Applying Information and Communications Technologies to Collect Health Data from Remote Settings: A Systematic Assessment of Current Technologies

Raj Ashar[a,*], Sheri Lewis[a], David L. Blazes[b] and JP Chretien[c]

[a] The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA

[b]Armed Forces Health Surveillance Center, Division of GEIS Operations, Silver Spring, MD, USA

[c]Division of Preventive Medicine, Walter Reed Army Institute of Research, Silver Spring, MD, USA

*Corresponding Author

Fax: +1 443 778 9188
E-mail address: Raj.Ashar@jhuapl.edu (R. Ashar).

The text of this manuscript has been published in the *Journal of Biomedical Informatics*, volume 43 (2010), pages 332-341. The unique identifier for the published article is doi:10.1016/j.jbi.2009.11.009.

Abstract

Modern information and communications technologies (ICTs) are now so feature-rich and widely available that they can be used to "capture," or collect and transmit, health data from remote settings. Electronic data capture can reduce the time necessary to notify public health authorities, and provide important baseline information. A number of electronic health data capture systems based on specific ICTs have been developed for remote areas. We expand on that body of work by defining and applying an assessment process to characterize ICTs for remote-area health data capture. The process is based on technical criteria, and assesses the feasibility and effectiveness of specific technologies according to the resources and constraints of a given setting. Our characterization of current ICTs compares different system architectures for remote-area health data capture systems. Ultimately, we believe that our criteria-based assessment process will remain useful for characterizing future ICTs.

Keywords: public health; disease surveillance; health surveillance; developing nations; data collection; communications networks; telecommunications; mobile technologies;

1. Introduction

The rapid emergence of the novel H1N1 influenza A in the spring of 2009 served as the latest indication that new and re-emerging infectious diseases pose a serious threat to global public health. This threat is compounded by international air travel, which facilitates disease transmission across distant urban centers more quickly than within countries [1, 2, 3]. Since global disease surveillance becomes critical for the early detection of outbreaks [2], it is mandated by international law: the revised International Health Regulations, which entered into force in 2007 and are legally binding on 194 countries, specify core national capacities for detecting and reporting events that may constitute a public health emergency of international concern [4]. Nations can enhance their compliance with the IHR, and so minimize the critical time in an outbreak from detection to disclosure, by establishing electronic surveillance systems based on ICTs for health data collection, transmission, and processing.

Though new infectious diseases are disproportionately likely to emerge from resource-limited countries that lack surveillance [5], most electronic disease surveillance systems are deployed in industrialized countries that have greater access to ICTs. Since infectious diseases do not respect political boundaries, all nations stand to benefit from advanced disease surveillance capabilities. That said, special efforts must be directed towards assisting resource-limited countries to implement surveillance tools that are appropriate for their local needs and resources [2, 6]. Capturing health data from remote settings, such as rural areas of resource-limited countries, may prove challenging due to their rugged environment and limited infrastructure. Since the United Nations' 2007 Revision Population Database projects that nearly half of the world's population will still be living in rural areas in 2015 [7], detecting infectious disease outbreaks in those areas will likely compel this challenge to be met.

The collection and transmission of health data from the point-of-entry (PoE) into a central surveillance repository are key functions for the success of any disease surveillance system [2]. Fortunately, technological advances have yielded ICTs which can now be used to capture health data from remote settings. For example, the International Telecommunications Union estimates that cellular phones reached nearly half of the developing world "at the end of 2008 – from close to zero only ten years ago" [8 p. 1]. Additionally, some technologies have been designed specifically for use in remote settings, such as the XO Laptop and satellite phones [9, 10]. Data captured using ICTs can be as, if not more, comprehensive than paper-based health data in two respects. It can be structured, which would allow for the automated submission and analysis of data, as well as the ability to reuse the collected data and combine it with other data sources to more clearly ascertain the situation under surveillance. It can also concisely store patient-level data details.

Several systems have been developed to capture health data from remote areas for disease surveillance [11, 12, 13, 14, 15]. These efforts suggest that no single ICT can satisfy every setting's data capture needs, but rather that different combinations of ICTs may work best in different settings. This article expands upon prior work by defining and applying a general process to characterize ICTs for health data capture across remote settings.

2. Background

A review of remote-area health data capture systems [11, 12, 13, 14, 15] indicated that they share three main components: communications networks (networks), data capture devices (devices), and data capture applications (applications). Networks are the physical communications infrastructure over which data are electronically transmitted, from the PoE to a central surveillance repository. Devices are usually[1] forms of hardware that physically collect data from the user, and transmit data over the network. Applications are software that provide

[1] Software programs can also technically be considered "devices" when they are composed of both applications and the logic to capture data.

user interfaces (UI) to the devices: they determine how the user can interact with a device, and specify how the user-entered data are encoded for transmission over the network. Given the three types of components, the system architecture is the integration of a particular set of component implementations into an end-to-end system that captures health data. Definitions of other key terms and table headings in this article can be found in Appendix A.

Generally speaking, the components relate to each other in three ways. In practice, the circumstances of the local environment and system deployment, such as cost limits, narrow the combinations of ICTs that can potentially be used together. First, the networks that are accessible from a remote setting constrain which devices and applications can be employed to capture data. Second, each application captures data with the same mode of latency, level of detail, and support for world languages across different networks. Third, a device that can communicate over a network typically provides support for all applications which are capable of running on that network. Based on the relationship among the components, a criteria-based process was developed to assess the ICTs available for each component, and help select those ICTs which satisfy the constraints of a given setting. The process is explained below.

3. Objectives

We created a process to characterize ICTs that remote-area health workers (users) could use to capture health data from remote settings. This process focuses on answering two overarching questions:

1. How *feasibly* could an ICT be integrated into a remote-area data capture system?

   Feasibility pertains to the level of resources that are needed for fielding an ICT in remote settings, including infrastructure, costs, and user training.

