

**EPI INFO™ CLOUD DATA CAPTURE SYSTEM
IMPLEMENTATION GUIDE**

Version 1.1
08/17/2017

VERSION HISTORY

Version #	Implemented By	Revision Date	Comments
1.0	Sachin Agnihotri	09/30/2016	Version 1.0 of the document
1.02	David Nitschke	08/08/2017	Updates and clarification
1.1	Sachin Agnihotri	08/17/2017	Additional implementation guidance

DRAFT

TABLE OF CONTENTS

1 INTRODUCTION.....	4
1.1 Purpose	4
1.2 Audience	4
2 SYSTEM DESCRIPTION	4
2.1 Key Features	4
2.2 System Operations	5
2.3 System Architecture	5
2.4 System overview.....	6
2.4.1 System Roles and Privileges.....	8
2.4.2 Data Collection Process.....	12
3 IMPLEMENTATION CONSIDERATIONS AND GUIDELINES	13
3.1 System implementation consideration	13
3.1.1 Single deployment vs multiple deployments of the system.....	13
3.1.2 Single Organization Vs Multiple Organizations	14
3.1.3 Single Form vs Multiple Forms	15
3.2 Epi Info™ project type considerations.....	15
3.2.1 Using Epi Info™ SQL Server Project in the System.....	15
3.2.2 Using Epi Info™ Access Project for the System	16
3.3 form sharing and data access rules consideration.....	16
3.3.1 Data collection effort involving users of a single organization	16
3.3.2 Data collection effort involving users of multiple organization	17
3.4 Epi Info™ sql server project implementation guideline	19
3.5 Epi Info™ Cloud data capture deployment guideline.....	20
3.5.1 Dedicated enterprise system for critical systems.....	21
3.5.2 Cloud Data Capture System as a service for small projects	25
3.5.3 Cloud Data Capture System for large and small projects.....	27

1 INTRODUCTION

1.1 PURPOSE

We developed the Epi Info™ Cloud Data Capture system to provide you with a secure web-based tool for distributed data collection and management of data on desktop computers and mobile devices.

1.2 AUDIENCE

The audience for this document includes:

- public health professionals who are users of Epi Info™ 7 for:
 - designing and developing data collection systems;
 - managing and coordinating data entry;
- database administrators who provide:
 - enterprise database services as needed
 - enterprise security configurations.

2 SYSTEM DESCRIPTION

2.1 KEY FEATURES

- **Web enablement:** Epi Info™ Cloud Data Capture lets you publish data collection forms from Epi Info™ 7 Form Designer to the web and to access those forms from any connected device such as personal computers and mobile devices running the Android or iOS operating systems.
- **Data Collection Forms:**
 - Quick response time: Your forms are immediately available to authorized users;
 - Updates: Changes are quick to make. You send them to all users simultaneously as soon as you republish the form;
 - Standardization: Since you control the forms you publish, all users are assured of using the same version which improves quality and facilitates real-time analytics.
- **Versatile entry devices:** You can collect data locally through a web browser from any operating system (e.g., Linux, Windows, Apple, and Chrome) or in the field with tablet computers or other mobile devices.

- **Multi-user data collection:** The system supports distributed and simultaneous data entry, so you and your team work together from the same instance for one or more ongoing studies or events.
- **Centralized data management:** The system can use a single database for all forms and data when not integrated with an enterprise system.
- **Enterprise database integration:** The system can be configured for real-time data synchronization to an additional database representing a specific public health event.
- **Distributed analysis:** Your analyses are done through the Epi Info™ 7 Desktop tools or the Epi Info™ Cloud Analytics system for remote queries.
- **Role based access:** You must be logged into the system by your user name and password. Your activities in the system, and the forms you can see are limited by your role in the organization and the data access level of the forms shared with you.

2.2 SYSTEM OPERATIONS

A typical use case for the Epi Info™ Cloud Data Capture system involves the need for authorized users to securely enter data over the web. The system is used to:

- publish surveys or questionnaires for small or large data collection efforts;
- add new users, manage their access levels within an organization, and assign forms for them to do data entry;
- Add, review, edit and delete data through the Epi Info™ Cloud Data Capture portal;
- Download data into a local Epi Info™ 7 project or directly accessing the data for a SQL Server-based form when automation is enabled.

