Electronic Data Collection *for* Global Health Research

Sabrina McCutchan
Research Design & Analysis Core
Version 1.1 • 20 August 2020

Duke GLOBAL HEALTH INSTITUTE | Research Design & Analysis Core

A Brief Introduction to EDC

Duke Global Health Institute (DGHI) researchers are collecting primary data in over forty countries around the globe every day.[1] Much of this work occurs in resource-constrained environments where implementing technology solutions has historically proven difficult. But now, DGHI researchers have various options for **electronic data collection (EDC)** in even the most remote parts of the world.

The goal of this guide is to equip readers with the knowledge and resources to adopt an EDC approach. Switching from paper-based to electronic data collection can improve data quality, accelerate the timeline between collection and analysis, reduce costs, and simplify collaboration for cross-cultural teams. Whatever the goal, read on to learn how to use EDC solutions in your research.

Contents

Introduction

Global health researchers have been collecting data on paper for decades. This approach remained necessary even as health research in the United States shifted to favor electronic data collection (EDC), in large part because many sites of global health research lagged behind in access to technological resources. Researchers have adapted both capably and creatively to these constraints. However, even the best-developed paper collection approach has drawbacks.

Transcribing data from paper forms into an electronic system for processing is time - and effort-expensive, and errors are unavoidable. Team members often make choices with later implications for the data, such as how to interpret messy handwriting or deal with values beyond a field's allowed range. These types of manual data processing introduce opportunities for data quality to degrade. A high volume of manual data processing can even bring the validity of later analyses into question, especially if changes to data values are not rigorously documented at every stage.

Paper-based data collection also necessitates the logistical work of managing physical records. Distributing blank instruments to field locations, protecting them from environmental factors that may cause damage, and

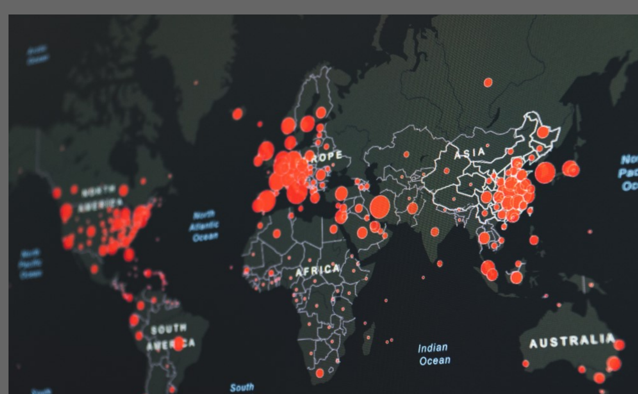
physically securing records containing sensitive data all require coordinated efforts across a research team. In short, paper-based data collection expends a great deal of time and energy on meticulous manual recordkeeping that teams might prefer to spend other ways.

Electronic data collection circumvents many of these challenges, improving data quality while reducing the logistical burden on research teams. This guide will equip readers to determine if an EDC approach can be used in their research, and if so, how to choose appropriate resources. First, it introduces key considerations and resources for those seeking EDC solutions. It then briefly defines relevant terms before addressing advantages and limitations of EDC approaches. It also reviews major considerations that guide choosing an EDC application suitable for a project. After reviewing a selection of EDC applications used at DGHI, it concludes with recommendations based on project needs.

The Research Design and Analysis Core (RDAC) at DGHI also offers consultations, referrals, and dedicated project support related to EDC. Researchers can request assistance using the online RDAC Faculty Support Request Form.

EDC and COVID-19

As COVID-19 spread globally in 2020, handwashing became the first line of defense against the virus. But good hand hygiene is not so easy for displaced populations like refugees and asylum seekers, many of whom live in camps with limited access to water. How could governments and NGOs track COVID-19 spread through these populations? Could they capture data quickly enough to respond?



The UN Refugee Agency (UNHCR) turned to EDC for solutions. They overlaid data from the WHO COVID-19 dashboard on top of Water, Sanitation, and Hygiene (WASH) data UNHCR collects with EDC.[2,3] This allowed real-time monitoring of COVID-19 positive tests in places with lower WASH scores, helping UNHCR spot and intercept outbreak risks.

What is electronic data collection?

"Electronic data capture" is a term originally coined in clinical trials to distinguish data collection using electronic devices from the typical paper-based approach.[4] It applies to non-clinical research as well, though people outside a clinical context may use "collection" interchangeably with "capture." Briefly, **EDC** is the process of using an electronic system to collect data. **Mobile data collection (MDC)** is a similar term for EDC conducted specifically through use of a phone, tablet, or other mobile device, usually in the field rather than in a controlled research environment.[5]

This guide refers to an individual software product that facilitates EDC as an **EDC application**. **EDC system** refers to *all* the software and hardware that collectively works together to accomplish electronic data collection. An EDC system typically includes one or more EDC applications alongside other applications for tasks like file storage and backup, plus hardware like electronic devices for accessing data and servers for storing data.