2. How *effectively* could an ICT be used to capture health data?

   Effectiveness pertains both to the timeliness and comprehensiveness of the data captured, as well as to the ICT's expected performance in the field.

Since the acceptable limits of feasibility and effectiveness will vary by remote setting, this work is meant to simply highlight the questions, tradeoffs, and design decisions that must be considered when architecting a remote-area health data capture system.

4. Methods

4.1. Overview

Each step of the process compares the available implementations for a single system component against a common set of cost and feature criteria. These technical criteria were established based on our domain expertise in computer science and public health, and are defined in Appendix A. To demonstrate how the process can be applied, commodity (commercially- or freely-available) ICTs that existed during the study (April-July 2009) were identified and characterized. Technical specifications were obtained primarily by browsing manufacturer and reseller websites for advertised costs and features.

4.2. Process for selecting feasible and effective data capture architectures

1. Identify the communications network(s) that are accessible in the proposed setting. Characterize accessible networks based on their performance and capabilities.

2. Assess applications supported by the accessible networks against criteria for data reporting and world language support.

3. Determine which devices are capable of communicating over the accessible networks.

4. Compare devices on the basis of operational criteria, such as cost, performance, and additional resource needs. Select those devices that are appropriate for the remote setting.

5. Consider different architectures for remote data capture, given the remaining networks, applications, and devices. The best approach will depend on the data capture environment, funding constraints, and time frame for deployment.

5. Results

5.1. Step 1: Identify and characterize accessible communications networks

Eight categories of networks were identified. These networks are listed by decreasing order of penetration in Table 1, which summarizes each network's level of penetration, features, and different generations (if applicable). Of the eight networks, only the cellular telephone and data network is deployed in varying generations across the world [16, 17]. The satellite communications networks were characterized by one service provider each, because those providers offer the widest level of global coverage. The satellite telephone network was based on the Iridium network [10], and the satellite broadband network was based on Inmarsat's Broadband Global Area Network [18].

Table 2 outlines each network's costs, including requisite hardware, and performance in terms of data speed. Two categories of terrestrial networks offer wireless high-speed Internet connectivity. The mobile broadband network was based on the IEEE 802.16e standard which is being adopted worldwide [19]. The 3G cellular telephone and data network works with a wider range of applications, but its data speed is roughly half that of the mobile broadband network. One other wireless high-speed network (long-range Wi-Fi) is not included because of its nascent stage of maturity [20, 21, 22].

Table 1

Network penetration and features.

Network	Generation	Penetration	Native applications	Extended applications	Comments
Manual delivery	N/A	Global	Physical media	N/A	Listed here to provide a baseline for comparison with electronic communications networks.
Satellite telephone (Iridium)	N/A	Global [10]	Voice, touch-tone, text message	Email, Web	Requires outdoor line-of-sight to satellite
Satellite broadband (Inmarsat BGAN)	N/A	Nearly global [18]	Voice, touch-tone, email, file, Web	VoIP (via streaming IP)	Requires outdoor line-of-sight to satellite
Radio (HF/UHF/VHF bands) [23, 24, 13]	N/A	Regional [24]	Voice	Email, file	In resource-limited settings, tend to be deployed primarily in response to military/NGO needs. Are generally not used by local populations. [24, 25, 26]
Cellular telephone and data (GSM)	All	49.5% [8]	Voice, touch-tone, text message		GSM family of standards is in global use. [27]
	2G			Email, file, Web	Modem hardware is increasingly built-into cell phones.
	2.5G		VoIP (latency possible), email, file, Web	N/A	
	3G and higher		VoIP, email, file, Web	N/A	
Fixed-line telephone	N/A	14% [8]	Voice, touch-tone	VoIP, email, file, Web	
Fixed-line broadband	N/A	2% [8]	VoIP, Email, File, Web	N/A	
Mobile broadband (IEEE 802.16e) [19]	N/A	< 1% [8]	VoIP, email, file, web	N/A	

Table 2

Network interface equipment, costs, and performance.

Network	Generation	Data Communications Hardware	Data hardware cost	Maximum Data Speed	Recurring Service Costs
Manual delivery	N/A	N/A	N/A	Hours	Cost depends on delivery method
Satellite telephone	N/A	Data kit	$140	9.6 kbps	$630 for 500 minute/1500 text message prepaid card
Satellite broadband	N/A	Satellite terminal	$1000	240 kbps-492 kbps (depending on satellite terminal)	$430 for 500 minute/52.6 MB prepaid plan
Radio	N/A	Radio modem	$200	9.6 kbps over HF radio	$0 [28, 26]
Cellular telephone and data	2G	Analog modem	$40	33.6 kbps	$1.33-$36.94 per month for 25 outgoing calls + 30 text messages*
	2.5G	USB data cable	$20	40 kbps	
	3G and higher	USB 3G adapter	$250	800 kbps	
Fixed-line telephone	N/A	USB modem	$30	48 kbps	$0.29-$32.43 per month for 30 local calls*
Fixed-line broadband	N/A	DSL modem	$50	384 kbps	$6.10-$1877.38 per month*†
Mobile broadband	N/A	WiMax adapter	$300	1500 kbps	(Data not available)

Legend

* Price range derived from the ITU Price Basket data for the 120 countries with the lowest GNI per capita [8].

† Monthly fixed broadband cost varies widely by country. Four countries had an estimated monthly service cost over $1000: Swaziland, Burkina Faso, Cuba, and the Central African Republic.

5.2. Step 2: assess applications supported by accessible networks against data reporting and world language criteria

Eight categories of data capture applications were identified, and are listed in Table 3. Three applications [voice, touch-tone, and Voice over Internet Protocol (VoIP)] are based on the real-time transmission of sounds, and so are the only ones that innately support any spoken language. Though the other five applications can asynchronously transmit data back to the repository, they lack innate world language support because they rely on predefined character sets. However, note that the file and physical media applications can digitally capture audio spoken in any language.

