of yet. I mean there's stuff in here that I'm not using at all, there's a lot of redundant data. Looking back on this, it could be a little more streamlined. But there's a lot of data here that isn't used yet."*

Many participants were also concerned about how field workers would perceive the amount of work required to fill in a survey and they devised strategies to make surveys appear easy to complete. Some participants prioritized density (see Figure 3, right), packing large amounts of information onto a single page to keep the survey as short as possible:

> *"We don't want them carrying around too many sheets of paper, so we try to get as much information as possible on one page. And that's also because of cost."*

Other participants prioritized for clarity and compromised on space (see Figure 3, left). Then, they broke each question into multiple parts so that it appeared as if the entire survey consisted of only a small number of questions:

> *"You can see that there's question 3a, 3b, 3c. We have stupid numbers of subquestions. And that was specifically asked for by the team, because psychologically if we're on question 7 in section 1, that's much nicer than question 23 of 49."*

In either case, the designer's expectation and understanding of who will use the survey in the field heavily influenced the complexity of the survey. For example, one participant designed two surveys to collect data regarding how rural farmers in Tanzania spend their time. The first is a household survey designed to be filled by trained field workers employed by the organization. Since the designer expected that these workers would be relatively well educated, the survey consists of many pages and includes complex features like conditional branching. By contrast, the second survey is a diary that will be filled out by rural farmers. The designer does not know the characteristics of these farmers, whether they will be literate, or their willingness to engage in the work. As a result, the survey is a single page that consists of images showing common household activities with a simple grid of bubbles for farmers to fill in (see Figure 4).

Since the consequences of misjudging the field workers' abilities may substantially reduce the quality of the data collected, designers use a variety of strategies to increase the likelihood that field workers and target populations will understand the questions. One common strategy is to copy question formats from pre-existing, tried and tested surveys:

> *"We are trying to write as little original content as possible and use past surveys that have already piloted these questions in a similar context."*

Another popular content layout strategy involved mimicking the format of government registers whose contents were to be transferred onto the survey. Keeping the structure of the two surveys identical may make it easier for field workers to simply copy the content from the register onto the survey.

1c: Designers struggle to create surveys using currently available tools and are willing to try new tools that optimize surveys for digitization and machine-readability

One goal in our study was to probe participants' reactions to the idea of optimizing their surveys for machine-readability and digitization. Although machine-readable surveys exist (such as Scantron[3]), none of the tools that our participants use currently support machine-readability. Instead, almost all of our survey participants reported using Microsoft Word as their primary survey design tool.

However, despite its prevalence, a large number of our interviewees were vocal about how challenging and frustrating they found the process of designing their highly structured surveys in Word. One participant who tried to switch from Word to Adobe InDesign shared that it had taken her several months to learn the new software but *"there is a startup cost to learning any new tool,"* and InDesign allowed her to *"create surveys that look very pretty"*. However, she also found the cost of purchasing the software to be significant (InDesign currently costs US $240 per year) and,

[3] Available at www.scantron.com

since designing surveys is typically collaborative, her team members would also need to purchase the software to edit the survey. Thus, despite her (time and financial) investment in the new software, and a personal preference for InDesign, she returned to using Word.

Although it is clearly challenging to incorporate new software into their workflows, all of our participants were eager to learn about viable alternative design tools. Moreover, they welcomed the idea of designing surveys that were optimized for machine-readability, with several assuming that automated digitization would be more accurate than human data entry. When participants used our prototype tool to create their surveys, all of the question types that they needed could be represented using features that the tool offered. In addition, almost all data fields used for analysis are already structured as numeric, bubble or checkbox fields, which could be automatically interpreted. Text, which cannot be automatically converted, must be saved in image format and/or manually transcribed by a person. However, several of our participants' surveys contained no text-based data, while some contained a single text field for the person's name, which was typically not used for analysis:

> *"Nobody ever looks back at the name. I don't see how the name field would be relevant, except maybe if you needed to contact the person later."*