2.3 SYSTEM ARCHITECTURE

The Epi Info™ Cloud Data Capture system has the following components:

- **Manager Service:** This is a web service that:
 - allows you to publish forms securely;
 - stores the meta-data needed to render the form in the browser;
 - securely downloads the collected data to the Epi Info™ 7 Enter module for Epi Info™ Access projects. Note that when automation is enabled and the project is using an Epi Info™ SQL Server database, the data are immediately available –separate downloading is not needed.
- **Epi Info™ Cloud Data Capture Web Application:** The web application lets authorized users collect and manage data.

- **Epi Info™ Cloud Data Capture Data Service:** The data service provides the web application with the meta-data to render the data collection form, and collect and manage data.
- **Epi Info™ Cloud Data Capture Database:** The system uses a relational SQL Server database to manage:
 - organizations and users within the organization
 - data collection forms published by the organizations; and
 - data entered by authorized users.

The diagram below summarizes these components and their interactions.

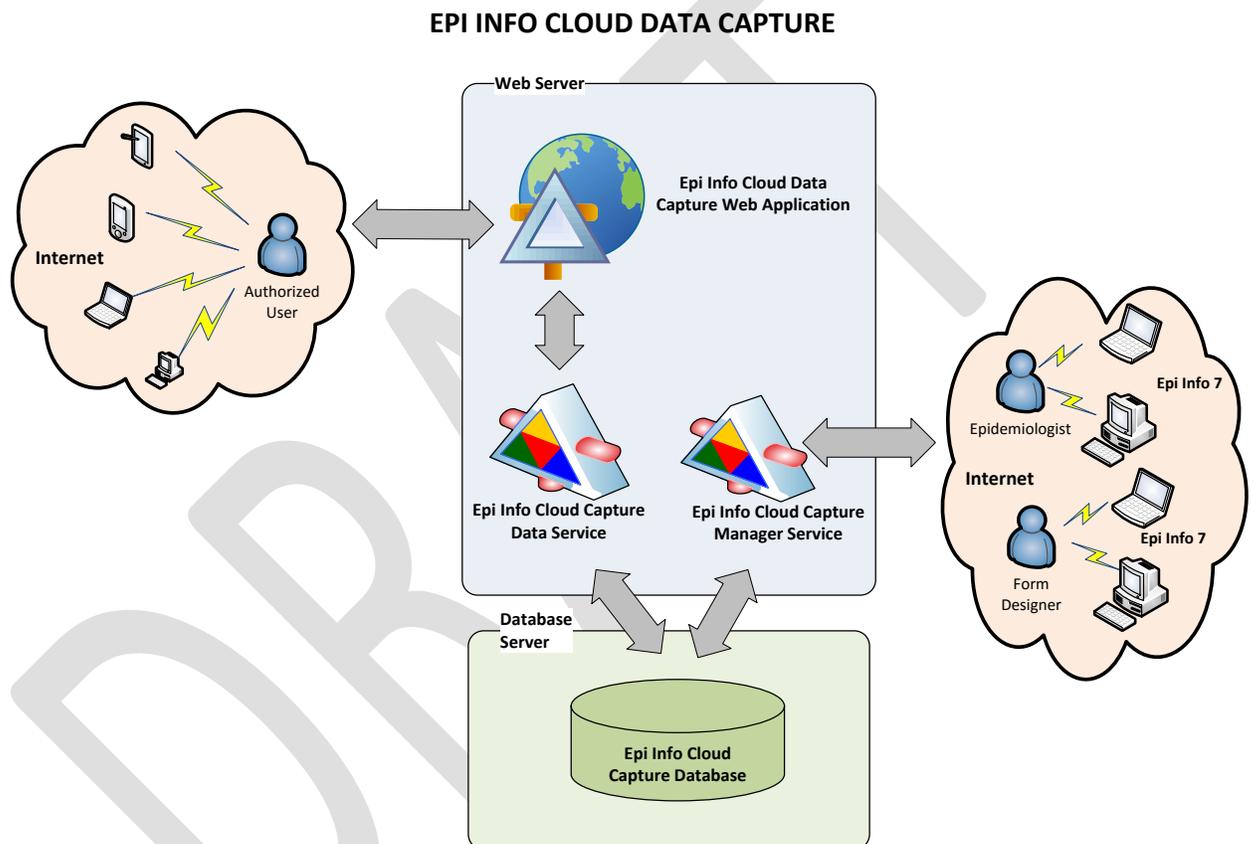


Figure 1: Overview of Epi Info™ Cloud Data Capture System

2.4 SYSTEM OVERVIEW

The Epi Info™ Cloud Data Capture system, at a high level, has the concepts of Host Organization and Super Administrator. A host organization is the very first organization created in the system and represents the primary organization that will manage the system. The Super Administrator is the very first