Spreadsheet applications like Excel or Google Sheets are not included in this definition of EDC applications. Furthermore, spreadsheet and EDC applications are different from **databases**. Imprecise definitions and overlap between these types of software can cause confusion, but this guide differentiates the three terms as follows:

Spreadsheet Application

Stores data in cells within singular tables. Cells may also contain operations that run on the data. Minimal options for querying and little data quality protection.
Examples: Microsoft Excel, Google Sheets

Database Application

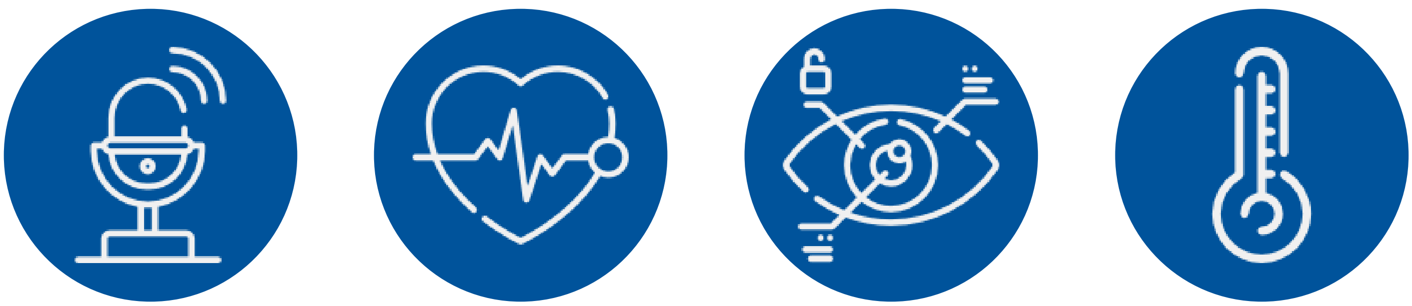
Stores data in a series of tables with defined relationships.[6] A database management system (DBMS) captures and manages data. More options for querying and designed to preserve data quality. Optimal for long-term storage of raw data.
Examples: Microsoft Access, MySQL, Oracle

EDC Application

An application that facilitates EDC. All capture data. Many store data in a database, manage it with a DBMS, display it in tables, and export it in formats compatible with spreadsheet applications.
Examples: REDCap, ODK, Qualtrics

Getting data from a database into a statistical analysis software like Stata or SAS often involves an intermediate step where data is exported into a spreadsheet-friendly format (.csv, .xls, etc). Note that because spreadsheet applications have few data quality protection features, it is advisable to perform data cleaning or analysis with a statistical analysis software rather than within a spreadsheet to prevent data quality degradation.

EDC applications can capture data in different formats. The applications featured in this guide are geared towards collecting **quantitative data** using questionnaires as instruments. This guide refers to questionnaires populated by research staff (either through observation or interview) as **forms** and those populated by a research participant as **surveys.** In clinical contexts, the former type of questionnaire is referred to as a **case report form (CRF).** Different disciplines may also use other preferred terminology.



Examples of other data formats that can be captured electronically include: audio and video recordings, images, biometrics (such as from wearable medical devices), and climate and other atmospheric data. For example, MyCap (a REDCap application) leverages Apple's ResearchKit to push interactive tests called "active tasks" to mobile phones, gathering data on the user's motor skills, cognition, and dexterity.[7] Some EDC applications can capture data in these other formats within a field embedded in a questionnaire. For EDC applications specializing in specific data formats, see *Conclusion.*

EDC vs. paper questionnaires

Until recently, paper questionnaires were the only way to collect data in resource-constrained settings. This made them a common fixture among global health research projects. Even when EDC applications became more pervasive, early technologies relied on having continuous internet connectivity, making them unsuitable for many global health contexts where research teams collect data in the field.

Greater offline capabilities, increasingly reliable performance, expanded features, lowering costs, and increasing technical literacy worldwide now make EDC an attractive alternative to paper questionnaires, though limitations remain. Consider the following advantages and limitations of EDC when deciding between EDC or a paper-based approach.

ADVANTAGES

Offline data collection

Many EDC applications can now collect data offline. Data are collected on a mobile device that runs the EDC application and stores data in local device memory until network connectivity is established (through a cellular data network, WiFi, broadband connection, etc). Once connectivity is established, the EDC application syncs locally-stored data to a designated server. Most sync automatically and also allow manual syncing and export of data files. Applications sometimes stumble over problems such as network connections dropping mid-sync or unexpected discrepancies between data stored locally and on the server. These issues can be reduced by testing the application thoroughly before data collection begins.

