

For Official Use Only

J U L Y 2 0 1 1

Getting “More Eyes” into Afghanistan



Experiments in Promoting Indigenous Self-Reporting of Local Conditions and Sentiment



RYAN PATERSON, DEPUTY DIRECTOR,
DARPA ADAPTIVE EXECUTION OFFICE

For Official Use Only

CONTENTS

Background and Purpose: Getting “More Eyes” in Afghanistan	1
Methodology: Exploring methods to gather, store, and display data to support assessment and planning of stability operations.	2
More Eyes Accomplishments: Successfully designed and implemented several crowd-sourcing tools	5
Success of “Broadcast out, SMS in” Concept	6
More Eyes Lessons Learned: The Team experiences	8
Appendix: Lessons Learned in Using SMS Text for Collecting Crowd-Sourced Data	11

Getting “More Eyes” in Afghanistan

Experiments in Promoting Indigenous Self-Reporting of Local Conditions and Sentiment

SUMMARY

Many dimensions of the conflict in Afghanistan would benefit from a reliable and timely assessment of regional (e.g., Province-, district- or village-level) conditions (for example, both ISAF's active COIN strategy and ISAF/GIRoA transition planning require reliable assessments of Provincial stability). Indeed, the essence of COIN lies in the “hearts and minds” of the local population whose conditions and sentiment are notoriously difficult to determine. During approximately one year starting Spring of 2010, the Defense Advanced Research Projects Agency (DARPA) conducted a series of “crowd-sourcing” pilots to learn how a combination of (a) modern communications devices (e.g. phones, smart phones, laptops), (b) culturally appropriate social networking and careful choice of participating sub-populations and goals, and (c) incentive strategies could result in such reliable self-reporting of local conditions and sentiment. This describes these efforts, including methodology, accomplishments, and some lessons learned that should be valuable for future efforts.¹

Assessment of counterinsurgency operations and stability operations requires the ability to gather data on the perceptions and attitudes of the local population. In order to understand trends, these measurements must be consistent and persistent over time. Traditionally, the U.S. agencies engaged in assessing such operations have gathered the necessary data through observation, interviews, and polls. These methods of data collection have been employed for many decades, but are imperfect and impractical for several reasons:

BACKGROUND AND PURPOSE

- 1.** Traditional data collection methods require a significant investment in sensors and people over long periods of time.
- 2.** Using these methods, it is extremely difficult to maintain consistency and objectivity over a long period to develop valid trends.
- 3.** Traditional methods of data collection can influence the data itself. Persons being interviewed or polled can be influenced by the process resulting in biased

Recent anecdotal evidence, including DARPA's experience with the Red Balloon Challenge in December 2009, indicates that it is possible to use social networking techniques to catalyze the local population to generate “white” data useful for assessing stability at multiple levels (e.g., regional, provincial, district, and village). White data is generated spontaneously by the local population in the course of everyday business, untainted by influence of outsiders. The “crowd-sourced” data includes network chatter on events directly relevant to the security and stability of a specific area (i.e., roadblocks, gunfire, crowd-gatherings). It also includes economic indicators, such as commodity pricing, crop yields, public health, and economic well-being. There are several efforts underway in Afghanistan to collect white data, including a number of databases populated and used by multiple organizations for disparate purposes.

Coinciding with this anecdotal evidence of crowd-sourcing potential, advances in mobile and web communications technologies (e.g., Android phones) provide a means for creating targeted social networks quickly and flexibly. An interesting hypothesis is that the prevalence of robust mobile networks together with a motivated population could be leveraged to generate significant quantities of white data that could be used to help assess local conditions quickly and inexpensively.

In March of 2010, at a meeting between USAID leaders in Afghanistan and the Director, DARPA, senior USAID leaders requested DARPA assistance to explore innovative techniques for overcoming the difficulties of collecting and displaying data to assess stability conditions. DARPA implemented the More Eyes program to explore the ability to generate crowd-sourced data, integrate it in a common repository with other white data, and develop innovative techniques to display this data to support stability planning.

¹The work reported in this White Paper was the result of the efforts of many individuals in the extended More Eyes team, including Todd Huffman, David Warner, the entire Synergy Strike Force Team, Peter Selfridge, Ted Dang, and Mark Anderson. Additionally a host of others participated in our efforts in Afghanistan and the U.S., Dale (Muddy) Waters, Kristina (Kina) Wihl, Elia (Cyrus) Adahi, Dennis Gallagher, Greg Fischer and others.

Methodology

Exploring methods to gather, store, and display data to support assessment and planning of stability operations.

MORE EYES

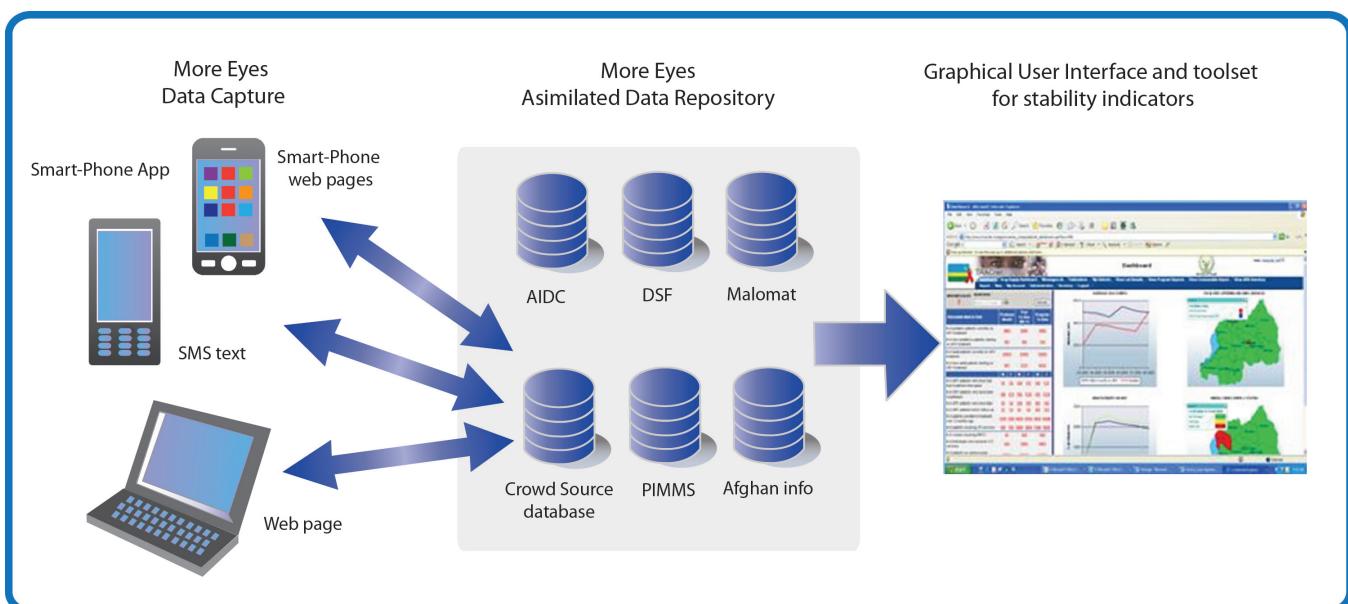
The More Eyes project was designed to explore methods to overcome the difficulties accompanying traditional methods to gather, store, and display data to support assessment and planning of stability operations. The methodology is displayed in Figure 1 below. The capability objectives and goals of More Eyes are aligned along three lines of effort:

- 1.** Increase speed and effectiveness of information collection through novel crowd-sourcing techniques.
- 2.** Assimilate crowd-sourced data with other existing sources of white data into a common repository.
- 3.** Create a graphical user interface for quasi-static data sets and high volume real-time data streams

More Eyes' first challenge was to generate and capture crowd-sourced data. The approach was to catalyze interaction among local residents by enabling and incentivizing the local population to participate in mobile phone and web applications. In theory, these interactions can reveal unbiased raw data that can be used to evaluate stability conditions on the ground. The challenge was to generate this data in formats that could be conveniently collected, standardized, and integrated.

Several pilot projects were developed to catalyze electronic communications among the local population. Each of these pilots was designed to provide prompt data directly from the local population that could be used to understand security and stability conditions in the field. Furthermore, More Eyes intended to leverage the most appropriate mobile communications technologies to elicit such information from remote areas that would otherwise be impossible or impractical to access using traditional methods. These pilots are described in more detail later in this report.

Figure 1: More Eyes Methodology



Pilot projects were designed with wide geographic and demographic overlap. More Eyes used this overlap to create the synergism and mutual-reinforcement to promote a self-sustaining process. More Eyes also identified and integrated existing sources of “white” data that provided useful insights into the security and stability conditions at the provincial and district levels.

The data gathered from crowd-sourcing techniques was integrated with data from existing sources – real-time feeds, if possible – in the More Eyes data repository. This repository consisted of several elements:

- Data Collection Interface (DCI) – Data from the sources described above was captured in a variety of formats, including:

1. Structured – Relational Databases

2. Unstructured – Documents, images, video and audio, and excel

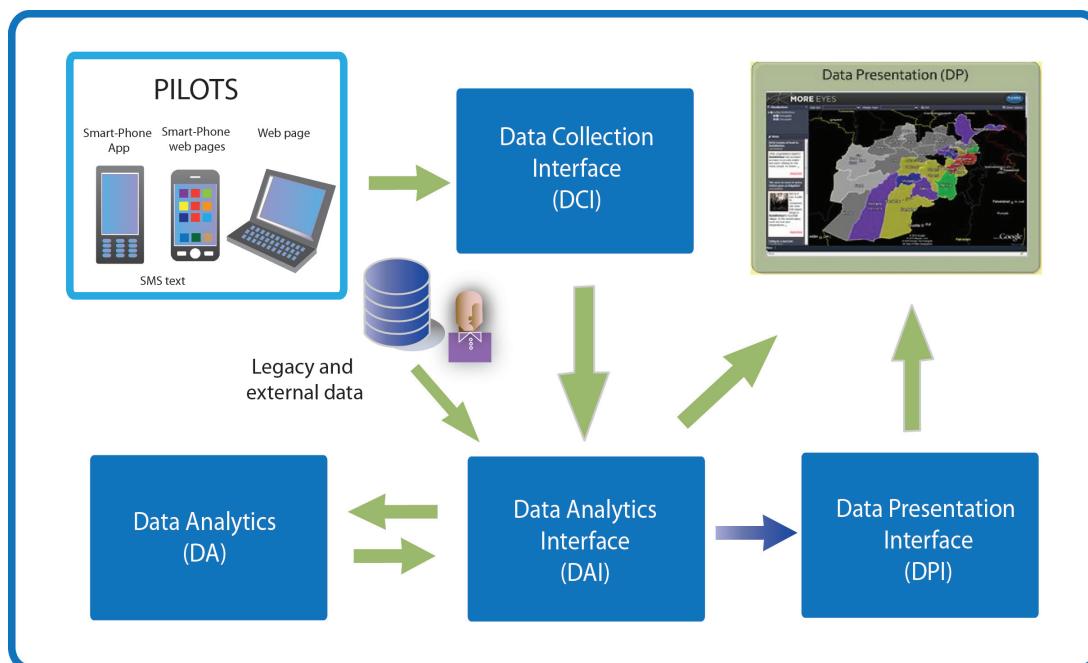
3. SMS and MMS services

- 4.** Open Data Kit services – from mobile apps developed for pilots
- 5.** Google Applications – from mobile apps developed for pilots
- 6.** Emails

To the maximum extent possible, structured data was standardized and sent to a structured Mongo database in the Data Analytic Interface (DAI). Unstructured data was maintained in a separate database in the DAI.

- Data Analytic Interface (DAI) – Once ingested, data was stored to facilitate common access for several functions. The DAI also contained a capability to transform the structured data using data analytic techniques to allow more efficient use for further analysis and display.

Figure 2: Data Repository Architecture



- Data Presentation Interface (DPI) –The DPI provided the interface for analysis and display tools to access the data efficiently. The DPI contained a geoserver, which configured the data for efficient geospatial analysis and display, and a Wildfire application to allow access to unstructured data. Many of the pilots depended on mobile or web display tools to feed data back to the public in a useful format. These tools accessed the data via the DPI.