Table 3

List and characterization of data capture applications.

Application	Mode of latency	World Language Support	Formatted Data	Patient-level Data	Explanation
Voice	Real-time	All		X	Will be labor-intensive to extract useful information from voice recordings.
Touch-tone	Real-time	All	X		Requires voice application in order to prompt the user for data.
VoIP	Real-time	All	X	X	Voice-over-IP. Has both voice and touch-tone capabilities. Examples include Skype and SIP softphones.
File	Store-and-forward	Some	X	X	Can store any form of electronic data: text , numerical, audio, images, etc.
Physical media	Store-and-forward	Some	X	X	
Text message	Store-and-forward	Some		X	
Email	Store-and-forward	Some		X	
Web	Store-and-forward[2]	Some	X	X	GUI-based software at the PoE which transmits data over the HTTP or HTTPS protocols.

[2] Classified as store-and-forward latency because data forms can be downloaded and asynchronously populated. However, users can interact with Web applications in real-time if a synchronous network connection is available.

5.3. Step 3: determine which devices can communicate over accessible networks

14 categories of devices were identified, and their support for the respective communications networks is listed in Table 4. Appendix B explains the potential utility of non-traditional devices which, to our knowledge, have not been deployed in remote-area health data capture systems. Two categories of devices were considered on the basis of products from a single vendor. The survey software category is based solely on the Voxiva HealthWatch software, as it is the only known instance of that category [29]. The satellite phone category considers Iridium satellite phones only, because Iridium offers the only globally-accessible satellite phone network.

In addition, laptop computers are split into three device categories for reasons that will be discussed in the next step's results. Finally, although the radio communications network encompasses the high frequency (HF), very high frequency (VHF), and ultra high frequency (UHF) bands, only HF radio was considered as a device category, because the HF band enables communication across the widest distances.

The data in Table 4 indicate that traditional laptop computers have the most versatile communications capabilities because they can connect to all networks, either directly or through additional communications hardware. Device categories that feature both computing and communications capabilities, such as the other laptop categories, personal digital assistants (PDAs), and smartphones, could support fixed-line and mobile networks with additional hardware. Notably, as the mobile broadband network gains broader market penetration, support for it will likely be built-into devices.

Since device support for networks drives which applications can be supported, each device can be classified into one of three groups depending on which latency of data capture it supports. Table 5 shows that three devices or software programs are only capable of real-time data capture, three devices are only capable of store-and-forward latency data capture, and the remaining eight devices are capable of data capture in both modes.

The three "real-time only" options differ from the other 11 in that they synchronously capture data and do not store any data themselves. Moreover, two of the real-time only options are "centrally-based" software programs running at the surveillance repository. That is, the user needs an intermediate device, such as a cell phone or satellite phone, to enter data while connected to the centrally-based software. Being hosted software, these two are the only data capture systems that cannot be damaged or lost in the field. At the other end of the communications spectrum, the three "store-and-forward only" devices serve primarily as storage media because their data can only be transmitted by manual delivery.

The eight devices that capture data in both modes provide the ability to transmit data directly from network-connected remote areas, and the flexibility to store data for manual delivery from areas which lack network access.

Table 4

Device and software support for communications networks.

Device	Manual delivery	Satellite telephone	Satellite broadband	Radio	Cellular telephone and data	Fixed-line telephone	Fixed-line broadband	Mobile broadband
IVR software		X	X		X	X	X	X
Survey software		X	X		X	X	X	X
HF Radio				X				
Satellite phone	X	X						
Cell phone	X				X			
Laptop (Regular)	X	A	A	A	A	X	X	A
Laptop (Resource-limited settings)	X		P		P	P	W	P
Laptop (Netbooks)	X		P		P	P	W	P
PDA	X		P		P	P	W	P
Smartphone	X		P		X	P	W	P
Network tablet	X						W	
USB flash drive	X							
2D barcode printout	X							
Digital pen	X							

X = Device typically has built-in support for that physical communications network.

A = Additional data communications hardware could enable that device to communicate over that network.

P = Possible in theory for device to communicate over this network with additional hardware. However, this configuration is not known to be tested or widely used.

W = Device can indirectly connect to fixed-line broadband over a local 802.11 (Wi-Fi) network.

Table 5

Devices and software classified by supported data capture mode.

Real-time only	Both real-time/store-and-forward	Store-and-forward only
IVR software	Satellite phone	USB flash drive
Survey software	Cell phone	2D barcode printout
HF Radio	Laptop (Regular)	Digital pen
	Laptop (Resource-limited settings)	
	Laptop (Netbooks)	
	PDA	
	Smartphone	
	Network tablet	

5.4. Step 4: compare and select devices based on operational criteria, appropriateness for setting

All 14 categories of devices were evaluated based on the operational criteria. Results of that evaluation are shown in Table 6. Though the lowest-cost devices can be acquired for tens of dollars or less, the cost for assembling an end-to-end system could be higher. Of the six devices which cost $100 or less, only the cell phone and smartphone can collect and transmit data without requiring any intermediate computer. Interestingly, data from devices that capture data exclusively in either real-time or store-and-forward mode must be ingested through an intermediate computer prior to transmission into the repository.

With regard to deployment concerns, battery-based devices tend to have a minimum charged life of two hours. However, manufacturers' claims for individual implementations suggest that the actual charged life of a battery-powered device will probably be higher. This study's definition for the Durability criterion was found to be stringent. Though each device category must be ruggedized by design to satisfy this criterion, the cell phone, which is not innately ruggedized, is flourishing in varied settings across the world [30, 8].

Considering the time needed to train a user, we estimated ranges for devices that support multiple applications. We estimated that devices which present computerized UIs would require a moderate amount of training time (1–3 days), and devices which require separate computers at the data collection PoE would require the most training time (2–4 days).