Better data quality

EDC applications help optimize data quality at the moment of collection. Collecting data electronically eliminates issues like ambiguous handwriting and the entry errors that occur during manual transcription. Built-in databases prevent data loss from badly formatted or mis-sorted spreadsheets.

Additional features allow the researcher to precisely define how data are encoded. Skip patterns and branching logic ensure that only relevant data are collected for each participant, e.g. only asking participants about pregnancy status if they report sex as female and age within a set range. Data validation features allow instrument designers to restrict the range of possible values per field, preventing impossible values like April 31st, and to define choice menus that prevent undesired write-in responses. Other features prompt data collectors to double-check values that are improbable but not impossible, such as if a person reports exercising seven days a week in the fifth week of a study when they typically exercise once a week.

Improved auditability

Data quality features greatly reduce errors in raw data but cannot reduce them to zero. In the event that errors do occur, EDC applications keep audit trails showing how a data record came to exist in its present state. Audit trails typically include usernames, date and time stamps, flags on impacted fields and/or records, and device serial numbers or IP addresses for every change made to a record. This information facilitates diagnosing where the data went wrong and deciding how to resolve errors. This is a strong advantage over spreadsheet applications like Excel, which often lack audit trail features and thus have limited ability to spot or reverse errors after the fact.

In 2013, University of Massachusetts students **debunked a famous economics article** after finding errors in the raw data spreadsheets.[8]

Saves time

All the above gets clean data into an analyst's hands quicker. With an EDC application, manually entering handwritten data into an electronic system is unnecessary (see *Figures 1* and *2*). Standardized data encoding from the electronic instrument eliminates time analysts would otherwise have to spend cleaning data downstream. Analysts don't have to manually recode outliers (like an age of 40 from a study of children) or obviously impossible combinations of variables (pregnant men) because the EDC application refused to accept entry of such data in the first place.

Builds data collection into project workflows

Project and data managers also benefit from features that save time and effort. EDC applications can tie some aspects of data collection and management directly into the larger project workflow. Many

Figure 1. Data collection using paper questionnaires.

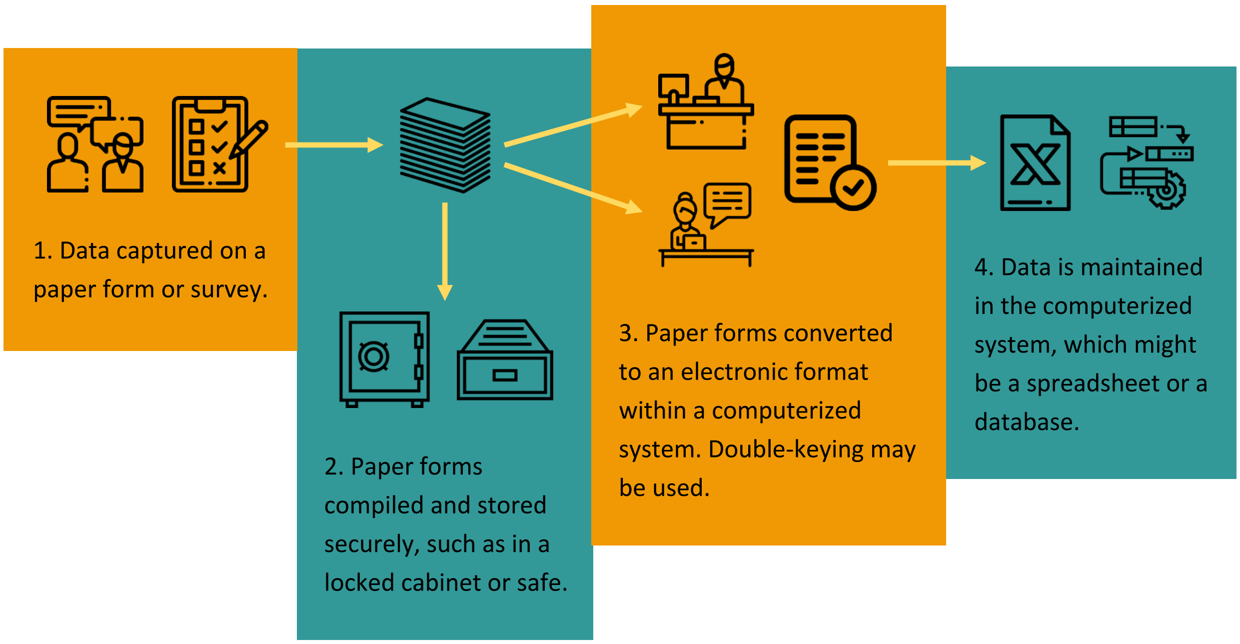
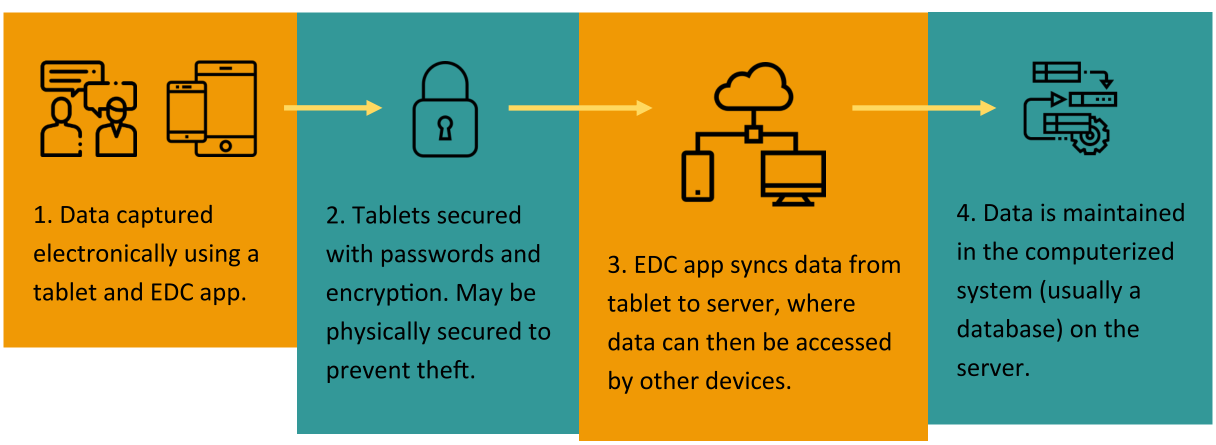