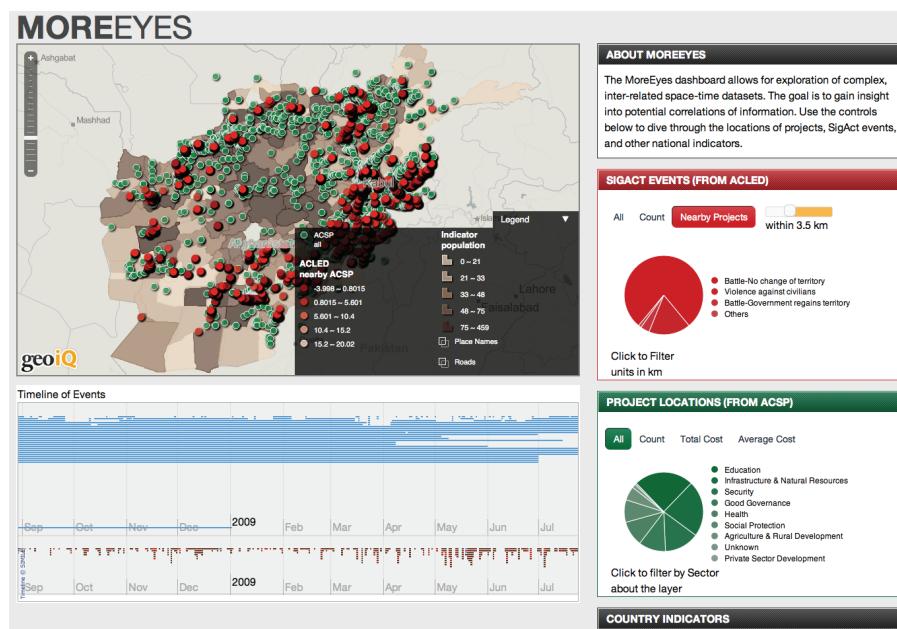
Ultimately, the intent of More Eyes was to generate and provide data useful for stability planning. This required combining data from existing sources in the More Eyes repository with real-time feeds from the pilots and the ability to interactively visualize the data in interesting ways. This visualization capability is the Data Presentation box (DP) in Figure 2, the DP dynamically queried and fetched data from the Data Presentation Interface also shown in Figure 2.

A Ushahidi-type (i.e., map based with an intuitive interface) tool was developed to provide an easy way for planners to visualize relevant data quickly, and to make judgments as to the stability conditions in regions of interest. We designed this tool to provide several capabilities we judged particularly important to do quickly and easily. These included:

- Simple drill-down: the ability to visualize, for example, all the development projects for Afghanistan and then the ability to visualize only those development projects (that meet a particular criteria such as cost or date of completion) that are in a particular district.
- Temporal display: the ability to quickly and succinctly visualize the temporal aspects of multiple data sets together; the temporal displays are particularly important in a dynamic environment.
- Multiple filtering: the ability to use multiple criteria from one data set (for example, the locations of development projects) to filter another data set, for example, SMS messages from a More Eyes pilot.

An initial mock-up version of the Data Presentation is shown in Figure 3.

Figure 3: More Eyes Visualization/Assessment Tool



More Eyes Accomplishments

Successfully designed and implemented several crowd-sourcing tools

SIGNIFICANT ACHIEVEMENTS

The More Eyes team achieved several significant accomplishments. First, More Eyes successfully designed and implemented several crowd-sourcing pilots. These included:

- ConnectJalalabad. Through this pilot, More Eyes helped commercial media partners understand their listenership and better target advertisers. Using a “broadcast out, SMS in” concept described in Figure 4, More Eyes partnered with a local radio station in Jalalabad to conduct a series of publicity campaigns and provided the resources to collect and process the SMS responses. More Eyes collected over 12,000 SMS messages through this pilot. Using this information, the radio station was able to develop a demographic profile of its listeners, and by understanding its listeners’ interests and behaviors, the radio station was better able to target its content and advertising to bolster listener loyalty and involvement. As a result of this pilot, a regional media conglomerate (TV and radio) has expanded this program to multiple stations specializing in women’s affairs and medical issues.
- Atmospherics Program - Afghanistan. More Eyes worked with Defense Intelligence Agency to facilitate data capture, storage and exploitation of data in the Atmospherics Program - Afghanistan (AP-A) initiative. The AP-A initiative uses selected local persons to passively observe and report on things they see and hear in the course of everyday activities. More Eyes provided mobile phones with tailored applications to facilitate their ability to create these reports and communicate them “upwards”. The applications also provided a conduit for posting these reports to the More Eyes common data repository allowing for a more detailed analysis of stability. To this point, over 200 geo-coded multi-paragraph (.25 to 1.5 page) reports have been generated and posted to the More Eyes repository.
- Nangarhar University. More Eyes built on relationships with the Nangarhar University Agriculture Department to promote reporting on the agricultural economy via mobile phones and tailored applications. When faculty, researchers and students traveled to rural areas, they were encouraged to use their mobile phones to send SMS messages reporting on local market prices, areas of agricultural land usage, and farmers’ concerns and issues regarding agricultural activity. This data is currently being published on a Ushahidi-type display at <http://nuagriculture.crowdmap.com>.

- Medical Clinics/Midwives. More Eyes also partnered with the medical community in Nangarhar Province to provide midwives with the ability to efficiently request consultation and report on women’s health issues in the rural areas they service. Midwives were provided with mobile phones containing preloaded applications that allowed them to access necessary medical information, communicate directly with experts at the clinics, and provide data on the medical conditions and issues in remote areas.

Each of these pilots had a specific focus and unique challenges. However, the More Eyes project deliberately created overlaps among these pilots in order to generate synergies by capitalizing on the advantages of each.

For example, techniques developed for ConnectJalalabad – broadcast out, SMS in – were applied to the Nangarhar University agricultural pilot. By partnering with a local not-for-profit radio station and the Ministry of Agriculture, Irrigation, & Livestock (DAIL), More Eyes conducted open-ended inquiries of farmers’ needs in Nangarhar Province. The radio station aired public service announcements and farmers texted responses via SMS. This created a unique and useful information loop that enabled both agribusiness and DAIL’s ability to meet farmer’s needs. More Eyes also fostered a sustainable relationship between the radio station and Nangarhar University in order to create similar efforts in other regions in Afghanistan.

In a similar example, More Eyes created synergistic overlap by partnering ConnectJalalabad with the medical community in Nangarhar Province. Working with the Ministry of Health, More Eyes partnered with a radio station to broadcast an announcement asking the public to report any medical symptoms they may be experiencing using SMS.

This technique served as another tool in the Disease Early Warning System (a Department of Health initiative) and will help improve the ability of the Department of Health to quickly track and react to spreading illnesses before they become epidemics. More Eyes is working with the Ministry of Health to expand this public service nation-wide.

