CDC Informatics Research and Development Unit (IRDU) Public Health Informatics Structured
Technology Evaluation

<u>EpiCollect: A Generic Data Collection Tool</u>

V0.5 – May 2011

1. Overview

1.1 At a Glance

EpiCollect is an open source, mobile app for collecting generic form data. While this initial release
suffers from significant limitations, the app shows promise as a way to rapidly design forms, collect data
in the field and share data within a project. It should be considered by informaticians in low resource
environments interested in collecting public health related data using forms.

1.2 Keywords

Generic Data Collection Tools, Forms, Mobile Apps, iOS, Android

1.3 Introduction

The practice of public health frequently needs to rapidly design and deploy forms in order to collect
routine or specific data pertaining to public health or population health. EpiCollect[i] is a form generation
and data collection tool designed by the Imperial College London funded by the Wellcome Trust. Form
generation takes place through EpiCollect's web site that allows users to rapidly design a form using a
web-based drag-and-drop tool and then download the form definition to iPhone/iPad and Android
mobile devices and smart phones. Data collected using these form definitions can then be shared with
other mobile users and / or uploaded to a web site for basic analysis and distribution. Data collected
using the form can be submitted to any web site specified by the form designer, but the EpiCollect
project provides a default web site, hosted on Google's AppEngine, for collecting and performing
analysis.

The current release is not intended for production scale, but was designed as a proof of concept to test
the viability of mobile devices for rapid form dissemination and use for data collection.[ii]

1.4 Public Health Business Process Alignment

Generic data collection tools align to almost every public health business process, but EpiCollect
specifically matches the use cases for:

- Data Collection;
- Process, Store and Analyze Data;
- Conduct Active Surveillance; and
- Conduct Public Health Investigation.

EpiCollect was designed to reduce the amount of time data spends in transit between public health
workers collecting data in the field and epidemiologists analyzing the collected data.

1.5 Determination of Evaluation Dimensions and their Weighting

The domain functionality weighting was increased to 50/100 based on the numerous functions required
by public health for form design and data collection. The Ease of Installation and Ease of Use dimensions

were each reduced to 5/100 because these dimensions are not as important, relatively, as the domain functionality for generic data collection tools. While important, Ease of Installation and Ease of Use are more straightforward for simple, direct mobile apps; therefore their weights are not as high as other dimensions of greater importance for generic data collection apps. All other dimensions retained the default weights of 10/100.

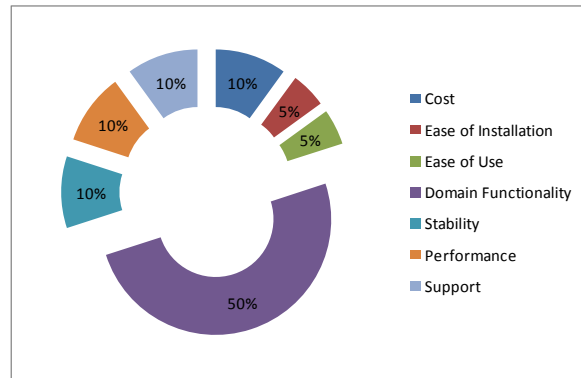| Dimension	Weight
Cost	10
Ease of Installation	5
Ease of Use	5
Domain Functionality	50
Stability	10
Performance	10
Support	10
Total	100

Table 1 Dimension Weights

Figure 1 Dimension Relative Weights

1.6 Alternatives

Many form tools exist within the public health community (e.g. Epi Info[iii], Outbreak Management System[iv], EpiSurveyor[v]) that provide detailed ability to design data collection forms and share data. In particular EpiSurveyor is designed as a web-based mobile system; its mobile client is developed for J2ME phones and has a commercial software license. That being said, EpiSurveyor warrants further analysis and is the closest alternative to EpiCollect.

1.7 Legal / license issues

EpiCollect is released free to use through the iTunes App Store and the Android Marketplace. Their Android mobile client is released under the Apache 2.0 open source license[vi] and the project has stated their plans to release the iOS mobile client and other project source code.