Figure 2. Data collection using an EDC application.



include features for scheduling data collection (e.g., emailing participants reminders to complete surveys) or moving data out of the application to the next point in the workflow (automated downloads/exports). Third-party applications can bridge the gap when these features aren't built in. It's also easier to distribute new versions of instruments electronically rather than in paper form, as they become instantly available upon upload to the application. All this frees project and data managers up to focus on other responsibilities besides record management.

Reduces cost

It's a misconception that EDC is prohibitively expensive in low-resource settings. Three things make it viable: 1) universities offer expanded pools of hardware and software tools to researchers at zero or reduced cost, 2) EDC applications have become more sophisticated and cheaper, and 3) technology proliferation has leapfrogged other types of development so that many populations worldwide now have access to technologies they previously did not. In many cases, the cost comparison of EDC versus paper forms is the cost of purchasing technology versus the cost of staff hours needed to achieve what an EDC system does automatically.

As alluded to in the introduction, the major advantages of using EDC applications are better quality data that produce better results, all while reducing time the research team spends on recordkeeping and logistics. That said, EDC has some limitations.

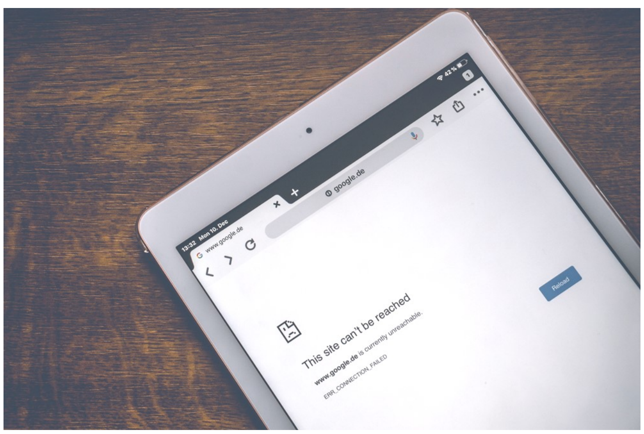
LIMITATIONS

Front-loaded time investment

Successfully deploying an EDC system requires lead time to design and test. Software compatibility issues or problems with data formats can bring research to a halt; these should be identified and resolved before data collection begins. This works in the researcher's favor, as changing an electronic instrument or database mid-collection is unadvisable due to repercussions for data quality. But project budgets and timelines may sometimes make it difficult to dedicate early effort for EDC system testing.

Requires greater technical literacy

Average technical literacy has increased with technology's global proliferation. With some trial-and-error, many people can navigate unfamiliar applications on laptops, tablets, phones, and other devices. However, if the group of people responsible for data entry (either staff or participants) has lower average technical literacy, the effort required to familiarize them with an EDC system may be prohibitively expensive. Also, in situations where it is challenging to build on a knowledge base—perhaps due to high staff turnover or other factors—it will be difficult to maintain elaborate EDC systems over time. Researchers in these conditions should favor simple systems with quick time-to-adoption.



