Published in final edited form as:

J Acquir Immune Defic Syndr. 2014 January 1; 65(0 1): S44–S49. doi:10.1097/QAI.0000000000000039.

Supporting research sites in resource-limited settings: Challenges in implementing IT infrastructure

Christopher Whalen, M.A.[†], Deborah Donnell, PhD^{*}, and Michael Tartakovsky, M.S.[†]

[†]Office of CyberInfrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD

^{*}Fred Hutchinson Cancer Research Center, Seattle, WA

Abstract

As Information and Communication Technology infrastructure becomes more reliable, new methods of Electronic Data Capture (EDC), datamarts/Data warehouses, and mobile computing provide platforms for rapid coordination of international research projects and multisite studies. However, despite the increasing availability of internet connectivity and communication systems in remote regions of the world, there are still significant obstacles. Sites with poor infrastructure face serious challenges participating in modern clinical and basic research, particularly that relying on EDC and internet communication technologies. This report discusses our experiences in supporting research in resource-limited settings (RLS). We describe examples of the practical and ethical/regulatory challenges raised by use of these newer technologies for data collection in multisite clinical studies.

Keywords

Information technology (IT); research infrastructure; data management; electronic data capture (EDC); mHealth; mobile technology

Introduction

Conducting, supporting, and sponsoring basic and clinical research in settings with poor infrastructure is a serious challenge for those working to advance use of technologies for data collection or capture and communication. Improving support for these technologies is critical for research clinics and laboratories in high income as well as resource-limited settings (RLS).

Sites with reliable infrastructure can increasingly rely on sustainable availability of internet connectivity for communication and data exchange to support research activities. However, many research sites in RLS still struggle with unreliable internet access. Building systems

Corresponding Author: Christopher Whalen Office of CyberInfrastructure and Computational Biology National Institutes of Health
National Institute of Allergy and Infectious Diseases 6610 Rockledge Dr., Suite 2102 Bethesda, MD 20814 Phone: 301-496-1603
FAX: 877 801 7159 cwhalen@mail.nih.gov.

Conflicts of Interest: None of the authors has a conflict of interest.

that can overcome this is challenging for the information technology professionals who support the investigators and their protocols. A contributing factor is a lack of skilled information technology (IT) support personnel on-site to assist researchers in overcoming the obstacles in reliable internet connectivity, to support advanced IT technologies and maintain the basic infrastructure needed to implement them [1]. We discuss the challenges in providing and maintaining reliable IT infrastructure at these sites, and present both the opportunities and some ethical considerations in implementing these new technologies for data management and communication.

Challenges for IT infrastructure

Reliable electrical power

Modern computing and scientific systems require continuous conditioned power to maintain the equipment and avoid loss of data [2, 3]. In RLS, one of the most difficult challenges in supporting IT infrastructure for field clinics and laboratories, and in cities has been the lack of reliable and high quality utility power.

There are strategies for delivering conditioned power regardless of location; examples include proper sizing of the power system, Automatic Voltage Switches, and Uninterruptible Power Supplies (UPS). Correct sizing of power systems and circuits for either a single device (such as a flow cytometer or server) or whole facilities is critical. A spike in power demand when a device starts, called inrush or startup load, is common for many devices such as refrigeration systems and laser printers. Inrush demand requires sizing a power delivery system such as UPS or custom-built inverter/rectifiers/batteries to accommodate the demand; failing to do so reduces the operational life of equipment or failure of systems using, conditioning and sharing that circuit. Use of tools like Automatic Voltage Switches (AVS) that cut power to the supplied systems (including UPS and inverters) when there is a fluctuation from the utility or generator can significantly improve operational life. Of the three modern types of UPS (“on-line”, “off-line” and “line interactive”), the “on-line” types provide the best continuous conditioning of power in areas with unstable power [4].