These findings suggest that, for the most part, our participants' surveys are *already* optimized for machine-readability. They simply need to be created using software that enables a machine to digitize the data. Participants also claimed that they would be willing to spend substantial amounts of extra time designing the survey if it meant that the survey would be machine-readable and ease the data entry process, particularly for large scale surveys:

> *"There would be 1200 of these. 10 hours upfront definitely outweighs typing in 1200 paper surveys. I think the time trade-off is in favor of [the tool]."*

They also argued,

> *"Using [the tool] might not take longer than doing it in Word, because creating, resizing, moving and aligning so many boxes in Word can be very challenging."*

Finally, we noticed that participants' thought processes were frequently shaped by the software that they use. For example, participants told us, *"I would do this as a table in Word"* or *"I use boxes because creating bubbles in Word is a nightmare."* When asked to design a survey using our prototype, they again thought in terms of the tool's features, *"Yesterday I would have said to use a handwritten number. Now, I would use bubble tallies."* Participants' experiences using our tool suggest that using software that encourages them to think in terms of digitization could change survey designers' perspectives and increase their awareness of digitization.

2. Field work: Collecting Data from Target Communities

In the second phase of the paper-digital lifecycle, paper surveys are used by field workers to collect data from target communities. Our findings highlight both the benefits and challenges of paper-based data collection. Although paper does not require power or Internet connectivity, field workers who use paper still face a variety of social and infrastructural challenges. In addition, despite working hard to design understandable surveys that are appropriate for use by field workers, many participants admitted that they do not fully understand the complexities of conducting field work in low-resource settings.

2a: Field workers conducting surveys in low-resource settings face infrastructural and social challenges

One of the primary benefits of using paper surveys to collect data in low-resource settings is that recording data on paper does not require electricity or Internet connectivity. However, the printed paper surveys still need be transported to target communities, and limited transportation infrastructure means that field workers frequently travel on foot, which can make it difficult for them to carry a large number of paper surveys:

> *"Our clinic assistant takes these binders, puts them in a suitcase and rolls them to our office. And then if something is missing she has to take them back. And they go back and forth. It's terrible. This poor girl...it's an enormous suitcase full of binders."*

In addition, field workers may also face social challenges operating within target communities. Participants reported that field workers would sometimes feel unwelcome or uncomfortable asking questions of a personally sensitive nature, and they expressed sympathy for the field workers:

> *"They are out in the field, having people slam doors in their face and telling them to go away. If they are interacting with somebody who is getting fidgety or bored, I imagine they would want to skip questions or rush through questions that have lots of options."*

However, although the physical properties of paper do make it feasible for field workers to skip questions or sections of the survey, this behavior is not desirable from the perspective of the researchers and practitioners since it would result in incomplete or missing data. Moreover, several participants told us that field workers frequently repurpose parts of the physical paper as they see fit. In one case, a field worker crossed out a particular question, wrote in a different question, and then recorded an answer to the new, handwritten question (see Figure 5). Again, although modifications like these are relatively easy to make on paper, they have not been intended or approved by the survey designer, and almost certainly complicate the data entry, since the value recorded on the paper no longer corresponds to the value expected by the digital entry form. To resolve these kinds of issues, participants expressed that it would be preferable to digitize the data in the field so that they could ask field workers clarifying questions if necessary.

2b: Survey designers may not fully understand the complexities of field work in low-resource settings

Although our participants spend substantial amounts of time thinking about the work performed in the field, several acknowledged that they had never met the field workers and

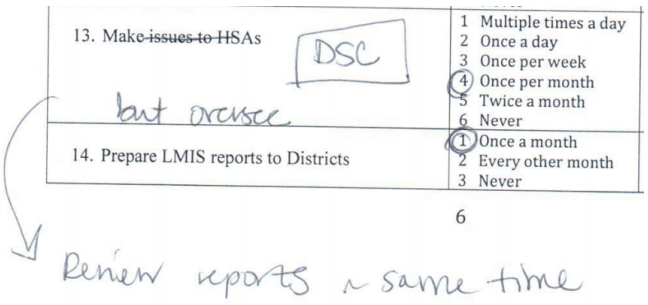

Figure 5. Part of a survey that shows how field workers can take advantage of paper's flexibility, which may make data entry challenging. In this case, a field worker repurposed part of the survey to record something new that the data entry worker may not be able to transcribe.