Workflow interruptions from technology failure

EDC is susceptible to interruptions from hardware and software failure. Batteries die, applications update, tablets overheat, and network connections drop. The downside of rapid software development is that two applications working together without issue one day may stop working together the next. This becomes more likely to occur, and more frequently, the more applications involved in the EDC system. Many of these interruptions can be prevented or resolved quickly. But in the worst-case scenarios they may result in data loss or project delays.

Limited languages for user interfaces

Most EDC applications support data collection instruments in the world's most commonly spoken non-English languages, including Spanish, Arabic, and Mandarin. Many can support *instruments* written in nearly all languages. However, the *user interfaces* for some EDC applications (e.g. dashboards, menus, help documentation) have limited language options. This may be a barrier for international research staff. Applications that have been around longer are more likely to have expanded language options for user interfaces; ODK Collect, for example, has been translated into over fifty languages.[9]

Choosing an EDC application

After deciding to pursue EDC, the next question is: which EDC application is best for the project? Consider the following factors in this decision.

Data structure

Will data be collected on different observational units, such as individuals and households? Are data longitudinal? If the answer to either of these questions is 'yes,' choosing an EDC application with a DBMS will dramatically simplify the task of merging and managing data.

Privacy and security

Global health research often collects **individually identifiable health information (IIHI).** This sensitive information is subject to national laws, regulations, and Duke institutional policies. Failing to protect research participants' private data can result in great harm to the participant, the researcher, and the institution. Choosing an EDC application that is sufficient for the privacy and security needs of a given project is crucial. However, it can be challenging to determine what exactly those needs are.

Privacy

The first assessment usually made at DGHI is whether the data are **protected health information (PHI)** under the HIPAA Privacy Rule.[10] HIPAA is US law, but several countries have now also implemented national privacy laws governing their citizens' data. Duke entities providing guidance for legal and regulatory matters are the Office of Counsel and the Office of Audit, Risk and Compliance. RDAC is also able to connect researchers to guidance.

Security

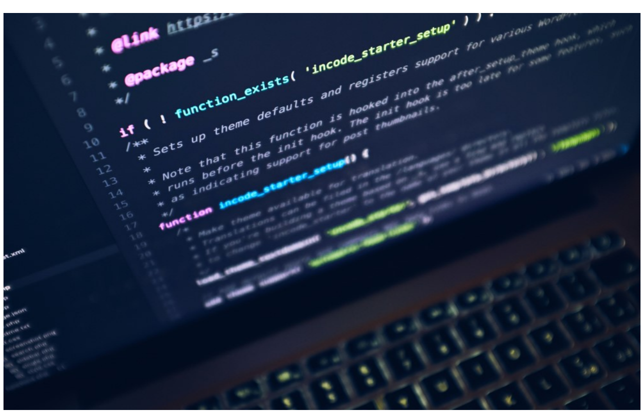
Security features vary by software and hardware. Consider access credentials, encryption, and any privacy policies (e.g. by company) or end-user agreements attached to the application. Even if an EDC application is itself secure, if the transmission channel that moves data from the application to

> Many applications say they encrypt data, but not all encryption is created equal.

a server is not encrypted, or if the server is not firewalled, data are vulnerable to intrusion. Researchers may consult with an IT security professional in Duke's Office of Information Technology (OIT) or Duke Health Technology Solutions (DHTS) for help securing an EDC system end-to-end.

Cost

The cost of an EDC system can be broken down into costs of software and hardware products and costs of staff who develop and maintain the EDC system. These all vary widely by project. In terms of software costs, some EDC applications are open-source, others have already been procured by Duke under an institutional license, and still others will require a one-time purchase or recurring subscription. Calculating the cost of staff effort should include software development, instrument creation, keying data, troubleshooting, and any other relevant activities.



Technical skill

As a general rule of thumb, there is a tradeoff between cost and ease of use for EDC applications. More user-friendly applications are more expensive and cheaper applications require more technical skill. More technical skill may also be needed when building EDC systems that involve multiple applications and/or pieces of hardware.

Lead time

How much time is there before data collection begins? Usually, if an EDC application is cheap in terms of product cost it will be more expensive in terms of time and effort required for set-up. Allowing lead time for testing gives researchers more flexibility to find the best performing application at the best price for their project.

E-consent

Consenting a research participant electronically has many advantages. **E-consent (eIC)** is especially useful for studies collecting participant-entered data where staff will not have much direct contact with study participants. Designing e-consent to hold up under legal scrutiny requires a holistic approach to questionnaire design, including navigability of the software delivering the consent and content of the communication, to facilitate the participant's ability to make an informed choice.