user created in the system. Both the host organization and super administrator are configured by the IT administrator or database administrator at time the system is initially deployed. As with all members in the system, the Super Administrator is identified by their email address and can be a public health professional or IT administrator. The roles are discussed in detail in the following sections.

The Super Administrator, by default, is also considered the Administrator for the host organization represented by “Org 1” in the diagram below. The system can be configured to have only one organization, or it can have several to many organizations according to your needs. If additional organizations are needed, the Super Administrator creates those organizations and designates the Organization Administrator, or Org Admin, for each new organization. The process that creates the organization also specifies its Org Admin. The Super Administrator by default, is also considered the Org Admin for the initial Host organization.

Organization Administrators are responsible for creating users for their own organization. This concept is described in the figure below.

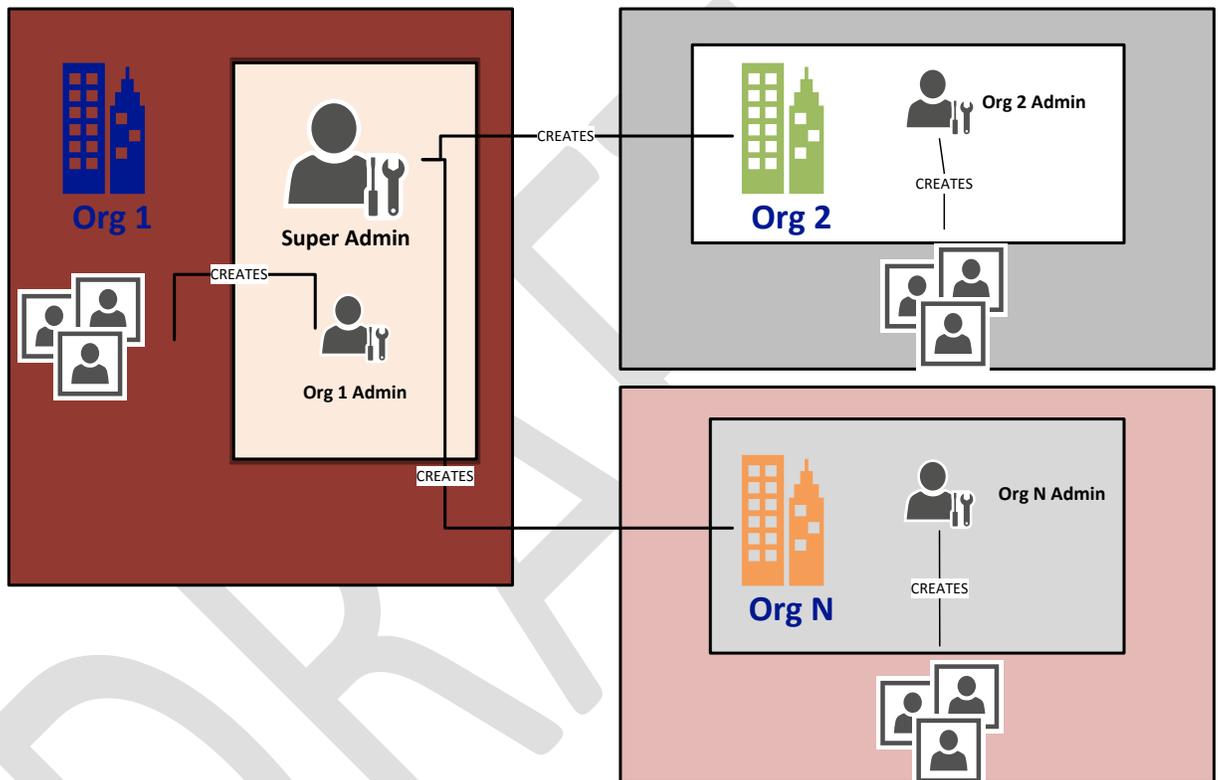


Figure 2: Organization and Users in Epi Info™ Cloud Data Capture System

Every user in the organization shares an Organization Key, which is unique to their organization. Everyone having an organization key has the privilege of publishing forms from Epi Info Desktop to the Epi Info™ Cloud Data Capture portal. This concept is described in the figure below.