2. Evaluation

2.1 Cost – 4/10

The EpiCollect project is free to use for commercial and non-commercial purposes and the Android application source code is released under the Apache 2.0 open source license. The source code for the iOS app and the web-based application has not been released, but the project has stated their intention to do so at an undetermined future time. There are currently no commercial support options available for purchase, there is no community forum, and the only support available is through emailing the project sponsors with questions and/or problems. As part of this evaluation, IRDU contacted the EpiCollect project and received a response within two days. Although the license cost is zero, support costs could be substantial as a program would need to provide their own resources to correct any defects found in the project.

2.2 Ease of installation – 5/5

The EpiCollect applications were extremely simple to find and install. This evaluation installed the iOS application on an iPhone 3GS running iOS 3.1.3 (7E18); an iPad Wi-Fi running iOS 3.2 (7B367); and a Google NexusOne running AndroidOS 2.2. Although not directly tested, the EpiCollect web site notes that the mobile applications have been tested on 19 devices including: iPhone 3G, iPhone 4, iPod Touch, T-mobile G1, HTC Hero, HTC Pulse, HTC Desire, HTC Legend, HTC Evo 4G, HTC Wildfire, Samsung Galaxy, Sony Ericsson Xperia X10, Motorola Droid, Motorola CliQ, Motorola Backflip and Dell Streak.

Applications were installed using the iTunes App Store and Android Marketplace by searching for the "EpiCollect" title. The program is 187KB and downloaded and installed in less than a minute.

2.3 Ease of use – 5/5

The EpiCollect applications are extremely simple to use and this evaluator was able to design and use forms within 5 minutes without referring to any documentation. A user familiar with Touch devices will recognize the navigation and control functions of the applications.

2.4 Domain Functionality – 25/50

While EpiCollect has basic functionality for data collection, it lacks many features that are required to meet public health data collection needs; specifically: security, vocabularies, related fields, business rules, in-app form design and advanced analytics.

Form Designer: The EpiCollect web site includes a very easy to use and straightforward form designer that lets you drag and drop fields onto your form. The forms are based on the W3C XForms 1.1 Specification[vii] and support field data types of Text, Long Text, Radio Selects, Checkbox Selects and Numeric.

Figure 2 Form Designer

Fields can be marked as required so that the form cannot be completed without specifying a value, but no other business rules are allowed on fields. By extension, no vocabularies can be used to validate form values based on value sets such as Demographics, Notifiable Diseases, etc. from controlled vocabularies such as SNOMED, ICD-9, etc. Not being able to specific value sets from vocabularies or code sets limits the ability to gather pre-validated data that conforms to the needs of a public health response (e.g., limiting input to only specific race, gender value sets). The form complexity supported is the main reason for the low score in domain functionality. Without functions like related fields, specific vocabularies, complex data types and business rule-based validation the public health workforce is limited for how they collect data.
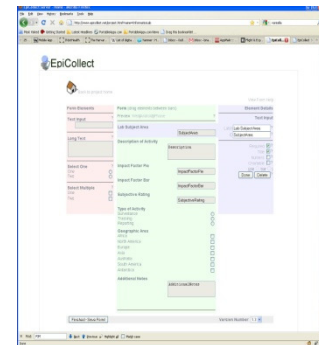
A strength of the form designer is that it is entirely web-based and runs on major browsers such as Internet Explorer 8, Safari 5, Firefox 3.6 and Google Chrome 7. The form designer uses Asynchronous JavaScript and XML (AJAX) effectively to allow users to reorder and modify fields dynamically without multiple server round trips. However, a desktop web browser must be used and the mobile apps cannot design or edit form definitions. Limited editing of form definitions can be performed using the mobile browser included with iOS/Android, but fields cannot be added without using a desktop browser. The

iPad includes Mobile Safari, but the form designer web site does not use a Touch-based Javascript library to allow form design.

The form designer allows for up to three fields per form to be tagged as chartable using the pie and bar chart components of the EpiCollect project home page.

A final useful feature is that the form designer allows the user to preview how the form looks in both iOS and Android prior to saving.

The form designer supports versioning of the form definition so that data collected from different versions can be noted for analysis. This is particularly useful as versions can be noted as minor (e.g., 1.1) or major (e.g., 2.0) so that the user can designate how major the changes to the form are for form users. However, the form version is stored as a field within the form record and all merging and harmonization must be done using separate tools. Records from different versions of the form are not harmonized automatically.