As indicated earlier, three categories of laptop computers were identified: standard laptops, resource-limited setting laptops, and netbooks. Though their capabilities are similar, each category has distinct price and feature characteristics. Standard laptops are the closest to desktop systems in terms of features: they have the greatest processing power among the three categories, and offer screen sizes that are comparable to those of desktop monitors. Since they are targeted to mobile consumers who have greater resources and need desktop-like performance, they are not innately designed for extreme environments. On the other end of the spectrum, the resource-limited setting laptop category includes products such as the One Laptop per Child (OLPC) XO Laptop and the Intel Classmate PC. These laptops have less processing power and small screens compared to standard laptops, and are targeted to children in rugged and resource-limited settings [31, 32]. The third class of laptops, called netbooks, has arisen as a response to the innovations of resource-limited setting laptops. Netbooks have similar features to resource-limited setting laptops and are intended primarily for activities that are less processor-intensive, such as browsing the Web. However, they lack innate durability. Given their feature set and lower price than standard laptops, they have exploded in popularity across both developed and emerging markets [33 pp. 71-72].

Table 6

Devices, as characterized by remote-area health data collection criteria.

Device	Cost	Battery life	Durability	Collection computer	Transmission computer	Training time	Comments
IVR software	$0 for open-source software	(n/a)	(n/a)		X	< 1 day	
Survey software	$1000s	(n/a)	(n/a)		X	1-2 days	
HF Radio	$200	6 hours	N	X (email, Web)	X (email, Web)		Radio operation may require local license.
Satellite phone	$1300	3 hours talk time	Y			1 day	
Cell phone	$35	2 hours talk time	N			1 day	Cost for a low-end unlocked GSM phone.
Laptop (Regular)	$380	2.5 hours	N			1-3 days	
Laptop (Resource-limited settings)	$200	2 hours	Y			1-3 days	
Laptop (Netbook)	$250	2 hours	N			1-3 days	
PDA	$300	4 hours	N			1-3 days	
Smartphone	$100	3 hours talk time	N			1-3 days	
Network tablet	$200	4 hours	N			1-3 days	
USB flash drive	< $20 for 4 GB USB drive	(n/a)	Y for mild jolts/falls, N for other stresses	X	X	2-4 days	
2D barcode printout	$400	Varies by printer	N	X	X	2-4 days	Cost includes barcode generating software and ruggedized printer. Cell phone camera and open-source ZXing software could be used to read barcode [34].
Digital pen	$100	2 hours	N		X	1-3 days	

5.5. Step 5: consider different architectures for remote data capture

To illustrate how the preceding results could be applied when designing a particular data capture system, the task of deploying a system for three hypothetical remote clinics in Southeast Asia with varying levels of network access is considered here. These clinics are referred to as A, B, and C, and are named in ascending order of network connectivity. Clinic A has no terrestrial access to existing telecommunications infrastructure, and therefore represents the most challenging setting from a data transmission perspective. Clinic B has access to a second-generation (2G) cellular telephone and data network. Clinic C has access to a mobile broadband network, in addition to a 2G or higher-generation cellular telephone and data network.

The primary purpose of these clinics is to detect and treat the patient population for influenza-like illness (ILI) given the recent identification of a novel H1N1 influenza that is currently sweeping through North America. As such, the timeframe for deploying a data capture link is on the order of days. In the long term, the clinics are intended to stay open after ILI incidences subside, and provide early detection of any local infectious disease outbreaks. Let us further assume that these clinics are staffed by infirmary technicians who have limited medical training [28], and that an outbreak of ILI is detected shortly after a data capture system is put in place.

These example circumstances are not meant to encompass all possible constraints a system designer might face, but rather they represent a continuum of network connectivity scenarios that would likely be encountered in the resource-limited world. The constraints of each individual setting, including cost limits, data timeliness, and desired level of comprehensiveness, will ultimately influence which approach is selected for deployment. We did not consider fixed-line network services among the example settings because mobile network services are growing more rapidly [8 p. 3].

5.6. Clinic A

Clinic A would have communications access only through the manual delivery, satellite, and radio networks. Though these networks have the greatest levels of penetration in the resource-limited world, they also have tradeoffs which may impede wide-scale deployment. Specifically, a tradeoff emerges between the timeliness of reporting and cost. At first, Clinic A may use a resource-limited setting laptop to collect simple counts of patients presenting with respiratory symptoms, and transmit that data using a USB Flash Drive. It may initially be acceptable for the clinic's data to arrive at a central repository several days after entry, via a manual delivery network like postal mail.

However, when an ILI outbreak is first detected, the need for two-way communications between Clinic A and central health authorities will be heightened. Central authorities may want to closely monitor ILI case counts, ILI-related deaths, and perform testing to determine if the illness is due to H1N1. Similarly, Clinic A personnel will likely wish to consult with experienced medical personnel regarding complicated cases. Consequently, the clinic will need to shift quickly to different data capture technologies that accelerate the time between data entry and transmission.

Though satellite communications can bridge the distance between the clinic and authorities, their prolonged use is probably unsustainable, because they have the highest equipment and service costs among all networks. Rather, satellite communications would be best utilized in acute short-term situations. Alternately, a radio link offers minimal ongoing service costs, but its up-front costs may be considerable. Setting up and operating a radio network would require specialized expertise, first with the establishment of a new terrestrial form of telecommunications infrastructure. Additionally, radio training for clinic personnel could remove them from the field for up to two working weeks [13].

If detailed patient data were to be collected on a laptop with either communications network, both networks are not known to work with resource-limited setting laptops, so the clinic may urgently need to obtain a standard laptop or netbook. On the whole, "future-proofing" Clinic A to avoid the need to shift technologies during an outbreak, while also minimizing ongoing service costs, would indicate that the radio network should be deployed from the start.