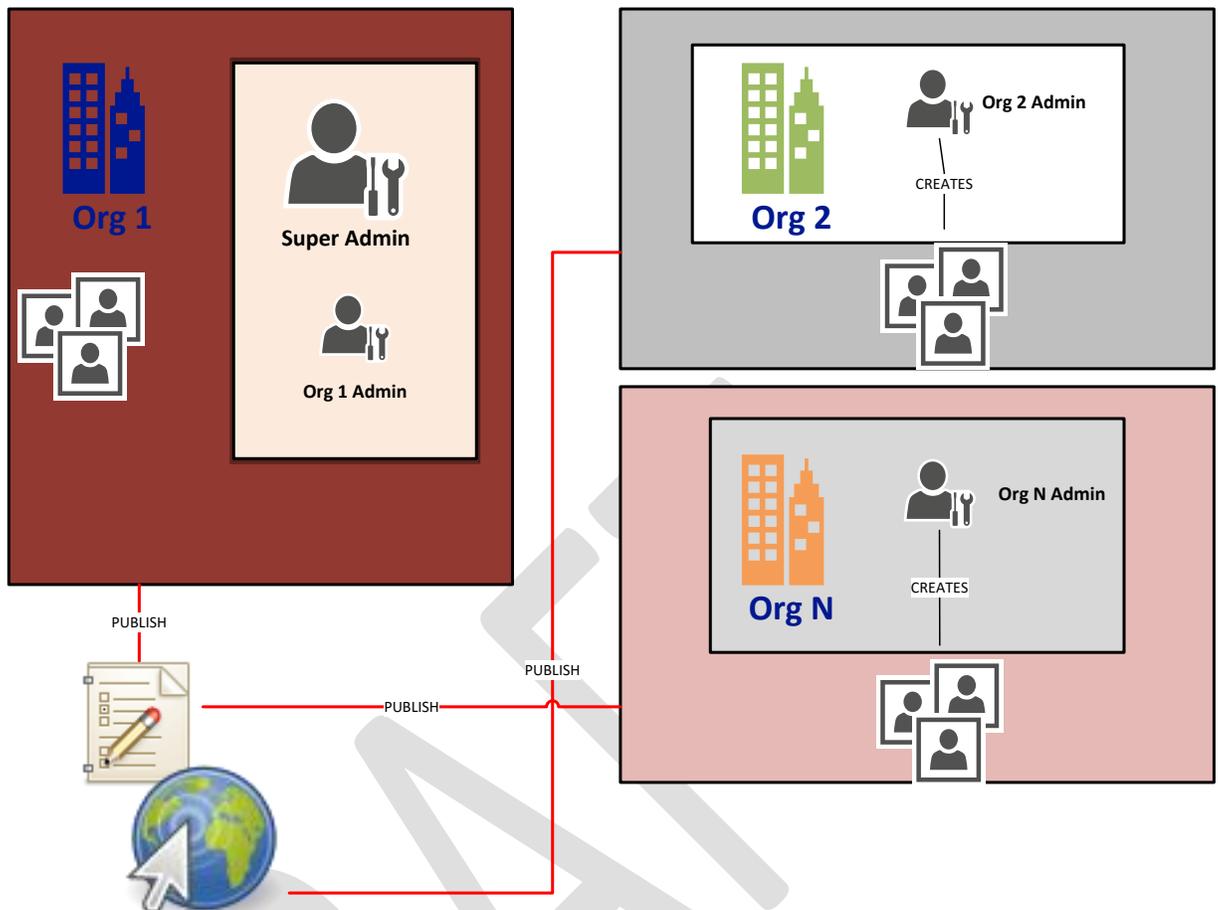


Figure 3: Publish process by user's part of Epi Info™ Cloud Data Capture System

2.4.1 System Roles and Privileges

The Epi Info™ Cloud Data Capture system has three primary roles:

1. Super Administrator
 - a. creates new organizations;
 - b. designates the Org Admin for new organizations.
2. Organization Administrator (Org Admin)
 - a. Creates or adds members only to their own organization
 - b. Designates members of their own organization as fellow Org Admin or only Analyst role.
3. Analyst
 - a. Performs data entry and data management in forms they are assigned;
 - b. Cannot add or change organization membership or role assignments.