did not fully understand their backgrounds or perspectives. This lack of understanding sometimes leads to a mismatch between the expectations of the researcher/practitioner and the capabilities of the field workers. For example, when we asked one participant if field workers had trouble filling out her complex survey, she told us, *"They're used to filling out complex registers,"* although she later revealed that *"60% of [the field workers] couldn't do it after two days of training."* In addition, several participants were concerned about the quality of their field workers and the potential for fraudulent data [1] and many expressed a desire to monitor field workers and track their progress as the data is collected.

Finally, participants described how there are frequently events that occur in the field that cannot be anticipated or predicted. Since our study focuses on analyzing paper-digital workflows from the perspective of the survey designers, additional research is necessary to comprehensively analyze the complexities of field work in low-resource settings.

3. Data Entry: Transcribing Data from Paper Surveys

All of our participants hire data entry workers to manually transcribe paper-based data into digital formats and identified data entry as a major bottleneck in their workflows. Participants described how intermittent power and unreliable connectivity complicate data entry and communication. However, despite these challenges, data entry workers receive less attention and training than field workers. Instead, researchers and practitioners attempt to control the data entry process by constraining the values that may be entered and by employing time intensive quality control techniques.

3a: Unreliable electricity and intermittent connectivity complicate the process of entering and communicating data

Many participants explained how unreliable electricity and intermittent Internet connectivity can complicate the process of transcribing data into digital formats:

> *"There is limited electricity and Internet in the data entry offices. They mostly use desktops, so when the power is out, they can't do data entry, and when the Internet is down, they can't enter data using REDCap."*

Our participants predominantly use two software tools to store their entered data: Microsoft Access and REDCap[4]. The benefit of using REDCap is that data is entered directly into an online database. This simplifies the synchronization of digitized data between offices in different locations, and allows stakeholders in other countries to view or analyze the data as soon as it has been entered. However, entering data into REDCap requires continuous access to the Internet, which means that if there is no connectivity, data entry workers are unable to enter any data. By contrast, Access is often chosen because it is capable of running locally on the data entry worker's computer, which means that (s)he can still enter data in the absence of an Internet connection. However, using Access complicates the process of synchronizing and communicating the data to remote researchers and practitioners:

> *"After entering the data, [the data entry worker] zips the database and uploads it to Google docs. It's awful. I wrote out step by step how to zip it. A real issue for us is figuring out how to get data back in a secure way."*

This finding illustrates how even digital information can be difficult to store and communicate in low-resource settings. In addition, in some situations, organizations do not have sufficient resources to perform data entry in the target country. Instead, several participants described how they outsource the data entry to external companies, which may introduce additional challenges:

> *"I know that for a paper survey that we worked on in Mexico, we ended up scanning all the surveys to somewhere in Thailand, and these Thai workers were inputting Spanish into their computers."*

3b: Survey designers may not fully appreciate the challenges of performing data entry in low-resource settings

Our findings reveal a mismatch between the expectations of the survey designers and the constraints experienced by data entry workers. Many survey designers assume that data entry is a relatively simple task that should happen seamlessly, with one participant describing the data entry worker as *"the comfy data person, sitting in an office, sipping a beverage."* However, in contrast to the assumption that data entry workers simply sit in front of a computer all day, we discovered that in fact data entry workers are often responsible for an array of additional tasks, such as preparing surveys for field workers, collecting and transporting completed surveys to the data entry office, communicating with field workers to resolve discrepancies, filing surveys for safekeeping, and preparing reports.

Moreover, all of our participants prioritized the needs of the field workers over the needs of the data entry workers, with most reporting that their field workers are usually well-trained and informed about the project's objectives so that they are motivated to collect good quality data:

> *"We try to set up the environment with high standards. Having a passion for the data they are collecting will help ensure the quality of data is high."*

[4]REDCap is an open-source cloud-based tool designed for research.