In addition to generating crowd-sourced data from these pilots, the More Eyes team identified, accessed, and integrated several other sources of white data generated and maintained by a variety of organizations in Afghanistan. These included:

- Afghan Infrastructure Data Center (AIDC) – AIDC is a USAID-sponsored database, managed by International Relief and Development, Inc., (IRD) designed to provide geospatial data on donor-sponsored development projects throughout the country. (<http://www.aidc.af/aidc/index.html>)
- District Stability Framework (DSF) data – DSF is a USAID-developed process to elicit information related to stability from the local population.

- AfghanInfo – This database is managed by the Afghan government Central Statistics Organization and provides statistical information about the people and economy of Afghanistan. (<http://www.cso.gov.af/afghaninf.html>)
- The Afghanistan Country Stability Picture (ASCP) – The ASCP project is an initiative led by NATO's International Security Assistance Force (ISAF) to develop and maintain a comprehensive geographic database of reconstruction and development activities across Afghanistan.
- Malomat - Launched by Roshan, in partnership with Mercy Corps and USAID's IDEA-NEW project, Malomat is a national price information system providing farmers, input suppliers,

Figure 4: “Broadcast out, SMS in” Description



Success of “Broadcast out, SMS in” Concept

The “Broadcast out, SMS in” (see Appendix 1) concept leverages the reach of the media organization to broadcast a common stimulus, and the ubiquity of mobile networks to allow individual responses. A partner organization, such as a retailer or a government ministry, develops a carefully crafted stimulus to elicit specific information from the radio station’s listeners. Once it is broadcast, listeners respond using structured SMS text. Analysis of the responses can provide valuable demographic and understand listeners’ interests and preferences. For the partner organization, this information that can be used to improve shape content, bolster brand loyalty, and target advertising.

Initially, the More Eyes team was told by experts in theater that they should not expect local Afghans to generate structured SMS text messages as a part of a crowd-sourcing pilot. Contrary to this advice, the More Eyes team implemented the “Broadcast out, SMS in”, which resulted in over 12,000 responses to date and is the most successful technique for generating crowd-sourced data.

traders, and wholesalers with access to commodity information in 11 provincial wholesale markets using a mobile phone, either through Short Message Service (SMS) or Interactive Voice Response (IVR) technology. (http://www.roshan.af/Roshan/Roshan_Community/Work/Communities/Malomat.aspx)

- Provincial Infrastructure Management Support System (PIMSS) - PIMSS provides the Government of the Islamic Republic of Afghanistan (GIRoA) with a fully intergraded and GIS-enabled decision support system to strengthen development planning, coordination and decision-making across the full spectrum of government. (<http://www.pimss.af/About-PIMSS.html>)

The More Eyes program designed and built a comprehensive data repository that contains all of the crowd-sourced data generated through the pilots, as well as allowing access

to the other sources of white data described above. Enabling convenient access to all of this information from one location provides coalition assessors and planners with a valuable and convenient tool for planning development and stability activities. The repository is portable and scalable, allowing users to use it in other theaters with and for different purposes, such as disaster response.

Through the More Eyes program, DARPA developed relationships with several relevant organizations in theater, including, USAID, DIA, multiple GIRoA Ministries and local government organizations. To foster and promote these relationships, DARPA established the DARPA Forward Cell (DFC) at Camp Julien, just outside of Kabul, to facilitate a continuous presence in the theater. The DFC provided DARPA personnel in theater with a robust base of operations, including communications, computing and billeting.

Figure 5: Members of the More Eyes team eating with Local Media Leaders



More Eyes Lessons Learned

The team experienced some successes and encountered multiple unplanned difficulties.

SUCCESS AND DIFFICULTIES

The More Eyes team experienced some successes and encountered multiple unplanned difficulties. These lessons learned fall into three general categories: dealing with the Afghan population, dealing with US and Afghan bureaucracies, and dealing with the data.

Each of the pilots required extensive personal interaction with the local Afghans in a culturally sensitive way. Recognizing this early on, DARPA contracted with several American personnel who had significant experience in working directly with the Afghan people on the ground in the Jalalabad area. More Eyes leveraged the existing contacts these personnel had at Nangarhar University and the Jalalabad medical community. Furthermore, their extensive experience in building personal relationships with local Afghans was a key factor in gaining access to telecom operators and radio broadcasters. This lesson learned cannot be overstated: experienced and culturally savvy personnel were a critical enabler for implementing the More Eyes pilots.

Successful pilots focused on the incentive to the local population, not the quality or the relevancy of the data to be harvested for stability planning. A successful pilot must focus first on getting the data to flow – getting the population to participate in any way – using whatever incentives are available. Once there is momentum in data flow, then the messages can be tweaked, if necessary, to elicit more relevant information.

Implementing successful pilots required local institutions to see value and “take ownership” of the process. To build these relationships and engender motivation, it was best to be very generous up front – to give first with no explicit conditions. ConnectJalalabad provided an example in point. The More Eyes team approached the radio station managers with a focus on their interests – increased listener loyalty and involvement, ability to determine listener interests and behavior, and the ability to collect valuable demographic data so that broadcasters’ advertisers can better target audiences.

More Eyes personnel installed servers to collect and process SMS texts at no cost to the station. As a result, the radio station became very motivated to make ConnectJalalabad work and the pilot provided the More Eyes team with access to a very large demographic group that generated a multitude of SMS responses.

Figure 6 shows some of the data that was gathered through the ConnectJalalabad pilot. In its initial stages, the objective of ConnectJalalabad was to exercise the “broadcast out – SMS in” concept and to improve the Afghan public’s proficiency in receiving a request for information and providing feedback. These graphics show the combined results of two events. The first was a sports trivia contest. The second was a popularity contest in which listeners were asked to SMS their favorite movie star, poet, cricket player, author, Member of Parliament, and singer. The chart on the left shows how the number of responses grew over time as the contest caught on. The chart on the right displays the distribution of responses from the various service carriers in Jalalabad. These results indicate that the “Broadcast out- SMS in” concept does have broad and diverse appeal.