5.7. Clinic B

Clinic B could access all of Clinic A's networks, as well as the 2G cellular telephone and data network. With its modest service costs and minimal required technical knowledge, adopting an architecture based on the cellular network when Clinic B is first established could provide sustainable, timely, and detailed two-way communications between the clinic and central authorities. Using a $40 cell phone, Clinic B could effectively transmit data in any language with a combination of the voice, touch-tone, and text message applications for tens of dollars per month. When the ILI outbreak occurs, more frequent communications could take place at a nominally-increased cost. Moreover, this architecture readily scales with increasing numbers of clinics.

5.8. Clinic C

In addition to Clinic B's networks, Clinic C would have access to the mobile broadband network. That network would allow the clinic to transmit large chunks of health data, such as images or instrument results, in real-time to the surveillance repository using a laptop or smartphone. Though the mobile broadband network currently has low penetration in the resource-limited world, the trend of decreasing cellular telephone network costs with increasing network penetration suggests that the mobile broadband network will become more affordable as it becomes more prevalent in the resource-limited world [8 p. 71].

6. Discussion

The criteria-based process described in this paper is useful for quickly estimating the costs and capabilities that can be realized when deploying a set of ICTs for remote health data capture. The criteria outlined at each step serve as a checklist of essential technical information for system design. Characterizations of existing ICTs demonstrate how the process could be used, and several scenarios encompassing the range of network connectivity situations that might be encountered in the resource-limited world were considered. These scenarios suggest that existing implementations of networks, devices, and applications could be integrated into feasible and effective remote-area health data capture systems.

There are several directions for future research. Since this study relied primarily on advertised specifications to characterize ICTs for capturing health data from remote areas, many salient deployment details remain to be identified. Experience with individual deployments could quantify essential practical knowledge about introducing ICTs into a setting, such as their actual price, performance, and required training time. Moreover, it is likely that deployment strategies will be shaped by a variety of factors external to technology. When choosing specific ICTs for deployment in a given setting, contextual feasibility is a foremost consideration among these external factors. The skill required to operate the chosen ICTs should match the technological and cultural "literacy" of users collecting data on the ground. Additionally, the local sustainability of a system that is deployed for an extended time period needs to be demonstrated, as high levels of long-term international support are usually not sustainable. Of course, justifying continued support for a system will likely entail measuring the public health impact as a result of surveillance [35 p. 76].

Several other external factors could also influence deployment strategies. Since the need for surveillance can be precipitated by a disaster, ICTs may be chosen based on how quickly and widely they can be deployed as part of the response. One example is the WiMax wireless networking technology, which was deployed in the wake of Hurricane Katrina to support emergency responders and the American Red Cross [36, 37]. In addition, new systems may be able to “piggyback” on existing disease surveillance infrastructure, such as networks of overseas military laboratories [38], which could determine the ICTs selected. Deployments of health data capture systems might also be coordinated with larger economic and social development initiatives [39].

More generally, ICTs can serve as invaluable tools that support many public health functions. For example, they could be used in remote villages to capture basic statistics, such as births and deaths, in addition to reporting data for specific disease surveillance needs. By arming health care workers with tools to collect and transmit data in a more timely fashion, data can be compiled at a central level for the purpose of early event detection, thus leading to more effective public health actions. Future work could also consider the benefits and distinct challenges of applying ICTs for health data capture in urban and semi-urban settings, and even transfer “lessons learned” back to the developed world.

Although this study considered only brand-new commodity ICTs, the full range of ICTs which could support remote health data capture is broader. For example, the United Nations Children’s Fund’s Bee system is a promising prototype device that combines computing and radio features [40]. However, it was excluded from this study because it is not yet a commodity. Government-off-the-shelf [41], proprietary, and research ICTs might also be adapted for data capture. Donations of services and refurbished equipment from the developed world, such as cell phones [42], may mitigate costs. It is certainly worth noting that regardless of how advanced an ICT is, technology alone cannot improve disease surveillance. Rather, successfully deploying a disease surveillance system will require innovatively integrating and adapting technology for each individual setting.

Given that ongoing technology advances yield greater performance and new capabilities for less cost, this study could be repeated on an annual basis with different results. ICT prices tend to decrease over time with the commoditization of sophisticated hardware, as well as the proliferation of open-source software systems and freely-licensed hardware specifications [33 p. 72]. Additionally, voice recognition software may eventually render the voice application capable of reliably formatting spoken data, enabling truly comprehensive data capture [43]. New types of communications networks could also gain significant market penetration. In particular, mesh networks have the potential to extend the range of a wireless network access point to non-line-of-sight devices if they are separated by a “critical mass” of line-of-sight devices [44].

A number of current projects are integrating and deploying ICT-based solutions specifically for disease surveillance in remote areas, such as InSTEDD, EpiSurveyor, AED-SATELLIFE, and efforts at The Johns Hopkins University Applied Physics Laboratory [45- 47, 15]. These projects, and similar new initiatives, could be documented from inception to deployment and evaluation across varied environments. The resulting literature would prove indispensable for capturing health data from many individual settings to build a sustainable global-scale public health surveillance capability.

7. Conclusion

We have described a criteria-based process that can be used to characterize ICTs for the feasible and effective capture of health data from remote areas. Awareness of global public health has grown increasingly important, considering today’s unmatched levels of international travel and corresponding outbreaks of infectious diseases. At the same time, ICTs have been developed that offer sophisticated computing and communications capabilities from remote settings. The convergence of both these aspects of modernization presents a unique opportunity to extend the reach and benefits of disease surveillance to those with minimal resources.

Acknowledgments

We thank Steven Babin, Jacqueline Coberly, Larry Frank, Wayne Loschen, Annie McGovern, Sue Pagan, Joseph Skora, and Rich Wojcik, for invaluable assistance which made this paper possible. This research was supported by the Stuart S. Janney Publication Program at the Johns Hopkins University Applied Physics Laboratory, and by the Armed Forces Health Surveillance Center, Division of GEIS Operations. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official, or as reflecting true views of the Department of the Navy, Department of the Army, the Department of Defense, or the Johns Hopkins University.