Mobile Applications: Form definitions are downloaded to and used to collect data through mobile apps that run on the Android and iOS platforms. While the different mobile operating systems required the

development of separate code bases, the user experience was very similar between the two platforms. By default, EpiCollect points to a demonstration form, but a user can specify their own form to be used through the settings panel. For this evaluation, this reviewer created a simple form called "InformaticsLab"[viii] to test the data collection capabilities.

Figure 4 Android Main Page   Figure 3 iPhone Main Page
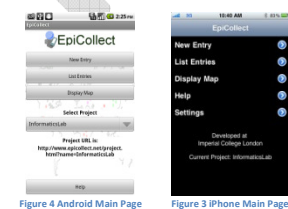
The app downloads a user specified form definition from either the EpiCollect server, or user-hosted server onto the local device. A user then accesses the form to create new records that are saved offline on the mobile device. Each record is tagged with GPS coordinates and can include a photograph. Data is stored locally on the device until the user chooses to synchronize their records with the server. This functionality is extremely useful as it allows the user to use the application in an "offline" mode and then only synchronize data when they are back within a Wi-Fi or cellular data network signal.
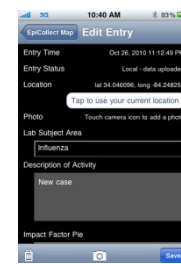

Figure 5 Form Entry Screen

Figure 6 In-App Google Maps

The mobile applications provide basic visualization functionality to display entries using Google maps software and can download records that have been created and synchronized by other users of the form definition.

While being able to share data easily and rapidly with other users of the form is extremely useful in an outbreak situation, this demonstrates EpiCollect's complete lack of security. Form definitions are not password protected, and data submissions are not authenticated and are transmitted over the network without encryption. While this may be acceptable for situations where data is not sensitive or in emergency situations where security concerns have been waived, this lack of

security controls presents a serious risk to confidentiality, data integrity and availability of the investigation.

By default, data is transmitted to EpiCollect's Google AppEngine server; however, a form definition can be privately hosted and data can be submitted directly to a private server. While this private server can implement encryption controls for transporting data (i.e., SSL/TLS), the mobile applications do not support userid/passwords, nor certifications for authentication of submissions. The integrity of the collected data cannot be confirmed without user authentication of who is submitting form data.

Project Home Page: EpiCollect provides a home page on their web site for each form that allows users to browse collected data, visualize data using Google maps, launch the form designer, and to provide basic project description documentation. EpiCollect provides the ability to use a Google user account for authentication to edit the form definition or update project documentation, but all collected data is publicly available without logging in.

The data can be browsed and sorted online for basic analysis and the web site allows the user to export all project data into XML or CSV through the web site and then imported into other applications or systems for analysis. In addition to basic browsing the EpiCollect site provides the capability to view entries using Google maps by geolocation and to visualize up to three fields using pie charts and bar charts. This geovisualization includes a slider to filter records based on date and time. All data can then be exported to Google Earth for additional geographic analysis.

This evaluation only did basic data entry and analysis, but an online demo is available online at spatialepidemiology.net[ix] that includes over 100 unique records created using the mobile app.
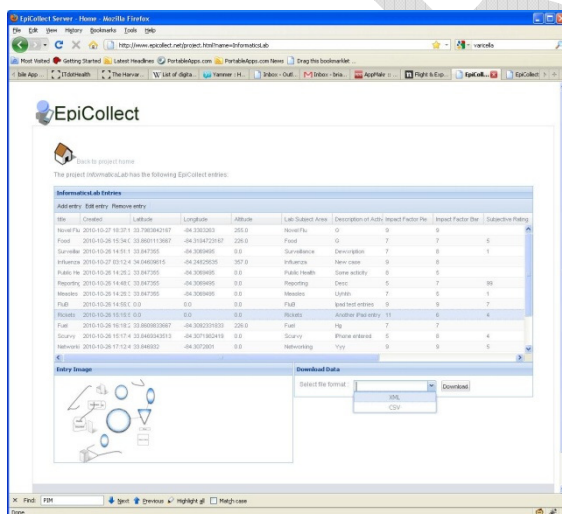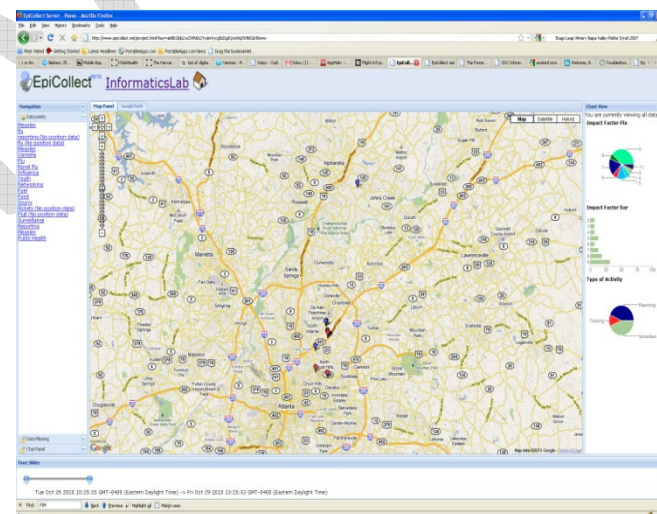