Figure 6: Data from ConnectJalalabad Pilot

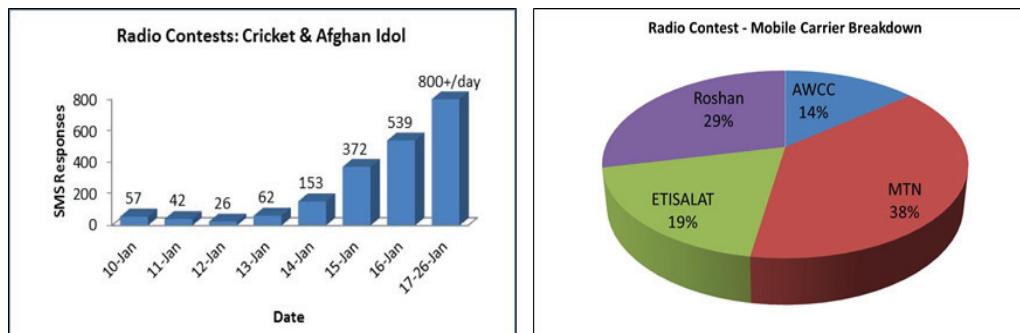


Figure 7 shows the results of the “Afghan Idol” contest. The topic was chosen to elicit the greatest expected response from the public. This pilot demonstrated the public’s ability to respond via SMS to a broadcast stimulus. Furthermore, it established a repeatable baseline for similar events in the future. Radio stations can now collaborate with partner organizations, such as government ministries, universities, and other business interests, to frame the subject and formulate a question that will elicit information more applicable to stability assessments. The SMS resources that More Eyes provided for the radio station have significantly impacted the broadcaster’s media sophistication and ability to engage listeners. This radio station and others are now developing future events to further refine these techniques.

Enabling the local population to use mobile and web technologies required special considerations. First, mobile applications must be simple to use and in the local language (Dari or Pashto). Furthermore, depending on the community of interest, many of the local Afghans using the phones to provide crowd-sourced data may be illiterate. In these cases, phone applications with visual interfaces may be needed. This may sound obvious in hindsight, but it was never a trivial issue. The combination of technical skill and cultural knowledge necessary to effectively respond to pilots like ours should not be taken for granted.

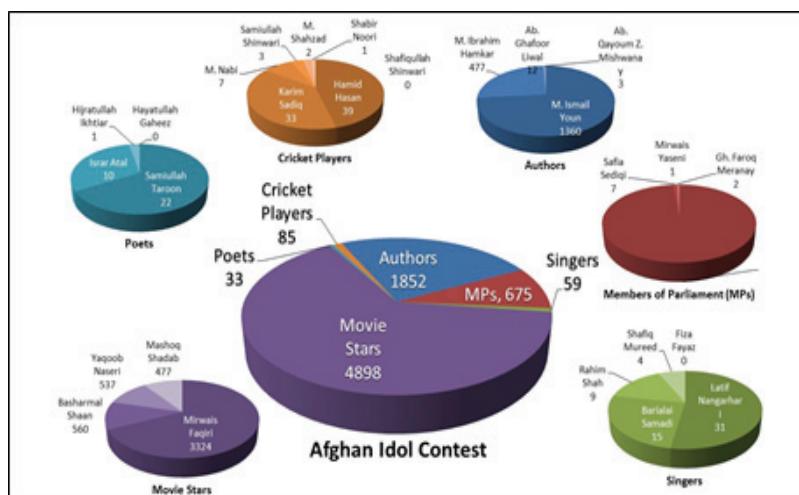
Security and discretion are also very important when working with the local population. It is important to ensure that any phones provided by the More Eyes program are simple and commonplace in Afghan society – preferably procured locally. Phones that stand out due to their appearance or advanced functionality can be an indicator of collusion with foreigners and can invite threats from local insurgents. Any mobile

phones supplied to the local Afghans should be acquired locally and the necessary applications must be simple and fast. Furthermore, phones provided by the More Eyes team should have a simple and quick feature to delete applications and data both locally and remotely. This will maximize user safety should the phone be lost or otherwise compromised.

The hope for More Eyes was that the pilots would grow to a point where the data flows would become self-sustaining. None of the More Eyes pilots reached this tipping point, although some have shown potential to do so. With more time and resources, this is an achievable goal. By their very nature, social networks are dynamic, evolving entities. Even after data flow becomes self-sustaining, operators will still be required to monitor and adjust activity to ensure that it continues to provide useful data.

Web and mobile communications systems limitations in Afghanistan constrained the design and implementation of pilot architectures. At the beginning of the project, the More Eyes team designed pilots based on what they believed to be very capable and ubiquitous web and mobile phone networks. The reality was quite different. The More Eyes team quickly learned that only 4% of the population had access and skills necessary to exploit the internet. Rural populations had even less, which effectively precluded any pilot that required web interaction. Also, the More Eyes team overestimated the capability of mobile services in Afghanistan. Mobile data capability varies from region to region across Afghanistan. Most areas outside of large cities, where most of More Eyes pilots were conducted, had limited telecommunications capability and could only support voice and SMS text communications. There were several challenges with ingesting, standardizing, and analyzing existing and crowd-sourced data:

Figure 7: Results from the “Afghan Idol” Radio Contest



- Sparseness/Uncertainty – Crowd-sourcing is most viable when there is a broad network of contributors that cover the demographic and geographic range of interest. One of the virtues of crowd-sourcing is that, in an uncertain environment, multiple data points viewed in aggregate will provide a more accurate, more precise picture than any single observation, even if the single observation is from a highly informed and capable source. Multiple data points with overlapping geospatial-temporal coverage allow analysts to develop correlations with high confidence, allowing them to imply and predict local stability conditions. With the possible exception of ConnectJalalabad, none of the pilots generated sufficient data to allow for such a robust analysis. Because of the limited data generated by the More Eyes pilots, analysts would not be able to exploit this capability.
- Applicability – Crowd-sourcing generated a lot of unfocused and superfluous data. By its nature, data generated from crowd-sourcing techniques does not necessarily apply directly to a particular information need tied to stability conditions. Thus, while some of the crowd-sourced data relates directly to evaluation of stability conditions, much of it is not directly useful for this purpose.
- Continuity – Stability operations are necessarily a multinational, interagency process. The More Eyes team encountered many obstacles in working among these bureaucratic stovepipes. Convincing these organizations to simply share their existing data was a huge challenge. Even USAID, as one of the ultimate beneficiaries of the More Eyes program, was reluctant to provide the liaison necessary to make More Eyes a success. Nevertheless, it is important to have a consistent team on the ground to interact with relevant organizations, such as GIRQA Ministries, USAID, ISAF, and DIA. The consistent and persistent presence enabled by the DARPA Forward Cell (DFC) was a critical enabler for developing and maintaining engagement with these organizations. These relationships provided the foundation for cooperation not only for More Eyes, but for several other DARPA projects in theater.