Appendix A. Criteria definitions

A.1. Step 1: characterize accessible communications networks

Network assessment criteria are separated into two groups: network features and technical specifications.

A.1.1. Network features

- **Network**: a type of communications network that could transmit health data from remote areas across the world. With the exception of the manual delivery network, each network identified is an electronic communications network that meets two requirements:
  - The network can transmit data across distances of 10 km or more.
  - The network can be accessed remotely with commercially-available equipment and is deployed publicly somewhere in the world.
- **Generation**: refers to individual versions, or generations, of a particular type of physical communications network.
- **Penetration**: level of that network's penetration in the developing world.
- **Native applications**: applications that the network, and devices compatible with the network, are intrinsically designed to support.
  - For a network composed of multiple generations, all of the network's native applications apply for the individual generations as well. The individual network generations can support more native applications, as listed.
- **Extended applications**: applications that a laptop computer, and potentially other devices, could use over that particular network with additional data communications hardware.
  - Use of extended applications could incur significant latency.
- **Comments**: additional notes about the communications network.

A.1.2. Network technical specifications

- **Network:** as defined above.
- **Generation:** as defined above.
- **Data communications hardware**: any separate, modem-like hardware that is required to connect a laptop computer, and potentially other devices, to this network.

- **Data hardware cost**: approximate minimum retail cost of a brand-new unit of data communications hardware that is compatible with this network. This price was for a low-end unit of compatible hardware advertised on the Web.
- **Maximum data speed**: maximum data bandwidth upload speed for an entry-level network service plan that could be expected when connecting to the Internet over this type of network in the developed world. This criterion is based upon advertised claims by data hardware vendors and service providers.
    - The upload speed is important because it affects the amount of time required to transmit data over a connection.
    - Maximum data speeds for network services offered in the developed world are listed to provide an upper bound on that network's possible performance.
- **Recurring service costs**: estimated recurring service cost in the developing world associated with the ongoing usage of this network once the reporting connection is set up.
    - For satellite network services whose costs are fixed worldwide, prices were based approximately on the minimum service prices that could be found on the Web.
    - For network services whose costs varies by country, price ranges were derived from the ITU Price Basket data for the 120 countries with the lowest GNI per capita [8].

A.2. Step 2: assess applications against data reporting and language criteria

- **Application**: a type of data collection and transmission application.
- **Mode of latency**: this criterion refers to the latency of data capture, and falls into one of two categories:
    - Real-time: must synchronously transmit the data over the network from the PoE in order to store the data.
    - Store-and-forward: can store data at the PoE without opening a network connection, and asynchronously upload that data later.
- **World language support**: falls into one of two categories:
    - All: this interface can be readily translated into any language that is written and spoken.
    - Some: the interface may not be readily translated to some languages that are written and spoken.
- **Formatted data**: whether the application could capture structured data that could easily be ingested and automatically analyzed.
- **Patient-level data**: whether the application could capture patient-level data, including identifying information.
- **Explanation**: additional explanation as to how that application can be used for data capture and transmission.

A.3. Step 4: compare devices based on operational criteria

- **Device**: a type of data capture device.
    - The devices considered can collect data from remote areas where the supply of electricity is intermittent.
    - Each device is a commodity, and is actively marketed by its source at the time of the study.
- **Cost**: cost in US dollars to acquire a single brand-new unit of this device.
    - This value was based approximately on the minimum cost for this device that could be found on the Web for sale in the United States.

    - **Battery life**: minimum battery life which could ordinarily be expected for continuous usage of a brand-new unit of this device. This value was based on manufacturers' and vendors' technical specifications.
    - **Durability**: whether this device is ordinarily designed to withstand environmental and usage stresses, such as:
        - Humidity.
        - Extreme temperatures.
        - Dust, dirt, and moisture.
        - Jolts/falls.
- **Collection computer**: whether a computer is required to actually collect data from the user at the PoE and write it onto the device.
- **Transmission computer**: whether a computer, and potentially other hardware, is required to persist data from the device into the surveillance repository.
- **Training time**: our estimate of how much time might be required to provide hands-on training to the health worker on how to use the device for data collection, based on our assessment of the "learning curve" associated with the device.
    - This estimate is listed to provide a baseline for comparison among different devices, and does not factor in the additional time required for training on collecting data for a particular report.
    - The estimate assumes that the health worker is literate in the language that data is being collected, but is not familiar with computers.
    - Since the touch-tone telephone has the most familiar UI of any ICT [8 p. 1], we reasoned that data capture systems based on this device would require the least amount of training time. As a baseline, we estimated this training time to be less than 1 day.

Appendix B. Explanation of non-traditional data capture devices

A.1. Interactive voice response (IVR) software

Voice menu-driven systems, like those used by many businesses ("Press 1 for sales, press 2 for customer support, etc.") are known commercially as interactive voice response (IVR) systems. Since IVR systems work with touch-tone phones, they collect data and respond to the caller's voice menu choices from the digits that the caller can dial on a touch-tone phone. Additionally, IVR systems may allow the caller to send voicemails. One example of an IVR system for health data capture is the Cell-PREVEN system deployed in Peru. It allows health workers to enter reports about population health via touch-tone, and leave voicemails with additional comments for central health authorities [11].

A.2. Survey software

Survey software can acquire a health worker's report from any one of several communications applications [48]. These applications include:

- Touch-tone data entry/voice (via a built-in IVR subsystem).
- Text messages.
- Email.
- Web.
- Radio.

A.3. USB flash drive

Capturing health data on USB flash drives might be a viable option for health workers with USB-compliant data collection devices, such as laptops or PDAs, who cannot connect to electronic communications networks from rural sites. Large amounts of data on these drives can be easily hand-carried by health workers for uploading later, perhaps from an Internet café or a central health office. Unfortunately, their compact size could put them at significant risk of theft or loss, and they need to be operated with care.