Figure 7 Simple Data Browsing

Figure 8 Geovisualization using Google Maps

During this evaluation, the EpiCollect project team was contacted by email and Dr. Aanensen, bioinformatician with the Imperial College of London School of Public Health, responded that many of the feature gaps identified in this evaluation are planned to be changed in the future. Although a release date for the new release is not yet scheduled, the updated features include: complex forms, multi-linked forms, skip patterns, increased validations (i.e., regex, sums) and relational database support.

2.5 Stability – 6/10

While generally stable, the application suffered from some abrupt crashes and did not perform consistently. For example, after extended use the iOS version would begin to error on touching a specific field within the form preventing entry of any further data without rebooting the device.

The ability to attach and upload photo did not function properly on the iPhone version, but both the iPad and Android devices allowed a photo to be attached to a record.

2.6 Performance – 7/10

Application performance was generally good with app response times all under 1 second to transition between views. However, after entering over 20 items without restarting the app, some fields began to be non-responsive and would not allow you to select specific values. This behavior was not consistently re-creatable, and would affect different fields during each usage session. For example, the InformaticsLab form contains a select-one field called "TypeOfActivity" with three possible values: "Surveillance", "Reporting", and "Tracking". Across multiple users, hardware devices and software versions the Tracking option was not able to be selected. The user would touch this option, but the app would not register the touch. Always, closing the form entry or restarting the app would clear the problem but it would reappear upon further usage.

This kind of inconsistent performance of the application may limit potential usage in the field.

2.7 Supportability – 4/10

EpiCollect's mobile apps run on the iOS and Android platforms. The web-based form designer and data analysis tools run within Google's AppEngine cloud-based python application server. Project owners can also optionally build and host their own data collection services on any platform that supports the HTTP protocol.

While the project has a significant contribution from the Imperial College London, there is no active community outside of the developers. While the application has received significant press by TreeHugger[x], BBC News[xi], Discovery News[xii], and BBC Digital Planet[xiii] the lack of active community could make support of the applications difficult over time.

2.8 Scoring & Visualization

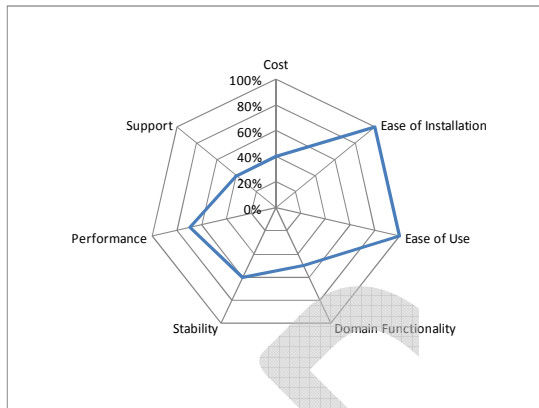| Dimension	Score / Weight	
Cost	4	10
Ease of Installation	5	5
Ease of Use	5	5
Domain Functionality	25	50
Stability	6	10
Performance	7	10
Support	4	10
Total	56 / 100 (56%)	