Options for power generation at remote sites or even at core facilities now encompass solar, wind, and traditional generators (sizing and need for battery capacity during dark or windless periods are critical during initial evaluation). Required capacity, availability of local maintenance, and the expected run time are the basic evaluation criteria for generators. Most commonly, sites install standby generators, designed to operate for a few hours at a time to bridge infrequent and short interruptions. Installing a standby generator at a location that suffers from long and/or repeated interruptions in power will result in frequent breakdowns and high maintenance. Continuous or primary generators, designed to deliver power for either constant or variable loads over long periods, are a more viable option [5].

Adherence to electrical codes is a critical aspect of supporting modern information and communication technologies (as well as laboratory equipment) in remote locations. Failure to ground electrical supply grids, network wiring, and antennas leaves equipment unprotected from power surges, yet ensuring correct grounding in locations with poor compliance to electrical codes can be difficult. Proper lightning protection is critical in

locations where there is uncertainty about the utility or facility grounding, since lightning surges can both damage sensitive electronic components and interrupt internet connectivity by causing interference on network cabling [6].

Buy Local

It is generally preferable to use locally supplied equipment because of ready access to local maintenance in the event of failure. However, this can mean committing to using equipment and devices that are only available in specific regions with uncertainty about quality and compatibility. For example, selecting touch screen devices for direct participant data entry in a trial in Brazil, Thailand and South Africa involved using three different devices marketed and sold in local markets, and entailed thorough testing of the interaction of each device with server/client based data entry software, and resolving device driver and software incompatibility. The alternate approach, selecting standardizing devices and equipment, may have avoided some software complexities, but inherently would have required long distance maintenance contracts in three continents, and incurred the hurdle of importing equipment and ensuring compatibility with local electrical, internet and cellular systems.

Internet connections and service providers

Rapid changes in the internet access market throughout the world, including RLS in recent years, have dramatically expanded the available choices of internet service providers (ISP). Most metropolitan areas have broadband internet access, although uncertain ISP technical competence and oversubscription can still cause uneven performance. Even in remote locations, there are now options for data and internet from mobile phone providers that provide reliable, albeit expensive, internet access. In addition, there are new satellite technologies available that face competition from terrestrial mobile networks and optical fiber providers, forcing improved performance and pricing.

Management of the local area networks is still the primary determinant of performance of an internet connection. Fortunately, new products on the market are providing much simpler means of managing internet connections. Cloud managed routers and content filtering provide for easy network monitoring and the ability to prioritize the traffic for applications and websites used directly for research [7]. The most common causes of network congestions is the use of multimedia and social networking and virus-infected computers that participate in spamming or 'bot network attacks on other internet sites. The two most effective ways to preserve the integrity of a network are restricting administrative access (so the users cannot install software or make changes to the system) and disabling autoplay (so that USB flash drives and CDs do not automatically run programs upon insertion). For extensive guidance on securing computers the US National Institute of Standards and Technology maintains extensive checklists as part of the United States Government Configuration Baseline (USGCB) [8].

Computing environment

Implementing and maintaining a server infrastructure in remote locations is both expensive and complicated. Unmanaged servers endanger the data, bandwidth, and the other devices on the network. Malware protection is essential for servers, desktops, laptops, and mobile

devices: systems with inadequate antivirus protection and lack of connectivity to the internet to download virus signatures for detection or system patches and updates present risks to the research data and to collaborating research partners [9]. Furthermore, systems running pirated operating systems result in unpatched systems that become virus reservoirs.

All clinical data guidelines require the backup of data for emergency recovery. Annex B to ISO 15189 Particular requirements for quality and competence specifies the quality management system requirements particular to medical laboratories states, “Efficient back-up should be in place to prevent loss of patient result data in case of hardware or software failure [10].” Unfortunately, tape is still the best solution for backups. While cloud systems are appealing because of seemingly ubiquitous access, restoring from cloud systems can be lengthy and difficult. Cloud based systems can contend seriously for bandwidth in parts of the world where that is still a constraint. Disk to disk systems have problems with portability, it is difficult to send disk backup systems off site to retain for disaster recovery. Thus, simple tape solutions remain the best option for data backup. Typically, the simplest default backup configurations are the best, and avoid vendor-specific hardware compression.