The section below describes the privileges assigned to each role

2.4.1.1 Analyst Role

A basic member of an organization is called an Analyst. The ‘Analyst’ name is used to suggest broader integration that can possibly exist with other Epi Info™ products. All analysts within an organization share an Organization Key. Analysts can publish forms to the Epi Info™ Cloud Data Capture system using the Epi Info™ 7 Form Designer. An Analyst who designs the form in Epi Info™ Form Designer and publishes to the Epi Info™ Cloud Data Capture system is considered the form publisher. The Form Publisher is the owner of the published form. As the owner of the form, the publisher can perform several administrative tasks for the published form such as assigning users in their organization the authority to collect data using the form. The figure below provides a listing of privileges the Form Publisher has:

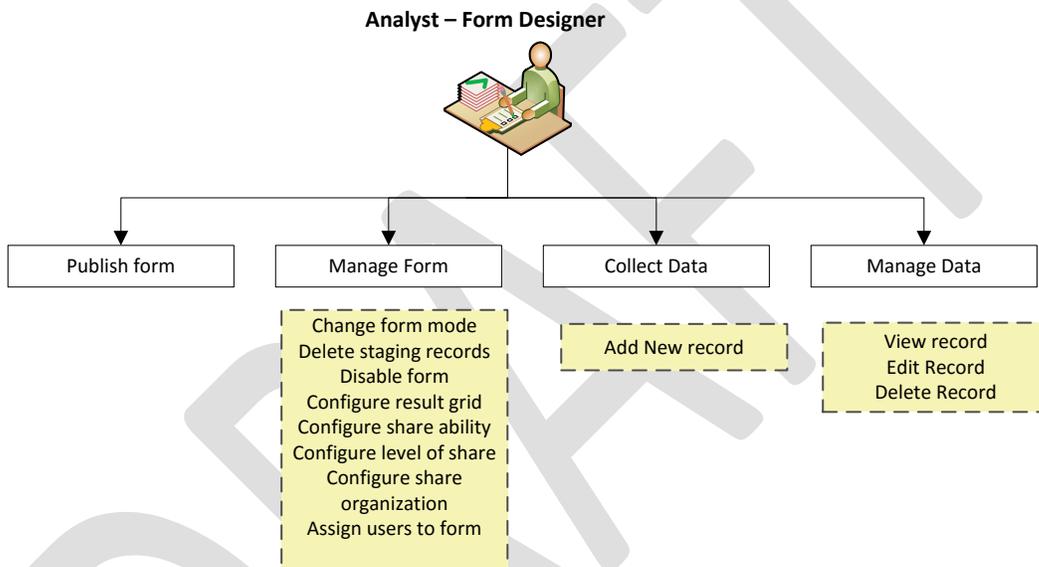


Figure 4: Form Publisher roles in the Epi Info™ Cloud Data Capture System

A member of the organization who has been given the authority to collect data on a form by the Form Publisher is considered a Data Collector. The figure below provides a listing of the authorizations a Data Collector has:

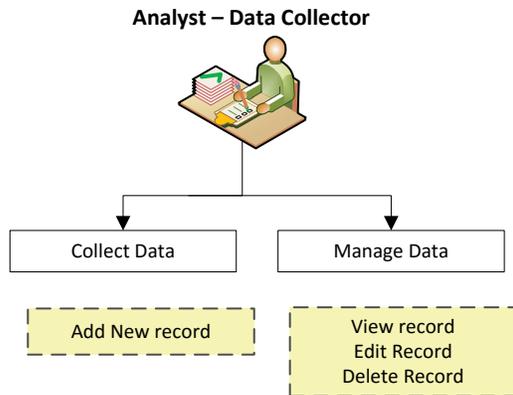


Figure 5: Analyst – Data Collector authorizations in the Epi Info™ Cloud Data Capture System

2.4.1.2 Organization Administrator Role

An initial Organization Administrator, or Org Admin, is designated at the time the organization is created by the Super Administrator. A person having the role of Org Admin has all the abilities of the Analyst Role. In addition to those abilities, the Org Admin can also do the following:

1. add/edit other users in their own organization;
2. promote or demote the role of a member in their organization between Analyst and Org Admin;
3. enable or disable a member's account in the system.