By contrast, when we asked if data entry workers understood the project's objectives, several participants told us that *"there really isn't any need for them to know about the goals of the project."* In addition, several participants admitted that they were unaware of how long it takes workers to enter data:

> *"I don't know how long it takes them to enter one of these forms. At one point I heard some number that was a little bit frightening."*

Given the lack of attention afforded the data entry workers, it is perhaps not surprising that the organizations in our study find it difficult to recruit and retain good data entry workers:

> *"We've had high turnover with data entry people. Data entry is a tedious job, it is boring, and so it's tough to get someone who wants to do it and wants to do it well."*

3c: Survey designers try to ensure data accuracy by constraining the values that may be entered and by employing time intensive quality control techniques.

Maximizing the accuracy of entered data is a priority for our participants and they described a variety of strategies that they use to ensure entered data is complete and correct. For example, many participants spend a large amount of time designing the data entry interface to look as much like the paper survey as possible. This allows data entry workers to more easily find the correct digital entry box for each value that they see on the paper, although it complicates the process of creating the digital data entry forms:

> *"After this paper survey was already designed, it took me about four months to recreate the survey in Microsoft Access. It took forever. It's a huge process."*

Moreover, many participants constrain the values that can be input into the database using pre-populated dropdown menus and complex validation rules:

> *"We have data validation, so when she's typing something in, if she says someone is 200 years old, it won't let her move on. But [setting up this validation] takes incredible amounts of work."*

However, despite constraining the values that may be entered by data entry workers, many participants still experience significant issues with the accuracy of entered data. To reduce these issues, data entry workers also spend substantial amounts of time and energy doing quality control. Some organizations use double data entry, which requires that each survey be transcribed twice. Alternatively, some participants described using a technique called *"line-listing,"* which involves holding the paper next to the screen and visually scanning the paper and digital data for discrepancies. One participant said that this process can add weeks of delay and that, in general, quality control is another major bottleneck.

Finally, in addition to transcribing data from paper surveys, data entry workers are often also responsible for finding and correcting data collection errors that have been made by the field workers. In some cases, this can be relatively simple, such as correcting spelling mistakes, while in others, the data entry worker is required to locate and communicate with the field worker in question before they can proceed with the data entry. Several participants also described the data entry workers as being *"at the mercy"* of the field workers. Illegible handwriting results in additional challenges for the data entry workers who have to decipher what has been written on the paper. Deciding the best way to resolve these issues is subjective and may result in additional stress for the designers and/or data entry workers.

4. Analysis and Storage: Making Sense of the Data

The reason for investing so much time and effort to gather data from target communities is so that researchers and practitioners can perform analyses to understand their impact, answer research questions and generate reports for stakeholders. Our analysis of this phase of the workflow highlights that it is challenging for organizations to quickly communicate digitized data to geographically dispersed team members for analysis and visualization. In addition, organizations often expend valuable resources to securely store paper surveys even after the data has been digitized.

4a: Communicating digitized data to geographically dispersed stakeholders is challenging

Our findings reveal several challenges that affect the synchronization and communication of digitized data between workers located in different countries. Many participants described that slow or unreliable Internet connectivity frequently delays the transmission of data from field offices in target countries to US-based offices where the analysis is performed. One participant shared that her organization uses software from Tableau to analyze data and create reports for stakeholders. However, since Tableau is relatively expensive, her organization purchased only a single license for the software, and, since the license is located in their Seattle office, the data must be communicated from the Mozambique office to the Seattle office before it can be analyzed. In addition, the paper surveys from which the data has been digitized are in Portuguese, and the Seattle-based worker who performs the analysis does not speak or understand Portuguese. However, she described,

> *"As it stands, you have the Portuguese version that they do data entry in, but the background code is in English. So then when we extract it, we have set up Tableau so that it feeds into the right worksheets in Tableau. So I see only numbers, and because the code is in English, I see that, oh, they have entered this into the tetanus field."*

Finally, after completing the analysis in Tableau, the Seattle-based worker emails a pdf of the results to her colleagues in Mozambique once a month.