Future Directions

Mobile Technologies and Electronic Data Capture (EDC)

Mobile tools and EDC technology could be a method for conducting clinical field research that does not rely heavily on local IT infrastructure and its attendant challenges. An attraction of direct EDC is eliminating the need for dual data entry, ultimately leading to reduced cost and time conducting clinical research in any environment, not only in RLS. International regulatory bodies are still developing guidelines for directly collecting data electronically for clinical protocols. Monitoring and oversight of studies that use these technologies is also still developing. While there are solutions available now that offer the ability to collect research data using tablets, laptops, and smart phones, not all of them are compliant with the appropriate regulatory framework [11].

Electronic Data Capture (EDC)

True EDC systems may virtually eliminate the need for paper source documents and provide real-time access to the data. In these protocols, capturing data directly to a remote database, offers many advantages for field operations, researchers, and clinical staff in the community. However, a serious concern for local investigators and institutions may be the physical location of the “source” data. Traditional data capture, with paper CRFs or locally installed servers, provides an in-country/at site “master” copy of the source data, even when electronic copies of the data in multi-site studies are stored centrally for primary analysis [12]. These are serious concerns with broad scope that require the attention of the international community. Two proposed methods for solving some of these problems are presented below. For protocols that capture the data to a remote server reliable access to the internet becomes critical and good local IT infrastructure even more important.

Recent advances in mobile technology and application development are facilitating data collection in remote locations with the ability to switch between “on-line” and “off-line” modes and reliable management of seamless data synchronization using a secure and

compliant remote server (public or private “cloud”). Compliant data management is a function of the application selected to collect data on the mobile devices and applications for mobile EDC need to reliably synchronize data collected offline, provide full audit trails to maintain regulatory compliance, and manage form version control.

Mobile devices

Device requirements for mobile data collection in RLS are quite daunting. Devices need to be rugged, have little attraction for theft or repurposing, be convenient for mobile use, have long battery life and/or convenient recharging options. Laptops, tablets and smartphones offer a range of screen sizes, data entry modalities and portability to match study use requirements. If used outdoors, high contrast screens are important. To minimize loss of data, capability for synchronization via cellular data networks is ideal. In studies collecting biological specimens, built-in barcodes are invaluable to enable direct linking of specimens to participant records at the time of data collection. Final selection of the devices seems to offer two approaches, a) select an expensive, durable device that is hardened for use in challenging environments and will suffer fewer hardware failures or breakage; or b) select a low cost device that can be inexpensively replaced if the device is lost or damaged. Ensuring compatibility between device operating systems and software has proved challenging; in our experience, slight mismatches between operating system versions, EDC mobile software, and local service providers can render the devices mysteriously inoperable, a problem that is complicated by the addition of external devices, such as barcode readers, Global Positioning System (GPS), or fingerprint readers.

Furthermore, integration with a mobile device management solution (MDM) is a critical part of any mobile EDC system or device. An MDM provides the data team with the means to control access to the device, maintain a user database for authentication to support audit trails, controlling use, updating applications, patching, and the ability to disable the device remotely if stolen or lost.

Protecting the devices from virus and malware requires planning a management system that will provide user authentication for the devices, as well as blocking the data collectors or other users from installing unapproved applications that can have spyware, malware, or adware included or consume costly bandwidth and storage [14]. Furthermore, mobile devices can become malware vectors (transporting viruses to tablets, phones, laptops, and computers), leading to compromised patient confidentiality or disabling the devices and delaying data collection.

An important tool available on mobile devices is the GPS that can insert the location of the data collection directly into the electronic form. Mobile devices acquire their location either using the satellite GPS (uses more battery power), their location relative to the local mobile towers (least accurate), or a hybrid of the two [15]. A note of caution, storing location of a participant's residence or place of work within the electronic database requires implementation of appropriate protections to avoid compromising confidentiality.