Summary statistics vs data quality checks

Most EDC applications can provide basic summary statistics on data, usually displaying them in tables or other formats on an application dashboard. Data quality checks are a more sophisticated capability to scan incoming data for values that may indicate data quality issues. Depending upon the EDC application, data quality checks may include looking for responses that break pattern (unexpected outliers), questions participants frequently leave unanswered (may indicate problems with survey design), or even custom rules to flag records meeting certain criteria for further review. Some applications, like Qualtrics, take it a step further and use machine learning and artificial intelligence to generate "predictive insights" based on incoming data.

Customizable account permissions

Customizability of user account permissions varies widely across EDC applications. Some have a handful of standard permissions schemes—usually abilities to view, create, edit, or delete data—and a range of options for how they can influence other accounts, such as by adding new users. Others offer more nuanced permissions schemas with finer control. Researchers on international or multi-site projects should pay special attention to how permissions options for collaborators outside Duke might differ from options available to those within Duke, especially if the application is managed by an institutional license.

Featured applications

This guide reviews three EDC application suites : REDCap, Qualtrics, and ODK. These are featured either because Duke has made them available to community members (REDCap, Qualtrics) or because the applications are widely used in global health research (ODK). See the table *EDC Applications at Duke* for a high-level comparison.

Microsoft Office is an application suite. It includes applications like Word, Excel, and Powerpoint.

While REDCap and Qualtrics are both specific EDC application suites, people use ODK to indicate both the OpenDataKit application suite and the numerous other applications built on this original open-source software. This guide calls such derivative applications **ODK-based applications** and reserves ODK for the original software. CartONG published a benchmarking of MDC applications in 2017 for the United Nations that includes a detailed comparison of different ODK-based (and other) applications.[11] These technologies change quickly and information may be outdated.

A major differentiating factor is that REDCap and Qualtrics have both already been approved for Duke community members to use. ODK and ODK-based applications do not have this same preapproval. Researchers desiring to use any EDC application not already approved at Duke will need to obtain approval before collecting any data. Duke's Office of Academic Solutions and Information Systems (OASIS) can offer guidance for the approval process.

THE BIG THREE

REDCap, short for **Research Electronic Data Capture**, is "a secure web application for building and managing [forms] and databases."[12] It was originally developed for clinical research data at Vanderbilt. REDCap's development within an academic research environment is reflected in the structure and features of its applications. The original REDCap application lets users create databases ("projects") that intake data using forms ("instruments"). Other REDCap MDC applications support offline data collection (REDCap Mobile) and participant-entered survey data (MyCap).[13]

EDC Applications at Duke

	REDCap	Qualtrics	ODK-based*
Duke institutional license	✓	✓	
Cost	$ - $$	$	$ - $$$
Types of Data			
PHI	✓	✓	✓†
Longitudinal data	✓	✓-	✓
Instruments			
Offline data collection	✓	✓	✓
Piping	✓-	✓-	✓
Non-English languages	✓	✓-	✓
Software Features			
Database Management System (DBMS)	✓		✓
Forms (data entry by staff)	✓		✓
Surveys (data entry by participants)	✓	✓	
E-consent	✓		
Stock reports	✓	✓	✓
Data quality checks	✓	✓	✓
Customizable account permissions	✓	✓-	✓

** features vary by application* *† requires DHTS/OIT approval* *✓- yes, but limited*

$ cheapest *$$ mid-range* *$$$ most expensive*

Universities, NGOs, and other entities gain access to REDCap by joining the REDCap Consortium. Each consortium member has their own instance of the REDCap software. "Each [REDCap] system is independently maintained and supported," so the Duke instance of REDCap does not grant access to other REDCap instances.[14]

At Duke, REDCap is maintained by the Duke Office of Clinical Research (DOCR). Infrastructure sunk costs, including server space and software programming support, are covered by the School of Medicine. Anyone with a Duke NetID may request free access to REDCap. DOCR also offers some free support services, including office hours. Information is available through the DOCR website.

REDCap is not always free. There may be some costs associated with using REDCap for research data collection, especially if: 1) the project's database is complicated enough to require additional technical support or 2) the project uses a REDCap application still being tested by Duke. In these situations, researchers may be asked to hire a DOCR programmer for some percent effort.

REDCap is approved to collect Protected Health Information (PHI) at Duke. Duke's REDCap stores data on School of Medicine servers compliant with HIPAA requirements. Other EDC applications that are not already approved must be reviewed and approved by Duke before collecting PHI. Importantly, this requirement also applies to other instances of REDCap because they store data on other servers.

Qualtrics is a private experience data (X-data) company offering a variety of products for data collection, management, and analysis.[15] Qualtrics products are primarily designed for companies performing marketing and operations analytics. Duke has an institutional license with Qualtrics for their Core XM product that makes it available to all Duke NetIDs. Like REDCap, the Duke instance of Qualtrics is approved for collecting PHI.