4b: It is difficult for organizations to securely store large quantities of paper-based surveys

Many organizations keep all their paper materials for reference and/or safekeeping and, although the materiality of paper makes it easy to use in the field, it is unarguably more challenging to store than digital data:

> *"There are literally rooms where it's just stacked ceiling to floor with old study surveys. And we have to get grants to get new space to store all the old surveys. It's tough. But that's what we do. This is a lot of very personal data.*

I have to spend hours blacking out people's names and personal identifiers. These are kids with HIV."

DISCUSSION

Having presented a close look at paper-digital workflows across the four stages of design, field work, digitization, and analysis, we now synthesize our findings to offer the following takeaways. First, we show that *both* paper and digital materials play vital roles in global development workflows and are necessary if these organizations are to operate effectively in target communities. Second, we show that, in the workflows we examine, the challenges of transitioning data between paper and digital materials reveal opportunities to design new tools to ease the burden of digitization. Third, we aim to increase awareness of the disconnectedness in global development workflows that largely results from geographical, cultural and socioeconomic differences that can only be understood, not eliminated. Finally, we argue that our findings are relevant for development initiatives across domains.

Paper and Computer Supported Cooperative Work

We find that global development work relies on essential affordances provided by both paper and digital materials, and argue that the two *must* coexist if these organizations are to successfully navigate the hurdles posed by poor infrastructure, low connectivity, cultural differences, and other socioeconomic constraints. Paper is cheap, easy to use in almost any environment, and provides stakeholders with visible and material evidence of data collected. However, in contrast to digital data, which can be stored in vast quantities with relative ease, paper materials must be transported by people and stored in warehouses. It is difficult to imagine how any organization would go about storing and navigating tera- or even gigabytes of paper-based data. In reality, many of the complex analyses and visualizations that help organizations to make sense of the data they collect are only feasible if it is in communicable, searchable and mutable digital formats.

However, our findings also show that global development organizations must pay attention to the material properties of digital data, since *"bits cannot escape the material constraints of the physical devices that manipulate, store and exchange them"* [2]. Organizations must purchase and maintain digital devices for their workers, which may introduce additional challenges. Several of our participants had previously tried providing field workers with laptops or tablets to collect data in digital formats. However, in many socioeconomically disadvantaged communities, possessing an expensive digital device could have social implications for field workers' relationships with local communities that may be unfamiliar or suspicious of new technologies. The material value of digital devices also makes them a target for theft, as described (in all seriousness) by one participant:

"In a previous survey that I worked on, it was not safe to send field workers out with tablets. It would be like, "please have them - steal them - I'm taking them around your village, I don't care about them." We couldn't send [field workers] out with technology."

Finally, infrastructural challenges also affect organizations' abilities to communicate and synchronize digital materials. According to Blanchette, people in resource-rich environments may assume that *"digital information can be reproduced and distributed at negligible cost and high speed, and thus, is immune to the economics and logistics of analog media"* [2]. We find that this is certainly not the case for global development organizations in low-resource environments, where intermittent electricity, poor Internet connectivity and costly data transmission may make the communication of digital materials expensive, slow and unreliable.

In summary, the reliance on both paper and digital materials, and the challenges of effectively coordinating the paper-digital lifecycle, suggest a need for further research that explores how to better support the complex and highly collaborative workflows that drive global development.

Opportunities for Design

One goal of our work was to probe participants' reactions to the idea of optimizing surveys for machine-readability and digitization. In particular, we wanted to investigate if survey designers at the top of the workflow may be willing to change their design practices, and potentially perform more work, to simplify the data entry process further down the workflow.

Researchers and practitioners currently view the process of transitioning information from paper surveys to digital formats as a major bottleneck in their workflow. They also expressed frustration with the tools that they currently use to design surveys and are willing to try new alternatives, even if they take time to learn or disrupt their existing workflows (several participants have already spent months searching for viable alternatives to Word). This suggests an opportunity to design new tools that ease the pain of converting digital to paper, and paper to digital. In the context of global development organizations, we found that survey design tools need to be affordable, provide creative control over the appearance of surveys, and be easily accessible by team members in multiple countries. Our prototype design probe (see Figure 1) appears to meet these criteria and provides a starting point on which to base the design of a new tool for creating surveys. In fact, shortly after we completed our study, we discovered that one of our participants had independently used the tool (which is freely available online) to design several surveys for an ongoing project in Tanzania.