Managing Patient Identifiable Data (PID) and source data in mobile data capture

Most investigators, data managers, and monitors are familiar with using the clinic record, patient identifier, locator information, and Case Report Form (CRF) as source documentation. However, in the mobile EDC environment, the first place where data are recorded is the electronic record in the Clinical Data Management System (CDMS), which thus becomes the source document [16, 17]. This has two immediate challenges: 1) source documents often have patient identifiable data (PID) and 2) there are specific regulatory guidelines for handling of source documents/data [18]. We describe two different approaches to this challenge. One solution, used in HPTN 071, is a single CDMS hosted at the software vendor with internal access controls that separate the source PID (with the identifiable information link to the patient key) from the research or CRF data. This internal barrier provides two distinct environments for the study data: one, with PID, accessible by those overseeing the fieldwork at the site; the other, with no access to PID, accessible by the statistical and data management center. National Institute of Allergy and Infectious Diseases' (NIAID) Laboratory of Parasitic Diseases used an alternate approach in a Tuberculosis-Filariasis co-infection study in southern India, where the support team built a separate database from a commercial product for the PID information collected by the mobile application. This database at the Indian collaborator site provides a function similar to a hospital Electronic Health Record system for the study. The anonymized research data (the CRF data) automatically uploads to the CDMS at a US data center. In addition to the regulatory challenges, the latter solution also overcomes objections often encountered in international clinical protocols that the data reside in the country of the participants. The Division of AIDS, as a study sponsor, requires the following for storage of source documents in electronic format: "All versions of application software, operating systems, and software development tools involved in processing of data or records need to be available as long as data or records associated with these versions are required to be retained [19]." Many countries have guidelines for the handling of specimens and biological agents in clinical protocols but there is less clarity regarding source clinical data. There are strict guidelines for using CRFs as source documents and their accessibility or storage at the site. How these guidelines will translate to mobile EDC are beyond the scope of this paper. In some regions, the electronic source data collected in a clinical protocol may be the only existing health record of the participants and the use and storage of this "document" may be important to both local health practitioners, and IRBs or ethics committees.

An example: Data collection in a large community-randomized HIV prevention trial

An HIV prevention trial planned for 21 communities in Zambia and South Africa (HPTN 071, the PopART study) proposed annual data collection in 52,500 households over a 6-month period. Mobile data collection was seen as the only viable solution, given the scope and environment of the study. The process for the implementation of mobile data collection for a group who were novices in mobile data collection, although experienced in data management for clinical trials, is described in Table 1. Some features of note in this project: 1) repeated visits to the same household member over a three year period led to a decision to implement fingerprint data capture as an aid for participant identification; 2) critical to the protocol was the collection and storage of biological specimens: a device barcode reader was a device system requirement; 3) accurate identification of randomly sampled households

required GPS for implementation; 4) real-time random selection amongst eligible household members was implemented locally on the device.

Improving the ability to analyze previously collected data

Over the past 20 years, sites have collected data for protocols and cohorts at remote field locations and entered the data using multiple database technologies with a separate database files for each round or protocol revision. This is how systems such as FoxPro, dBase, DB2, and FileMaker stored their data. A recent project integrated 14 rounds of cohort census data from multiple FoxPro files into a single SQL database. With each annual census, the survey questions and labels were changed, added, deleted, or reused for slightly different questions. The multiple files and inconsistent labeling made longitudinal analysis of the historic data very difficult because it required custom queries to run against each database and specialized knowledge of the evolution of the questionnaires and field labels. In modern EDC systems, the stronger data management protocols require a single database that dramatically improves current and future accessibility of data. The single SQL database that was formed as noted above met these requirements and has simplified the data analysis process [13].

CONCLUSION

Conducting clinical research in RLS is challenging. Digital devices and laboratory equipment rely on continuity of quality power supplied by well-designed and maintained sources, whether utility, local diesel generator, or renewable energy. The performance of these devices, such as tablets, mobile phones, and laptops, depend on the networks to which they connect, which in turn depend on IT systems that are updated and secure.