Qualtrics' marketing orientation distinguishes it from other EDC applications in a few ways. First, it's a survey application optimized for participant-entered data. Second, it's well-suited for qualitative data. Third, because many users are companies needing persistent access to customers, Qualtrics offers more robust options than REDCap or ODK in terms of automating and customizing direct-to-individual communications. Fourth, Qualtrics has “smart” features built into the application that REDCap and ODK lack. It uses artificial intelligence and machine learning to provide feedback about survey design—including number, length, and type of questions—and to proactively perform dynamic data analysis and report on patterns in data.[16]

Despite these advantages, Qualtrics is not particularly well-suited for quantitative research data collection. The biggest technical disadvantage of Qualtrics Core XM is its lack of a DBMS that links data records across surveys. Researchers collecting data on different observational units, such as individual- and household-level, will require alternative means of merging records--such as linking a custom database to Qualtrics through API or exporting data out of Qualtrics before merging in a third-party application like Stata.[17] Data quality features like piping, skip patterns, and value range restrictions are much less powerful than in REDCap or ODK.

ODK, short for **Open Data Kit**, is a suite of free, open-source software for data collection in “resource-constrained environments.” It was developed for humanitarian organizations requiring EDC applications that work in contexts with restricted technology access and prioritizes low-cost, user-friendly EDC support. The ODK suite includes applications for building instruments (ODK XLSForm), managing servers (ODK Aggregate), and more.[18]

Though ODK is free software, the total cost of an EDC system using ODK is unlikely to be zero. Savings on applications are offset by the comparatively labor-intensive task of configuring a custom EDC system. Some software programming knowledge is needed to make full use of ODK features. A custom system also requires finding server space, which may incur additional cost.

ODK has comparatively weak privacy and security safeguards. Forms are unencrypted by default, though encryption can be turned on.[19] Researchers using ODK are responsible for configuring their EDC systems, especially data transmission protocols and server firewalls, to apply appropriate security measures. Doing this requires some knowledge of information technology infrastructure.

Many ODK-based applications require less custom configuration than open-source ODK. They may include 1) more sophisticated user interfaces, 2) server space, and/or 3) secure transmission channels between the data collection application and server. The owners of these applications range from for-profit companies (SurveyCTO, Ona) to not-for-profit groups (KoBoToolbox is maintained by the Harvard Humanitarian Initiative). Exact features vary. Those interested in an ODK-based application should research applications individually using the criteria outlined earlier in this guide.

ODK and ODK-based applications require Duke approval before they can be used to collect PHI.

Duke University may be liable for any mishandling, data breaches, or disclosures made by third-party applications. Duke policy often requires a formal legal agreement with the owners of third-party applications storing Duke data to protect both the institution and researcher, especially if the data being collected are sensitive. Securing approval commonly involves Duke's Office of Research Contracts, Procurement, and sometimes OASIS.

Conclusion

Electronic data collection is more accessible than ever before for the global health researcher. New offline data collection capabilities, reduced cost, improved performance, and features that support contemporary data privacy and security needs make EDC an attractive alternative to paper-based data collection in most cases. A well-designed EDC system efficiently collects high-quality data and empowers researchers to spend more time on scientific discovery and less time on operational logistics.

RECOMMENDATIONS

Use EDC instead of paper forms when possible to streamline data collection and management

Allow enough lead-time before data collection to test EDC applications

Store data in databases for improved data quality and auditability

Use **Duke REDCap** or **Qualtrics** for projects collecting **PHI**

Use **Duke-approved EDC applications** like REDCap or Qualtrics when lead-time is short

Discover other options through the following Duke resources:

- Center for Data and Visualization Sciences (CDVS)
- Clinical and Translational Science Institute (CTSI), especially myRESEARCHnavigators and the Duke Mobile App Gateway
- Duke Office of Clinical Research (DOCR) for all REDCap-related questions
- Office of Academic Solutions and Information Systems (OASIS), part of DHTS
- Research Design and Analysis Core (RDAC)
- Schedule a consultation with RDAC's Data Manager to discuss using EDC for your next research project at DGHI

Glossary

EDC
"Electronic data collection" or "electronic data capture." The process of using an electronic system to collect data.

MDC
"Mobile data collection." EDC conducted specifically through use of a phone, tablet, or other mobile device, usually in "the field" rather than a controlled research environment

EDC application
An individual software product that facilitates EDC.

EDC system
All the software and hardware that collectively works together to accomplish electronic data collection.

Spreadsheet application
An application that stores data in cells within singular tables. Cells may also contain operations that run on the data.