Table 2 Scoring Table

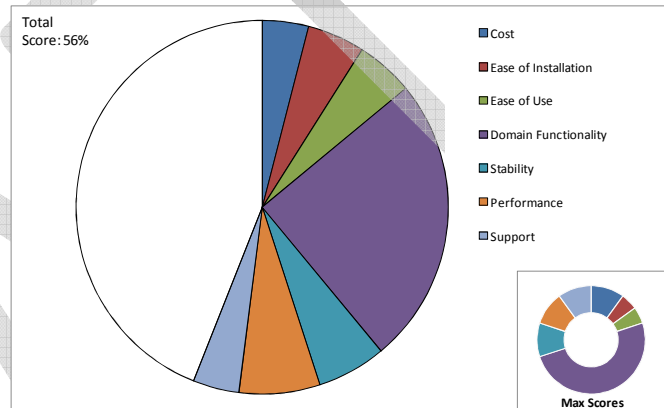**Figure 9 Dimension Ratings (Unweighted)**

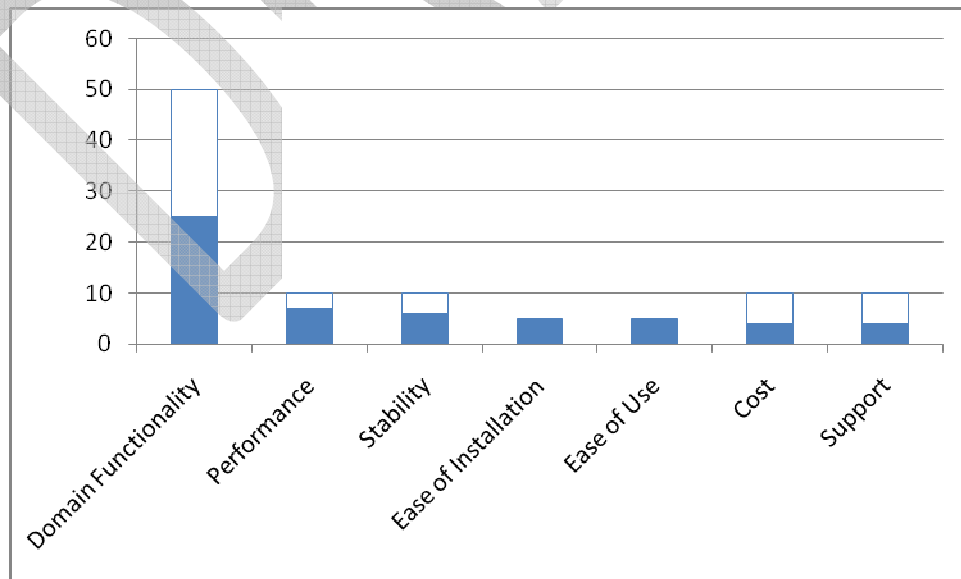**Figure 10 Dimension Ratings (Weighted)**

Figure 11 Ranked Raw Scores

3. Recommendation / Conclusion

3.1 Overall Recommendation

EpiCollect presents an intriguing opportunity for public health workers: the ability to rapidly create, share and collect data using a diverse mobile platform; but hindered by security and functionality concerns. While this initial release could be useful in emergency situations where device selection is limited and existing smart phones must be used, the lack of security means that the collection of any sensitive or potentially sensitive data is wrought with privacy implications. This means that the tool should be limited to training or simulations until the security concerns are addressed in future releases.