Emerging EDC and mobile technologies can make conducting clinical research easier, and improve the quality and timeliness of data collection. These technologies promise faster access to validated data, although this depends critically on reliable access to the internet and support for local IT infrastructure. Implementation of these new technologies creates new challenges for protection of participant confidentiality and the regulatory frameworks. Finally, despite their promise, these technologies are not a panacea for regions with poor IT infrastructure and technical support: without a reliable communication infrastructure to store and access data, researchers may not be able to utilize EDC in research studies.

Acknowledgments

The authors would like to thank Dr. J.J. McGowan of the NIAID Office of Science Management and Operations (OSMO), and Dr. Katherine Zoon of the Division of Intramural Research respectively for their backing of the international IT infrastructure support program at the NIAID.

Source of Funding:

Portions of this research was supported by the HIV Prevention Trials Network, sponsored by the National Institute of Allergy and Infectious Diseases (NIAID), the National Institute of Mental Health (NIMH), and the National Institute of Drug Abuse (NIDA), Office of AIDS Research, of the National Institutes of Health (NIH), Department of Health and Human Services (DHHS), grants UM1-AI068617 (Deborah Donnell). This research was also supported in part by the Office of Science Management and Operations (OSMO) of the National Institute of Allergy and Infectious Diseases (NIAID), Bethesda, MD.

Disclaimer:

The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the National Institutes of Health. Use of trade names is for identification purposes only and does not constitute endorsement by the National Institutes of Health and Prevention or the Department of Health and Human Services.

REFERENCES

1. Best Paying Jobs in Kenya. Salary Explorer.
2. Seymour J. The Seven Types of Power Problems. Schneider Electric. 2011:1–21.
3. The Institute of Electrical and Electronics Engineers I. IEEE Recommended Practice for Powering and Grounding Electronic Equipment. The Institute of Electrical and Electronics Engineers, Inc; New York, NY: 2005. p. 1-603.
4. Emadi SBBaA. Uninterruptible Power Supplies: Classification, Operation, Dynamics, and Control. Applied Power Electronics Conference and Exposition, 2002. APEC 2002. Seventeenth Annual IEEE. 2002; 1:7.
5. Siu SK, Lopopolo J. Compatibility, Sizing, and Design Considerations for Generators and UPSs in Tiers I, II, III, and IV Topologies. Ieee Transactions on Industry Applications. 2011; 47:2324–2329.
6. Michelson, M. Electrical Contractor. 3 Bethesda Metro Center; National Electrical Contractors Association; Suite 1100, Bethesda, MD: 2008. Grounding & Bonding in the Data Center New TIA Requirements.; p. 20814-5372. See more at: <http://www.ecmag.com/about-us%23sthash.MrAHNRzq.dpuf>
7. Rubens, P. Enterprise Networking Planet. IT Business Edge; Foster City, CA: 2013. Cloud Wi-Fi Could Become Key for Network Management..
8. The United States Government Configuration Baseline (USGCB). National Institute of Standards and Technology;
9. Padmavathi GD,S. A Survey on Various Security Threats and Classification of Malware Attacks, Vulnerabilities and Detection Techniques. The International Journal of Computer Science & Applications. 2013; 2:7.
10. Standardisations IOF. ISO 15189 Medical laboratories -- Requirements for quality and competence. International Organization for Standardization; Geneva: 2012.
11. Administration FaD. Draft Guidance: Electronic Source Data in Clinical Investigations. Food and Drug Administration; 2012. p. 10
12. CDISC Standards & Electronic Source Data within Clinical Trials. Edited by (eSDI) eSDIP: Clinical Data Interchange Standards Consortium (CDISC). 2006
13. Beaulah SA, Correll MA, Munro RE, Sheldon JG. Addressing informatics challenges in Translational Research with workflow technology. Drug Discov Today. 2008; 13:771–777. [PubMed: 18617017]
14. Felt, AP.; Finifter, M.; Chin, E.; Hanna, S.; Wagner, D. Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices. ACM; Chicago, Illinois, USA: 2011. A survey of mobile malware in the wild.; p. 3-14.
15. Singhal M, Shukla A. Implementation of Location based Services in Android using GPS and Web Services. International Journal of Computer Science Issues. 2012; 9:237–242.
16. IWG) GIWGG. Edited by Agency EM. European Medicines Agency; London, UK: 2007. Reflection paper on expectations for electronic source data and data transcribed to electronic data collection tools in clinical trials..
17. Ohmann C, Kuchinke W, Canham S, Lauritsen J, Salas N, Schade-Brittinger C, et al. Standard requirements for GCP-compliant data management in multinational clinical trials. Trials. 2011; 12:85. [PubMed: 21426576]
18. use ICoHotropfh. NOTE FOR GUIDANCE ON GOOD CLINICAL PRACTICE. European Medicines Agency; 2002. Guideline for Good Clinical Practice.; p. 1-48.
19. DAIDS. Requirements for Source Documentation in DAIDS Funded and/or Sponsored Clinical Trials. DHHS/NIH/NIAID/DAIDS/Office for Policy in Clinical Research Operations (OPCRO); Bethesda, MD: 2007. SOURCE DOCUMENTATION REQUIREMENTS.