A.4. 2D barcode printout

Printing out and sending a physical copy of data via messenger or postal mail may be an option for transmitting health data from very isolated locations to central data collection centers. Ideally, the printouts would arrive at the collection centers in an encoded format that could be automatically scanned in, instead of entered by dedicated data entry personnel. One solution might be to print out health data encoded in two-dimensional (2D) barcode format. 2D barcodes can encode large amounts of information vertically and horizontally, such as the "bulls-eye" barcode that parcel delivery services place on address labels. This solution would be comprised of several components: 2D generating software and a ruggedized printer on the health worker's end, and a 2D barcode scanner to read in the barcodes on the central collection center's end.

A.5. Digital pens

Digital pens are portable data input and storage devices that can be used like ordinary pens. When a user writes with a digital pen, the handwriting is stored digitally on the device, and can later be uploaded to a computer. In some cases, the pen data can be automatically formatted when it is uploaded. Since digital pens have a familiar form factor, they may lend themselves for use in remote environments where health workers do not have a computer or phone handy, but can hand-carry the pen to a networked computer for later transmission to the repository.

A.6. Network tablet

Network tablets are handheld computers designed primarily to browse the Internet over a Wi-Fi network connection. They are capable of storing a limited amount of data, and may be able to download and run third-party applications. Examples of this device category include the Nokia N810 Internet tablet, Apple's iPod touch, and Archos Internet Media Tablet. Since these devices rely on Wi-Fi network connections, which can be rare in remote areas, they may be best utilized for collecting data which can be uploaded later.

References

[1] International Society for Infectious Diseases. ProMED Digest. June 19, 2009, Vol. 2009, 298.

[2] Lewis S, Chretien J-P. The potential utility of electronic disease surveillance systems in resource-poor settings. Johns Hopkins APL Technical Digest. 2008;27(4):366-373.

[3] Khan K, Arino J, Hu W, Raposo P, Sears J et al. Spread of a novel influenza A (H1N1) virus via global airline transportation. N Engl J Med. 2009;361(2):212-4.

[4] World Health Organization. International Health Regulations (2005), Second Edition. Geneva, Switzerland : WHO Press, 2008. Available at http://whqlibdoc.who.int/publications/2008/9789241580410_eng.pdf.

[5] Jones K, Patel N, Levy M, Storeygard A, Balk D, Gittleman JL, Duszak P. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990-3.

[6] Chretien J-P, Burkom HS, Sedyaningsih ER, Larasati RP, Lescano AG, Mundaca CC, Blazes DL, Munayco CV, Coberly JS, Ashar RJ, Lewis SH. Syndromic surveillance: adapting innovations to developing settings. PLoS Med. 2008;5(3):e72.

[7] United Nations Population Division. World Urbanization Prospects: The 2007 Revision Population Database. [Online] 2007. [Accessed: July 22, 2009.] http://esa.un.org/unup/.

[8] International Telecommunication Union. Measuring the Information Society: The ICT Development Index. Geneva, Switzerland : ITU, 2009. Available at http://www.itu.int/ITU-D/ict/publications/idi/2009/index.html.

[9] One Laptop per Child. One Laptop per Child (OLPC): Laptop Hardware > Features. [Online] [Accessed: June 30, 2009.] http://laptop.org/en/laptop/hardware/features.shtml.

[10] Iridium Satellite LLC. Iridium. [Online] 2007. [Accessed: April 29, 2009.] http://iridium.com/about/about.php.

[11] Curioso W, Karras B, Campos P, Buendia C, Holmes K, Kimball AM. Design and implementation of Cell-PREVEN: a real-time surveillance system for adverse events using cell phones in Peru. AMIA Annu Symp Proc. 2005;176-180.

[12] Odero W, Rotich J, Yiannoutsos C, Ouna T, Tierney W. Innovative approaches to application of information technology in disease surveillance and prevention in Western Kenya. J Biomed Inform. 2007;40(4):390-7.

[13] Martinez A, Villarroel V, Seoane J, Del Pozo F. A study of a rural telemedicine system in the Amazon region of Peru. J Telemed Telecare. 2004;10(4):219-25.

[14] Morris T, Pajak J, Havlik F, Kenyon J, Calcagni D. Battlefield Medical Information System-Tactical (BMIST): The application of mobile computing technologies to support health surveillance in the Department of Defense. Telemedicine and e-Health. 2006;12(4):409-16.

[15] Lewis S, Coberly J, Wojcik R, Loschen W, Frank L, Ashar R et al. Public health informatics tools for electronic disease surveillance in resource-limited settings. International Meeting on Emerging Diseases and Surveillance, Austria, Vienna. 2009.

[16] Ziff Davis Publishing Holdings Inc. Cellular generations Definition from PC Magazine Encyclopedia. [Online] [Accessed: May 7, 2009.] http://www.pcmag.com/encyclopedia_term/0%2C2542%2Ct%3Dcellular+generations&i%3D55406%2C00.asp.

[17] EE Times Asia. 4G hype won't phase out 2G/3G networks. [Online] April 22, 2009. [Accessed: July 23, 2009.] http://www.eetasia.com/ART_8800570428_499488_NT_f44aaa08.HTM.

[18] Inmarsat Global Limited. BGAN: Global voice and broadband data. [Online] February 2009. [Accessed: April 29, 2009.] http://www.inmarsat.com/Downloads/English/BGAN/Collateral/bgan_overview_brochure_EN.pdf?language=EN&textonly=False.

[19] IEEE Standards Association. IEEE 802.16e Mobile WirelessMAN Standard is Official. [Online] December 7, 2005. [Accessed: May 8, 2009.] http://standards.ieee.org/announcements/pr_p80216.html.

[20] Patra R, Nedevschi S, Surana S, Sheth A, Subramanian L et al. WiLDNet: Design and Implementation of High Performance WiFi Based Long Distance Networks. Cambridge, MA : 2007. NSDI '07: 4th USENIX Symposium on Networked Systems Design & Implementation. pp. 87-100.