The figure below illustrates the privileges of an Org Admin:

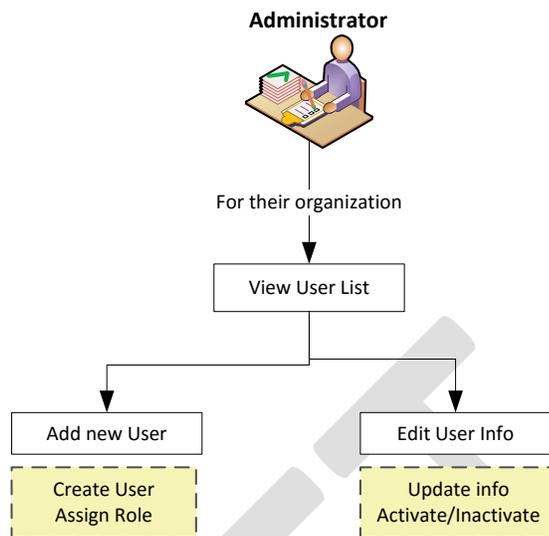


Figure 6: Org Admin’s role in the Epi Info™ Cloud Data Capture system

2.4.1.3 Super Administrator Role

The primary responsibility of the Super Administrator is to add new organizations to the system and to edit, enable, or disable existing organizations.

The IT administrator that sets up the system in the beginning configures the very first organization in the system. We refer to this as the “host organization”. In setting up this host organization, the IT administrator adds the email address of the very first member in the system—the Super Administrator. In doing so, the IT administrator associates the Super Administrator to the host organization as this organization’s Organization Administrator, or Org Admin. Since the Super Administrator is the Org Admin of the host organization, the Super Administrator has the abilities described above, but only within the context of the host organization. The Super Administrator cannot add members to other organizations. Only an organization’s own Org Admin can add members to its organization.

The figure below illustrates the abilities of a Super Administrator:

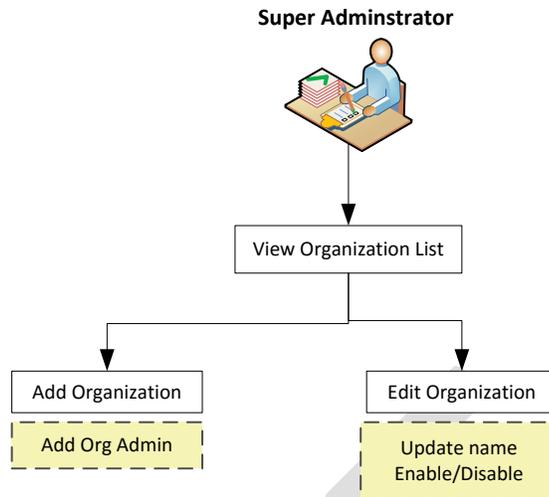


Figure 7: Super Administrator role in the Epi Info™ Cloud Data Capture System

2.4.2 Data Collection Process

The Epi Info™ Cloud Data Capture system supports the following workflows for data entry: As an Analyst, you publish the form from the Epi Info™ 7 Form Designer to Epi Info™ Cloud Data Capture portal using your organization key. The form is immediately available in a web portal for data entry.

When forms are published, they are initially in a draft mode. Forms in this mode display a “Draft” watermark across their pages. You enter records in draft mode during the development and testing phases of the form’s design before the actual data entry begins.

As the form’s publisher, you can assign the form to other members in your organization to assist with testing, piloting, or for data entry. If assigned to your form, the members can enter and manage records regardless of their role in the system.

When the iterative process of testing and editing the form is finished, and the form is functioning as you intend, you can delete the draft records using the administrative features available in the form’s Settings menu. Only the publisher of the form has access to this Settings menu. After you have deleted the draft records, you can change the form’s mode to Final mode.