Data Flow

Table 1

Process steps for implementing mobile data collection for HPTN 071, using a contracted software vendor

Steps	Process
Software vendor selection	<ul style="list-style-type: none"> ■ Collection of requirements and specifications for software (e.g. CFR Part 11 compliance; off-line/online synchronization) ■ Review of vendors ■ Execution of pilot using 2 vendors ■ Review of pilot findings, selection of final vendor
Device selection	<ul style="list-style-type: none"> ■ Define requirements: barcode scanning, offline GPS, SIM card/WiFi/Bluetooth, USB ports, ruggedness, battery life ■ Coordinate joint expertise of software and hardware vendors to ensure platform-device compatibility ■ Select device type: phone, tablet, laptop, mini-laptop (ranged from \$500 - \$5500 per device) ■ Select Operating System: Android/Windows ■ On-site testing for compatibility and acceptability ■ Timeline for availability: varied from 2 weeks to 3-4 months
Implementation of data collection	<ul style="list-style-type: none"> ■ Development of data collection plan and procedures ■ Preparation of document specifying finalized data collection instrument to hand-off to vendor ■ Specification of required device support for field process (e.g. verification of house location, scanning of barcodes on specimens) ■ Implementation of electronic CRF forms by vendor ■ Testing and validation on devices by vendor ■ Translation of questions into 6 languages ■ Specification of user roles with tablet and central database access ■ Specification of web-based reports by user role ■ Site set-up of secure locations for storing, charging, and sync'ing devices
Field and UAT testing	<ul style="list-style-type: none"> ■ 3 rounds of User Acceptability Testing (UAT) for Version 1 + 1 week of post-training field practice ■ 2 rounds of UAT testing for Version 2 ■ Acquisition and import of field site household census data
Site Training	<ul style="list-style-type: none"> ■ 1 training in South Africa ■ 3 training sessions in Zambia

Table 2

Executive Summary	
• Modern computing and scientific systems require continuous conditioned power to maintain the equipment and avoid loss of data. In RLS, one of the most difficult challenges in supporting IT infrastructure for field clinics and laboratories, and in cities has been the lack of reliable and high quality utility power.	
• An adherence to electrical codes is a critical aspect of supporting modern information and communication technologies (as well as laboratory equipment) in remote locations.	
• It is generally preferable to use locally supplied equipment because of ready access to local maintenance in the event of failure. However, this can mean committing to using equipment and devices that are only available in specific regions with uncertainty about quality and compatibility.	
• Rapid changes in the internet access market throughout the world, including RLS in recent years, have dramatically expanded the available choices of internet service providers.	
• Management of the local area networks is still the primary determinant of performance of an internet connection.	
• Emerging EDC and mobile technologies can make conducting clinical research easier, and improve the quality and timeliness of data collection.	
• EDC challenges include clarification of source data and careful separation of patient identifiable data and study data.	