Optimizing surveys for machine-readability would undoubtedly impact the entire paper-digital workflow. In the design phase, since designers think in terms of the software they are using, a tool that encourages them to optimize for digitization could shape their thought processes and increase their awareness of the digitization process. In the field work phase, filling in machine-readable surveys could create additional work, since field workers may need to complete surveys more neatly to ensure they are accurately interpreted by software. This, in combination with the visual appearance of the survey, could also make field workers more aware that collected data needs to be digitized. In the data entry phase, the majority of workers' time would no longer be spent typing data from

paper into a computer. Instead, workers would be responsible for scanning the surveys, double-checking the interpretation of critical fields, and transcribing the few data items that are not machine-readable. Several participants also suggested that they would teach the data entry workers to perform simple, immediate analyses on the data so that they could monitor the field workers and identify problems. Thus, optimizing surveys for machine-readability could improve the digitization bottleneck in two ways: first, by automatically interpreting machine-readable data, and second, by making *all* workers more aware of the digitization process.

Cross-Cultural Cooperative Work

Successfully coordinating work processes across multiple geographies and cultures poses several challenges for global development organizations. In particular, the four distinct phases of the paper-digital lifecycle that we examined suggest that these workflows are somewhat disconnected. Remote locations, infrastructural challenges, cultural differences and changing time zones all exacerbate this problem. Field workers are trained to understand the importance of collecting data from target communities, but not the challenges associated with digitizing the collected data. Data entry workers are expected to seamlessly digitize large numbers of paper surveys without understanding the broader implications of their work. Survey designers often do not understand the perspectives of the field and data entry workers, and may have misplaced expectations or not fully appreciate the constraints experienced by their remote colleagues.

We also find that the nature of workers' roles within global, cross-cultural workflows are, to borrow from Massey, subject to *"a highly complex social differentiation. There are differences in the degree of movement and communication, but also in the degree of control and initiation"* [24]. Researchers and practitioners have a higher degree of control than field workers, who have a higher degree of control than data entry workers. This hierarchical structure impacts the resulting attitudes of and relationships between workers in a variety of ways. For example, our findings reveal disparities between the realities of data entry work and the desires and expectations of researchers and practitioners. Our participants frequently expected the data entry process to be seamless, and expressed frustration and confusion as to why data entry workers found it difficult when *"all they have to do is type this stuff in."* They frequently attributed delays in data entry to the workers' backgrounds and attitudes, rather than the challenging nature of the work, with one commenting that *"it turns out data entry in Malawi is not super-accurate."* Another participant commented that when she performed data entry, she would *"turn on techno and move through it."* Though this tactic may work for her, it may not necessarily help data entry workers operating in remote locations, who may be struggling with computer safety issues, intermittent electricity, and poor Internet connectivity.

Although we have limited our analysis of cross-cultural issues to those that specifically affect the paper-digital workflow, our findings highlight rich opportunities for future research that focuses on more fully understanding the impact of culture in global development work. In addition, the differences in backgrounds, experiences and attitudes of culturally diverse workers suggest that it may be beneficial to increase workers' awareness of their cultural and social differences. Although it may not be possible to entirely overcome these differences, our paper contributes to a greater awareness of such challenges, in the belief that a greater awareness can strengthen work processes and practices.

Limitations and Generalizability

Our paper provides a close look at the challenges presented by the relevance of paper and the desirability of digitization in the context of global development workflows. Since our participants were all engaged in development work but drawn from a wide range of organizations and a variety of domains, we argue that our findings are domain-independent, and researchers and practitioners engaged in multi-country, cross-cultural research in any aspect of development could benefit from our work. In addition, by offering a comprehensive examination of the various abstract types of survey questions and how they can be optimized for digitization, we aim to be of help to survey designers as well.