Database
An organized collection of data stored and accessed electronically. The most common type, a relational database, models data as rows and columns in a series of tables.

DBMS
"Database management system." The software that interacts with end users, applications, and the database itself to capture and analyze the data.

CRF
"Case report form." A paper or electronic questionnaire used in clinical trials.

Forms
Questionnaires populated by research staff.

Surveys
Questionnaires populated by research participants.

IIHI
"Individually-identifiable health information." A subset of health information that identifies or can reasonably be used to identify the individual.[20]

PHI

"Protected health information." IIHI becomes PHI when a covered entity creates, maintains, or receives the information.

eIC

"Electronic informed consent." Using electronic systems and processes that may employ multiple electronic media to convey information related to the study and to obtain and document informed consent.[21]

ODK-based applications

A software application based off of original open-source ODK software.

References

1 Duke Global Health Institute. (n.d.). *Where we work.* Duke School of Medicine. Retrieved June 18, 2020, from https://globalhealth.duke.edu

2 United Nations High Commissioner for Refugees. (2020). UNHCR WASH [Data dashboard]. Retrieved from https://wash.unhcr.org/wash-dashboard-for-refugee-settings/

3 CartONG. (2020, April 6). *CartONG'S work in support of the COVID-19 pandemic response*. https://www.cartong.org/news/cartong-work-support-covid-19-pandemic-response

4 Hamad, F. (2017). Health information systems: Clinical data capture and document architecture. In C. Urquhart, F. Hamad, D. Tbaishat, & A. Yeoman (Eds.), *Information systems: Process and practice* (pp. 233-253). Facet Publishing.

5 Satterlee, E., McCullough, L., Dawson, M, & Cheung, K. (2015). *Paper-to-mobile data collection: A manual* [White paper]. FHI360. https://www.fhi360.org/resource/paper-mobile-data-collection-manual

6 Oracle. (2020). What is a database. Retrieved June 18, 2020, from https://www.oracle.com/database/what-is-database.html

7 Apple Research & Care. (2020).

ResearchKit. https://www.researchandcare.org/researchkit/

8 Thomas Herndon. (2020, June 14). In *Wikipedia.* Retrieved from https://en.wikipedia.org/w/index.php?title=Thomas_Herndon&oldid=962515252

9 Brinkhurst, M. (2019, August 27). Taking data collection offline with Open Data Kit and Mapbox. *Medium.* https://blog.mapbox.com/taking-data-collection-offline-with-open-data-kit-and-mapbox-4228ad172e9f

10 U.S. Department of Health and Human Services. (2013, July 26). *Summary of the HIPAA privacy rule.* https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

11 CartONG. (2017). *Benchmarking of mobile data collection solutions* [White paper]. https://blog.cartong.org/wordpress/wp-content/uploads/2017/08/Benchmarking_MDC_2017_CartONG_2.pdf

12 REDCap. (n.d.). *Software.* Vanderbilt. Retrieved June 18, 2020, from https://www.project-redcap.org/software/

13 REDCap. (n.d.). *REDCap mobile device applications.* Vanderbilt. Retrieved June 18, 2020, from https://projectredcap.org/software/mobile-app/

14 REDCap. (n.d.). *About.* Vanderbilt. Retrieved June 18, 2020, from https://projectredcap.org/about/

15 Qualtrics. (2020). *What is experience management.* Retrieved June 18, 2020, from https://www.qualtrics.com/experience-management/

16 Qualtrics. (2020). *Online survey software.* Retrieved June 18, 2020, from https://www.qualtrics.com/core-xm/survey-software/

17 Rahul. (2018, August). *Integrating Qualtrics with SQL database* [Online forum post]. Community. https://www.qualtrics.com/community/discussion/1738/integrating-qualtrics-with-sql-database

18 ODK. (2020). *Software.* Retrieved June 18, 2020, from https://getodk.org/software/

19 ODK Docs. (2020). *Encrypted Forms.* Retrieved June 18, 2020 from https://docs.getodk.org/encrypted-forms/

20 Electronic Frontier Foundation. (n.d.). *Individually identifiable health information.* Retrieved June 18, 2020, from https://www.eff.org/glossary/individually-identifiable-health-information-iihi

21 Food and Drug Administration. (2016). *Use of electronic informed consent: Questions and answers.* U.S. Department of

Health and Human Services. https://www.fda.gov/media/116850/download

Photos from Unsplash, in order of appearance:
NASA @nasa, Franki Chamaki @franki, Hiva Sharifi @hivadotme, Martin Sanchez @martinsanchez, Markus Spiske @markusspiske, Jonathan Kemper @jupp, Javier Allegue Barros @soymeraki, Luca Bravo @lucabravo

Icons from flaticon.com, by Freepik, Eucalyp, Prosymbols