3.2 Adoption Timeline

This technology is still in the emerging stage, but based on the dedication of the development team it could be ready for use within the public health toolkit within 2-3 years.
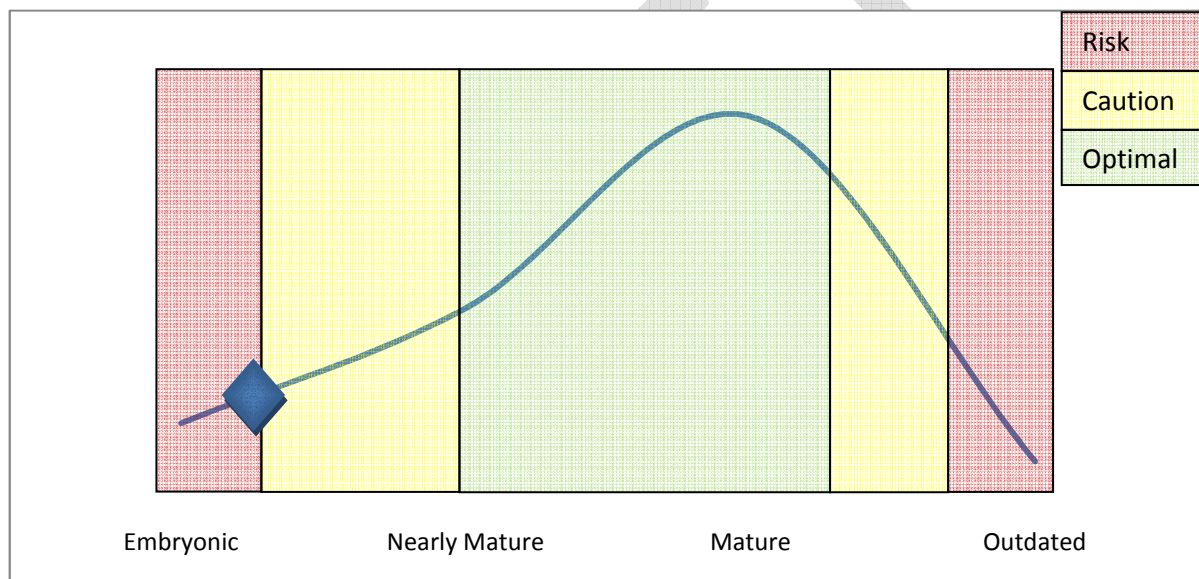

Figure 12 Technology Position on Adoption Curve

3.3 Public Health / Health IT Adoption Impact

This tool- and lightweight, mobile based tools like it- will have a moderate impact on public health practice. While the applicability of this technology in the field is limited as described above in sections 2.4 Domain Functionality and 2.7 Supportability, the open source nature and adherence to W3C form specifications means that either this tool or a tool derived from its open source code base will likely be used for rapid event response or in developing nations that do not have dedicated hardware for data collection.

Since the data collected by this technology is easily exported to more sophisticated analysis tools (SAS, SPSS, etc.) tools such as EpiCollect will be valuable during future public health events.

Appendix A – Peer Review

TBD

Appendix B – Excel Calculation Worksheet

EpiCollect Public
Health Informatics St

[i] EpiCollect http://www.epicollect.net/ Accessed October 20, 2010

[ii] Aanensen DM, Huntley DM, Feil EJ, al-Own F, Spratt BG (2009) EpiCollect: Linking Smartphones to Web Applications for Epidemiology, Ecology and Community Data Collection. PLoS ONE 4(9): e6968. doi:10.1371/journal.pone.0006968

[iii] Epi Info http://wwwn.cdc.gov/epiinfo/ Accessed November 19, 2010

[iv] Outbreak Management System http://www.cdc.gov/ncphi/OutbreakManagement.html Accessed November 19, 2010

[v] EpiSurveyor http://www.datadyne.org/episurveyor Accessed November 19, 2010

[vi] EpiCollect Source Code Repository http://code.google.com/p/epicollect/ Accessed October 21, 2010

[vii] W3C Forms Working Group http://www.w3.org/MarkUp/Forms/ Accessed October 20, 2010

[viii] InformaticsLab EpiCollect Project Home Page http://www.epicollect.net/project.html?name=InformaticsLab Accessed October 26, 2010

[ix] EpiCollect Data Geovisualization Analysis Demonstration http://www.spatialepidemiology.net/epicollect/demo/ Accessed November 19, 2010

[x] "EpiCollect App for Android Puts Laboratories on Phones...Your Phone" 09/18/2009 http://www.treehugger.com/files/2009/09/epicollect-app-for-android-puts-laboratories-on-phones-your-phone.php Accessed November 19, 2010

[xi] "Mobile app sees science go global" 09/16/2009 http://news.bbc.co.uk/2/hi/8258501.stm Accessed November 19, 2010

[xii] "Scientific Data Collection Goes Mobile" 10/22/2009 http://news.discovery.com/tech/wide-angle-scientific-data-collection-goes-mobile.html Accessed November 19, 2010

[xiii] "Digital Planet Podcast" 09/22/2009 http://www.bbc.co.uk/programmes/p0049k4w Accessed November 19, 2010