Some of our findings may also hold for collaborative work outside the field of development. Commenting on these, however, lies outside the scope of our paper. We do stress that the workflow challenges that accompany global, cross-cultural research in general are exacerbated when low-resource environments are involved. For example, linguistic barriers can impact non-development work as well. However, when they are additionally accompanied by low literacy and awareness, the situation is much worsened. Finally, we study paper-digital workflows from the perspective of researchers and practitioners responsible for driving global development initiatives. On-going research is exploring additional issues from the perspective of other stakeholders.

CONCLUSION

This paper presents a study of the collaborative practices surrounding paper-digital workflows as enacted by organizations engaged in global development initiatives. We used a mixed methods approach to examine these workflows that span cultures and geographies, organizing our findings according to the different stages of data as it is sought, collected, digitized, and analyzed. We highlight the tensions that arise between the ubiquitousness of paper and the desirability of digitized data, also discussing the inherent affordances of paper and digital materials, and contribute a nuanced understanding of the accompanying challenges and tradeoffs. Taken together, our findings could influence the design of new tools that aim to bridge the gap between paper and digital materials in the context of global development. In addition, our findings will be useful for survey designers, researchers and practitioners interested in global, cross-cultural research and practice.

ACKNOWLEDGEMENTS

This work was funded by NSF Grant 1111433 and a Facebook Graduate Fellowship. We thank all our participants and the anonymous reviewers for their valuable feedback.