CONCLUSIONS:

The More Eyes concept was a high risk/high payoff program that resulted in a combination of successes and failures. More Eyes showed that crowd-sourcing is a viable and relatively inexpensive way to acquire unbiased, raw data that can support stability planning. However, it is not a simple or unambiguous capability to implement in the field. The end-to-end process of conceiving of pilots, developing applications and architectures, and working with local institutions to implement them resulted in many lessons learned. These lessons learned will provide continuity as DARPA transitions the management of the More Eyes program to other organizations.

A relatively immature business environment exists in Afghanistan, requiring the More Eyes team to spend a lot of time laying the foundation and educating the key players on the business process. Business processes and culture common in the developed world, such as cooperation toward a mutual benefit, simply did not exist, even when the More Eyes team provided all of the necessary equipment and expertise. For example, the ConnectJalalabad pilot asked the Afghan telecoms to cooperate through the Afghan Telco Regulatory Agency (ATRA) to provide common SMS text numbers, allowing responses to be collected in one database regardless of the carrier. ATRA was very slow in accommodating this (apparently) unprecedented request, causing a significant barrier to implementing the pilot. In another example, the customs system in Afghanistan was slow and uncertain. The More Eyes team found it extremely difficult to ship mobile phones to Afghanistan via commercial channels. Eventually, they came to rely on couriers and military transport to avoid the bureaucracy of the customs system. Afghanistan lacks the business culture that is taken for granted in the developed world. Any future crowd-sourcing projects should identify and address these issues early on.

Even in this environment, however, More Eyes was successful at initiating crowd-sourced data flows through its pilots. The applications and architectures put in place to support the pilots began to bear fruit in the form of collected data. When developing social networks, success depends on a technically skilled and culturally knowledgeable cadre. Technical capability alone was not sufficient to implement these pilots. Furthermore, success requires a persistent presence on the ground to build and maintain the relationships with local institutions. This cannot be done remotely. The More Eyes pilots did produce some useable data, but they did not achieve the desired volume of data necessary to support stability planners. Although the pilots are on the right track, additional time and resources would be needed to reach this goal.

The More Eyes project successfully created a repository for capturing, storing and integrating white data and a visualization tool useful to stability planners. Although these tools were developed for the Afghanistan Theater of Operations, they were designed to be scalable, portable and adaptable for potential use in other theaters and for operations other than COIN (i.e., Foreign Humanitarian Assistance, consequence management). Given the time and resource constraints for More Eyes, these tools were never exercised or tested outside of Afghanistan.

Appendix 1 – Lessons Learned in Using SMS Text for Collecting Crowd-Sourced Data

Summary: The More Eyes project explored the utility collecting crowd-sourced data using SMS text. SMS texting is an attractive method because it is widely available and relatively simple to use. On the other hand, SMS is unstructured text. Unless a degree of structure is imposed, each individual will decide how to write his message. Unstructured SMS messages can be very difficult to deal with on the scales required for crowd-sourcing.

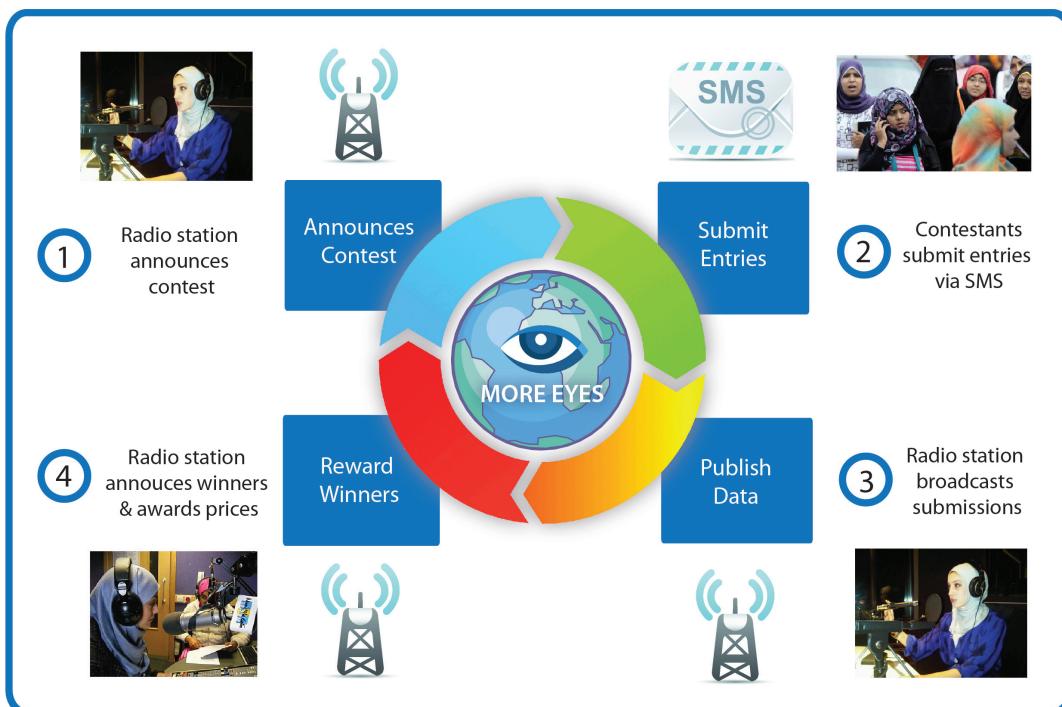
The More Eyes team devised and tested several methods for imposing different levels of structure. Two of these experiments were done through radio contests, and one was done to support a group of midwives to provide medical support to women in remote areas. The results were instructive.

Radio Station Contests: The More Eyes team worked with a local radio station to evaluate the proficiency of the Afghan public to use SMS, and to assess the types of data that they would provide. The radio station held week-long contests & opinion polls to stimulate the public and to assess its proficiency in submitting information via SMS. The concept was for

the radio station to broadcast a question asking its listeners to respond via SMS text. Figure 1-1 illustrates this process. Ultimately, the objective was to integrate these SMS text responses into the More Eyes data repository to support assessment of regional stability. The radio station conducted two contests. One centered on sports trivia (cricket) where the winners of the contest were given cell phone minutes. The other contest was a popularity contest called “Afghan Idol” where the winners were recognized at a ceremony at the end of each week. Each of these contests asked for a different level of standardization in listeners’ responses.