[21] Miller C. Research@Intel - Intel (r) Rural Connectivity Platform becomes a reality. [Online] March 10, 2008. [Accessed: May 10, 2009.] http://blogs.intel.com/research/2008/03/rural_connectivity_platform_be.php.

[22] Miller C. Research@Intel - Intel® RCP goes commercial. [Online] November 24, 2008. [Accessed: May 10, 2009.] http://blogs.intel.com/research/2008/11/rcp_goes_commercial.php.

[23] Kantronics, Inc. Kantronics Radio Modems/TNC's. [Online] 2007. [Accessed: May 5, 2009.] http://www.kantronics.com/modems.html.

[24] Marshall W. Radio E-mail in West Africa: The Complete Version. Linux Journal. [Online] November 2002. [Accessed: April 24, 2009.] http://www.linuxjournal.com/article/6299.

[25] Q-MAC Electronics Pty Ltd. www.maritime-index.com - HF Radio bridges the gap between Peacekeepers and NGOs - KNM Media LLP. Maritime-index.com. [Online] KNM Media LLP. [Accessed: April 24, 2009.] http://www.maritime-index.com/index.php?pg=128.

[26] International Institute for Communication and Development. ICTs in developing countries: Booklet IV - Examples of applications -- International Institute for Communication and Development (IICD). [Online] December 11, 2002. [Accessed: April 24, 2009.] http://www.iicd.org/articles/IICDnews.import13.

[27] GSM Association. 3 Billion GSM Connections On The Mobile Planet - Reports The GSMA ~ GSMA World. [Online] April 16, 2008. [Accessed: May 9, 2009.] http://gsmworld.com/newsroom/press-releases/2008/1108.htm.

[28] Martinez A, Villarroel V, Seoane J, Del Pozo F. Rural telemedicine for primary healthcare in developing countries. IEEE Technology and Society Magazine. 2004;23(2).

[29] Voxiva. Voxiva: The Power of the Internet, The Reach of the Phone. [Online] [Accessed: June 30, 2009.] http://voxiva.com/solutionspage.php?catname=HealthWatch.

[30] Corbett S. Can the Cellphone Help End Global Poverty? New York Times. April 13, 2008. Obtained from http://www.nytimes.com/2008/04/13/magazine/13anthropology-t.html?pagewanted=all on June 30, 2009.

[31] One Laptop per Child. One Laptop per Child (OLPC): Laptop Hardware. [Online] [Accessed: June 30, 2009.] http://laptop.org/en/laptop/hardware/index.shtml.

[32] Intel Corporation. Second-Generation Intel-Powered Classmate PCs: Purpose-Built Netbooks for Education. [Online] 2008. [Accessed: June 30, 2009.] http://www.classmatepc.com/pdf/ClassmatePC_ProductBrief_SecondGeneration_Celeron_Eng_LoRes.pdf.

[33] Kraemer K, Dedrick J, Sharma P. One Laptop Per Child: Vision vs. Reality. Communications of the ACM. June 2009;52(6):66-73.

[34] Google, Inc. zxing - Project Hosting on Google Code. [Online] [Accessed: July 27, 2009.] http://code.google.com/p/zxing/.

[35] Dias M, Brewer E. How computer science serves the developing world. Communications of the ACM. 2009;52(6):74-80.

[36] Sinha R. Mobile Magazine » FCC Pushes WIMax OK for Katrina Victims, Intel supplies the hardware. [Online] September 9, 2005. [Accessed: November 9, 2009.] http://www.mobilemag.com/2005/09/09/fcc-pushes-wimax-ok-for-katrina-victims-intel-supplies-the-hardware/.

[37] Intel Corporation. 2005 Information Technology Annual Performance Report. [Online] 2006. [Accessed: November 9, 2009.] http://www.intel.com/it/pdf/2005-apr.pdf.

[38] Chretien J-P, Gaydos J, Malone J, and Blazes D. Global network could avert pandemics. Nature. 2006;440(7080):25-26.

[39] Parikh T. Designing an Architecture for Delivering Mobile Information Services. [Online] 2007. [Accessed: November 9, 2009.] http://people.ischool.berkeley.edu/~parikh/papers/parikh-thesis.pdf.

[3640] United Nations Childrens Fund (UNICEF). UNICEF - UNICEF in emergencies - 'Bee' system to bring improved connectivity. [Online] August 20, 2008. [Accessed: June 10, 2009.] http://www.unicef.org/emerg/index_45259.html.

[41] Defense Acquisition University. AKSS - Ask A Professor - Government off the Shelf. [Online] February 13, 2009. [Accessed: July 29, 2009.] https://akss.dau.mil/askaprof-akss/qdetail2.aspx?cgiSubjectAreaID=14&cgiQuestionID=27290.

[42] EARTHWORKS. Recycle My Cell Phone. [Online] 2007. [Accessed: July 29, 2009.] http://recyclemycellphone.org/faq.shtml.

[43] Martin J. From Speech to Text - PC World. [Online] October 10, 2007. [Accessed: June 30, 2009.] http://www.pcworld.com/article/138262/from_speech_to_text.html.

[44] WiMAX Industry. WiMAX Non-Line-of-Sight Solutions. [Online] 2008. [Accessed: June 30, 2009.] http://www.wimax-industry.com/sp/tsc/1j.htm.

[45] InSTEDD. InSTEDD. [Online] [Accessed: June 30, 2009.] http://instedd.org/.

[46] DataDyne.org. EpiSurveyor | DataDyne.org. [Online] [Accessed: June 30, 2009.] http://www.datadyne.org/episurveyor.

[47] AED SATELLIFE. AED Satellife - Center for Health Information and Technology. [Online] [Accessed: June 30, 2009.] http://www.healthnet.org/.

[48] Prahalad CK. The Fortune at the Bottom of the Pyramid. Wharton School Publishing, 2006. Excerpt obtained from http://voxiva.com/articlespage.php?id=4 on June 30, 2009.