REFERENCES

1. Birnbaum, B., Borriello, G., Flaxman, A. D., DeRenzi, B., and Karlin, A. R. Using behavioral data to identify interviewer fabrication in surveys. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '13 (2013), 2911–2920.
2. Blanchette, J.-F. A material history of bits. *J. Am. Soc. Inf. Sci. Technol. 62*, 6 (June 2011), 1042–1057.
3. Bondarenko, O., and Janssen, R. Documents at hand: Learning from paper to improve digital technologies. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '05 (2005), 121–130.
4. Bossen, C., and Jensen, L. G. How physicians 'achieve overview': A case-based study in a hospital ward. In *ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14 (2014), 257–268.
5. Captricity. http://captricity.com/.
6. Chen, K., Kannan, A., Yano, Y., Hellerstein, J., and Parikh, T. Shreddr: pipelined paper digitization for low-resource organizations. In *ACM Conference on Computing for Development* (2012).
7. Dell, N., Breit, N., Chaluco, T., Crawford, J., and Borriello, G. Digitizing paper forms with mobile imaging technologies. In *ACM Conference on Computing for Development* (2012).
8. Dell, N., Crawford, J., Breit, N., Chaluco, T., Coelho, A., McCord, J., and Borriello, G. Integrating ODK Scan into the Community Health Worker Supply Chain in Mozambique. In *International Conference on Information and Communication Technologies and Development*, ICTD '13 (2013), 228–237.
9. Dymetman, M., and Copperman, M. Intelligent paper. *Electronic Publishing, Artistic Imaging, and Digital Typography* (1998), 392–406.
10. Fitzpatrick, G. Integrated care and the working record. *Health Informatics Journal 10*, 4 (2004), 291–302.
11. Fitzpatrick, G., and Ellingsen, G. A review of 25 years of cscw research in healthcare: Contributions, challenges and future agendas. *Computer Supported Cooperative Work (CSCW) 22*, 4-6 (2013), 609–665.
12. Gaver, B., Dunne, T., and Pacenti, E. Design: Cultural probes. *interactions 6*, 1 (Jan. 1999), 21–29.
13. Guimbretière, F. Paper augmented digital documents. In *ACM Symposium on User Interface Software and Technology (UIST)* (2003), 51–60.
14. Gupta, A. *Red Tape: Bureaucracy, Structural Violence, and Poverty in India*. Duke University Press, Durham, NC, 2012.
15. Haller, M., Leitner, J., Seifried, T., Wallace, J. R., Scott, S. D., Richter, C., Brandl, P., Gokcezade, A., and Hunter, S. The nice discussion room: Integrating paper and digital media to support co-located group meetings. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI '10 (2010), 609–618.
16. Harper, R., and Sellen, A. Paper-supported collaborative work. *Rank Xerox Research Centre, Cambridge Laboratory* (1995).
17. Heiner, J. M., Hudson, S. E., and Tanaka, K. Linking and messaging from real paper in the paper pda. In *ACM Symposium on User Interface Software and Technology (UIST)* (1999), 179–186.
18. Hinds, P., and McGrath, C. Structures that work: Social structure, work structure and coordination ease in geographically distributed teams. In *ACM Conference on Computer Supported Cooperative Work*, CSCW '06 (2006), 343–352.
19. Hull, M. Documents and Bureaucracy. *Annual Review of Anthropology 41* (2012), 251–267.
20. Johnson, W., Jellinek, H., Leigh Klotz, J., Rao, R., and Card, S. Bridging the paper and electronic worlds: the paper user interface. In *ACM SIGCHI Conference on Human factors in computing systems* (1993), 507–512.
21. Kolko, B. E., Hope, A., Brunette, W., Saville, K., Gerard, W., Kawooya, M., and Nathan, R. Adapting collaborative radiological practice to low-resource environments. In *ACM Conference on Computer Supported Cooperative Work*, CSCW '12 (2012), 97–106.
22. Luff, P., Heath, C., and Greatbatch, D. Tasks-in-interaction: Paper and screen based documentation in collaborative activity. In *ACM Conference on Computer-supported Cooperative Work*, CSCW '92 (1992), 163–170.
23. Luff, P., Heath, C., Norrie, M., Signer, B., and Herdman, P. Only touching the surface: Creating affinities between digital content and paper. In *ACM Conference on Computer Supported Cooperative Work*, CSCW '04 (2004), 523–532.
24. Massey, D. A global sense of place. *Marxism today 35*, 6 (1991), 24–29.
25. Medhi, I., Tewari, A., Jain, M., and Cutrell, E. The Fate of a Digital Slate: Unexpected Issues with Deployment in Rural India. *User Experience Magazine 11*, 2 (2012).
26. Parikh, T., Javid, P., Sasikumar, K., Dhosh, K., and Toyama, K. Mobile Phones and Paper Documents: Evaluating a New Approach for Capturing Microfinance Data in Rural India. In *ACM SIGCHI Conference on Human factors in computing systems* (2006), 551–560.
27. Piper, A. M., Weibel, N., and Hollan, J. D. Introducing multimodal paper-digital interfaces for speech-language therapy. In *ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '10 (2010), 203–210.
28. Ratan, A., Chakraborty, S., Chitnis, P., Toyama, K., Ooi, K., Phiong, M., and Koenig, M. Managing Microfinance with Paper, Pen and Digital Slate. In *International Conference on Information and Communication Technologies and Development (ICTD)* (2010).
29. Rosner, D., and Ames, M. Designing for repair?: Infrastructures and materialities of breakdown. In *ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '14 (2014), 319–331.
30. Rosner, D. K. The material practices of collaboration. In *ACM Conference on Computer Supported Cooperative Work*, CSCW '12 (2012), 1155–1164.
31. Schmidt, K. The critical role of workplace studies in cscw. In *Workplace Studies: Recovering Work Practice and Informing System Design*, University Press (2000), 141–149.
32. Sellen, A. J., and Harper, R. H. *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, USA, 2003.
33. Singh, G., Findlater, L., Toyama, K., Helmer, S., Gandhi, R., and Balakrishnan, R. Numeric Paper Forms for NGOs. In *International Conference on Information and Communication Technologies and Development (ICTD)* (2009).
34. Tabard, A., Mackay, W. E., and Eastmond, E. From individual to collaborative: The evolution of prism, a hybrid laboratory notebook. In *ACM Conference on Computer Supported Cooperative Work*, CSCW '08 (2008), 569–578.
35. Underwood, H., Sterling, R., and Bennett, J. The Design and Implementation of the PartoPen Maternal Health Monitoring System. In *ACM Conference on Computing for Development* (2013).
36. Veeraraghavan, R. Dealing with the digital panopticon: The use and subversion of ict in an indian bureaucracy. In *International Conference on Information and Communication Technologies and Development*, ICTD '13 (2013), 248–255.
37. Wellner, P. Interacting with paper on the DigitalDesk. *Communications of the ACM 36* (1993), 87–96.