Cricket Trivia Contest: Each day for one week the radio station broadcasted a new trivia question related to Cricket, asking the public to answer via SMS text, and promising mobile minutes to those who answered correctly. While the contest started out slow, a change in format significantly increased the number of respondents. The change in format not only yielded more useful data, but also a much higher number of SMS responses. The history is instructive:

Figure 1.1: “Broadcast out, SMS in” Concept



- Early in the contest, the radio station broadcast the following question (translated from Pashto):

"On International Cricket Race Day, which player made 60 runs?

Mohammad Shahzad
 Karim Khan Sadeq
 Asghar Estanekzai
 Hamid Hussan

Dear listeners you can send the right answer to this number 0777601412 from now till 7:00 pm and the result of your answer will be announced from 7:30 pm till 8:00 pm."

The answers were not labeled by letter (a, b, c, or d), and the question did not ask for responses to be put into a specific format. As a result, the responses were completely unstructured. Some respondents implied a letter label to their response, but most wrote out the full name (with multiple spellings). Also, most respondents felt the need to identify themselves by name and where they live. Cleaning up this data for analysis required significant resources and probably even created additional errors. Table 1-1 is a sample of the responses to this question.

Table 1-1: Sample Results from the Cricket Trivia Contest with no Choice Labels 2

ID	Sender	Date Time	Message
151	excised	1/4/2011 19:36	Asghar im pareena
148	excised	1/4/2011 19:06	SALAM PARWEZ WAMA OALA SPA GEZA KAREM SADEQ WAHALE DA.
143	excised	1/4/2011 18:48	KAREM KHAN SADIQ 6 DA SCOTLIND KHLAF .AMANULAH.KAMAWAL
141	excised	1/4/2011 18:43	asghar stanekze iem majeedullah
137	excised	1/4/2011 18:05	B
105	excised	1/4/2011 15:05	Hamid aminzai Karim khan sadeq
101	excised	1/4/2011 12:10	E
98	excised	1/4/2011 11:51	PA YW WRAZMY SYALYO K LOMRY 6 KREM SADIR WALY , HABIB QASAB
74	excised	1/4/2011 8:56	//Hamed hasan farida

- By the fifth day, the radio station began to label the choices by letter (a, b, c, and d):

"Answer the following question about the cricket race, which were held in South Africa. Which player got the most runs?

- a) Nowrooz Mangal
 b) Mohammad Nabi
 c) Hamid Hassun
 d) Mirwise Ashraf

You can send the right answer to this number, 0777601412 from now till 7:00pm."

Table 1-2: Sample Results from the Cricket Trivia Contestwith Choice Label

ID	Sender	Date Time	Message
677	excised	1/8/2011 12:52	Merves aentezar (B)mohamdnabe Sapa bndar
643	excised	1/8/2011 12:23	D :Merwise Ashraf this message has been sent by Aimal Hairan
641	excised	1/8/2011 12:21	D
613	excised	1/8/2011 11:52	Atiq hashmi d
602	excised	1/8/2011 11:41	D
601	excised	1/8/2011 11:41	B.miras.ashrap.'..ibrahim.salarzy.kunr
598	excised	1/8/2011 11:38	C
594	excised	1/8/2011 11:34	(MERWAIS ASHRAF) (ZAMA NOME .QISMATULLAH)
591	excised	1/8/2011 11:29	D
589	excised	1/8/2011 11:27	D
576	excised	1/8/2011 11:16	D hikmat ullah
577	excised	1/8/2011 11:16	C

By labeling the choices (a, b, c, and d), the question implied responses could be submitted by letter. As a result, many responses contained the letter of their choice, which significantly simplified data standardization. This also resulted in a significantly higher number of responses, probably because it made it easier to respond. Nevertheless, many responders still felt compelled to provide their name and location. Table 1-2 below shows a sample of the responses to this question.

The experience of the sports trivia contest pointed to the need for the broadcast itself to specify the precise response format desired. Building on this experience, the Afghan Idol contest asked for specifically formatted responses.

- **Afghan Idol Contest:** In this contest, the radio station asked the public to vote for the best candidate in the following categories: author (au), actor (a), cricket player (cp), singer (s), member of Parliament (mp), poet (p), district governor (dg), governor (g), and youth leader (yl). The radio broadcast asked listeners to respond using a very specific format:

In celebration ceremony of its second birthday, once again, Best Author, Best Poet, Best Cricket Player, Best Movie Star, Best Singer of the year and best previous Member of Parliament from Nangarhar,are selected by having specific voting codes.

All of Safa Radio listeners can SMS the codes of their favorite nominees from now until the celebration day, and select the personalities by their own votes.

Send your SMS to....
0777 601 412

For the most part, respondents followed the instructions in the broadcast, but there were still some inconsistencies. Some respondents used periods, underscores, or slashes where the instructions directed them to use a space. Like the sports trivia contest, some appended their name, location, or phone number. One even inserted an emoticon (:-)). Despite these issues, the data was generally much easier to preprocess for integration and further analysis. Table 1-3 shows a sample of the results.

Table 1-3: Sample Results from the Afghan Idol Contest

ID	Sender	Date Time	Message
2335	excised	1/22/2011 14:48	A 2011 2
2670	excised	1/23/2011 7:44	P 2011 1
2680	excised	1/23/2011 9:13	p 2011 1
12226	excised	1/20/2011 15:09	mp 2001 4
12215	excised	1/20/2011 14:55	Mp 2011 4
12203	excised	1/20/2011 14:38	::-)au.2011.:
12196	excised	1/20/2011 14:32	mp 2011 4
7131	excised	1/17/2011 6:46	0777601412MP 2011 4
6336	excised	1/16/2011 11:14	s 2011 2
6333	excised	1/16/2011 11:12	Cp 2011 3
6325	excised	1/16/2011 11:07	cp 2011 3
6305	excised	1/16/2011 10:51	s 2011 1
5501	excised	1/15/2011 9:23	Cp_2011_4
1329	excised	1/21/2011 19:12	My name is Aimal. Cod number.CP 2011 .4
9383	excised	1/18/2011 13:39	CP/2011/7

Midwife Pilot: The More Eyes team leveraged its contacts in the medical community to implement a crowd-sourcing pilot focused on midwives. In eastern Afghanistan, midwives provide key medical care to women in remote areas who do not have access to other medical facilities. The concept was to provide midwives with the mobile phone technology and applications to report via structured SMS text the health conditions they encounter. The mobile phones also provided a means to consult directly with doctors in their home clinics. In aggregate, these communications could provide valuable information on the state of women's health – and thus regional stability – in these remote areas.

Since midwives are generally more literate and well educated than the general public, it was possible to use a more complex method to elicit information. Midwives were asked to send highly structured SMS texts – a sequence of numbers that represented the answers to a group of predetermined questions. Figure 9 shows the structured checklist (translated from Pashto) that midwives were asked to follow in transmitting their SMS texts.

Table 1-4 shows a sample of the results of the Midwife Pilot. The SMS texts were very structured, allowing for easy preprocessing and ingestion into the More Eyes data repository.

How are you coping:			
<input type="radio"/> =Excellent	<input type="radio"/> =Good	<input type="radio"/> =Fair	<input type="radio"/> =Poor
How many normal deliveries did you perform in a home this month?			
Enter number amount.			
How many normal deliveries did you perform in a clinic this month?			
Enter number amount.			
How many complicated deliveries did you perform in a home this month?			
Enter number amount.			
How many complicated deliveries did you perform in a clinic this month?			
Enter number amount.			
Did you use a partograph?			
<input type="radio"/> =Yes	<input type="radio"/> =No	Did you perform post partum care?	
<input type="radio"/> =Yes	<input type="radio"/> =No	Did you perform ante-natal care?	
<input type="radio"/> =Yes	<input type="radio"/> =No	Did you perform post natal care?	
<input type="radio"/> =Yes	<input type="radio"/> =No	Did you provide any of the following family planning services? If so, which ones?	
<input type="radio"/> =Condoms	<input type="radio"/> =Pills	<input type="radio"/> =Injection	
<input type="radio"/> =All These	<input type="radio"/> =None of These	<input type="radio"/> =Other	
Did you refer any women to Jalalabad or Asadabad hospital?			
<input type="radio"/> =Yes	<input type="radio"/> =No	Were any women referred to you?	
<input type="radio"/> =Yes	<input type="radio"/> =No	Did you provide vaccinations this month?	
<input type="radio"/> =Yes	<input type="radio"/> =No	How many maternal deaths did you experience this month?	
Enter number amount.			
How many still births did you experience this month?			
Enter number amount.			
Send SMS to +93 77 008 6311			

Figure 1.2: Midwife Pilot SMS Checklist

Table 1-4: Results of Midwife Pilot

ID	Sender	Date Time	Message
23827	excised	5/17/2011 12:03	15 1 14Feb08 31Dec07 1 3 2 3 2 0 0 0 2 0 0 0 0 2 0 0 0 2 1 2 , 2 , 2 , 1 , 2 2 2 2 1
23825	excised	5/17/2011 11:44	1 2 1 21Mar010 29Dec05 7 , 8 2 2 1 , 3 , 7 , 8 2 0 0 0 2 0 0 0 0 2 0 0 0 2 2 2 , 2 , 2 , 2 , 2 2 2 2 1
23604	excised	5/15/2011 12:27	1 2 2 27Mar010 1Der09 2 , 4 2 2 1 , 2 , 3 , 4 , 8 2 0 0 0 2 0 0 0 0 2 0 0 0 2 2 2 , 2 , 2 , 2 , 2 2 2 2 1
23603	excised	5/15/2011 12:20	1 2 1 12Oct09 18Dec08 2 , 4 2 2 1 , 2 , 4 , 8 2 0 0 0 2 0 0 0 2 0 0 0 2 2 2 , 2 , 2 , 2 , 2 2 2 2 1
23601	excised	5/15/2011 12:02	1 2 1 21Mar010 19Mar06 2 , 4 2 2 1 , 3 2 0 0 0 2 0 0 0 2 0 0 0 1 2 2 , 2 , 2 , 2 , 2 2 2 2 1
23599	excised	5/15/2011 11:50	1 2 1 27Mar010 1Dec09 1 , 3 3 2 1 , 2 , 4 , 8 2 0 0 0 2 0 0 0 0 2 0 0 0 2 2 2 , 2 , 2 , 2 , 2 2 2 2 1
23480	excised	5/14/2011 12:29	1 2 1 16Jan11 30Dec10 2 , 4 2 2 1 , 2 , 3 , 4 , 8 2 0 0 0 2 0 0 0 0 2 0 0 0 1 1 2 , 2 , 2 , 1 , 2 1 1 2 1
23478	excised	5/14/2011 12:20	1 2 1 16Jan11 30Dec10 1 , 3 4 2 1 , 4 , 8 2 0 0 0 2 0 0 0 0 2 0 0 0 2 2 2 , 2 , 2 , 2 , 2 1 2 2 1
23477	excised	5/14/2011 12:05	1 2 1 27Mar10 1Dec09 1 , 3 3 2 1 , 3 , 4 , 8 2 0 0 0 2 0 0 0 0 2 0 0 0 2 2 2 , 2 , 2 , 2 , 2 2 2 2 1
23473	excised	5/14/2011 11:54	1 2 1 1Apr10 1Dec09 4 , 8 2 2 6 2 0 0 0 2 0 0 0 0 2 0 0 0 2 2 2 , 2 , 2 , 2 , 2 1 2 2

SUMMARY AND CONCLUSIONS:

The More Eyes team learned several things from these pilots:

- 1.** Because SMS text widely available and relatively simple to use, it is a valuable tool that allows the local population to communicate quickly.
- 2.** Because SMS is so simple, individuals will respond in a unique formats – unless a standard format is imposed – making data integration and follow-on analysis very difficult and time consuming.
- 3.** Depending on the type of data being collected, it is possible to impose different levels of standardization on the SMS responders. The More Eyes team was able to impose structure using two methods:
 - By encouraging respondents to use in a very specific format (Afghan Idol Contest) .
 - By forcing SMS texters to use a structured checklist (Midwife Pilot).

While some respondents did not always follow the instructions exactly, using these techniques resulted in “cleaner” data that was much easier to preprocess and integrate into the More Eyes data repository.

- 4.** The number of responses rose significantly when the initial broadcast asked listeners to respond in a standard format. Although there are other possible explanations (i.e., learning), this correlation could imply that listeners were reluctant to respond when there were fewer rules, but were more comfortable responding when told exactly how to do so.
- 5.** An educated group, such as midwives, is capable of providing very structured and detailed data when given specific instructions.
- 6.** Most significantly, the More Eyes team demonstrated that it was possible to get a large group of people to provide structured data via SMS, even in Afghanistan, where the majority is illiterate.