The figure below describes this process:

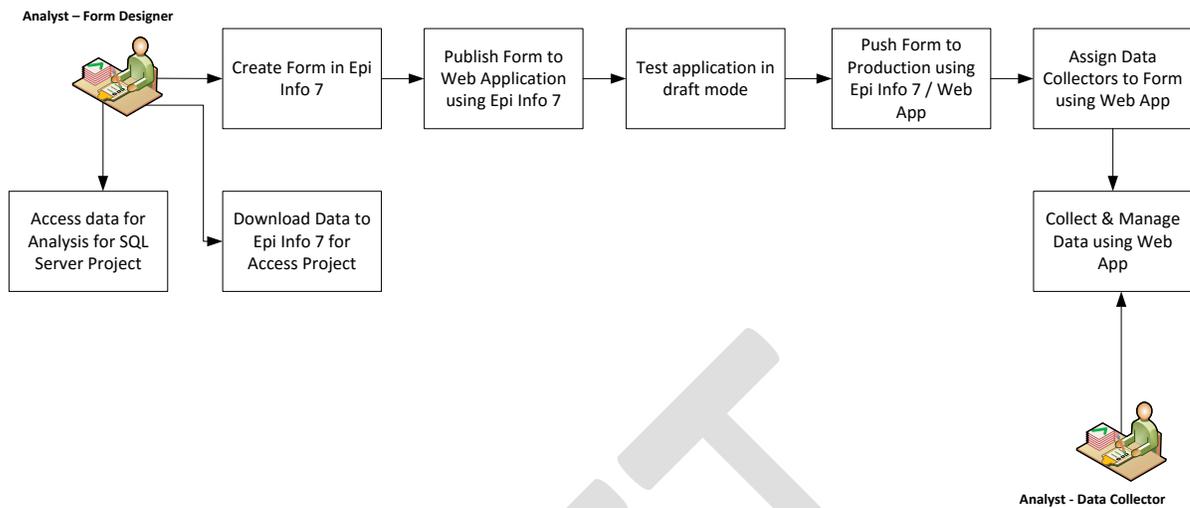


Figure 8: Data Collection process involving an Analyst user in Epi Info™ Cloud Data Capture System

3 IMPLEMENTATION CONSIDERATIONS AND GUIDELINES

3.1 SYSTEM IMPLEMENTATION CONSIDERATION

Epi Info™ Cloud Data Capture system is an open source software program and free to use. It provides considerable flexibility in terms of the number of organizations that can be used within the same system and the number of data collection forms that can be supported by the system. It is up to the organization to determine their own specific use case for the system.

3.1.1 Single deployment vs multiple deployments of the system

3.1.1.1 Single Deployment Considerations

With a single deployment strategy, you would set up the Epi Info™ Cloud Data Capture system as one instance that serves the entire enterprise. Your system acts as a multi-tenant system. You can have multiple organizations and each organization is responsible for publishing their forms and entering data. Alternatively, you can have just one organization in the system with multiple data collection forms published in that organization. Either way, this single deployment scenario results in one application that is used for several purposes.

Resources for both database and web servers are shared. This includes memory, CPU and storage. Since the database is centralized, it stores data for multiple data collection efforts. System performance is proportional to the usage of the system. As such, capacity planning should be done by the IT staff. At the time of a system upgrade, this option requires an upgrade of just one instance of the web application and one centralized database.

3.1.1.2 Multiple Deployment Considerations

The multiple deployment strategy of the Epi Info™ Cloud Data Capture system is where the system is deployed each time a specific data collection purpose occurs. For example, you might deploy it for a specific outbreak, then again for a different public health event, and subsequently for a separate surveillance activity. In this case, you would end up with potentially several web sites created on the web server—one for each specific data collection purpose. These multiple deployments can reside on a single web server, each with its own website and each deployment having its own dedicated database. Optionally, for a large data collection effort, a given website can be configured on a virtual/physical web server. This way you can keep the use case simple in terms of having a dedicated site and a dedicated database.

Whether the sites are on the same server or different servers, your dedicated CPU cores and memory can be assigned according to their expected usage. Having a centralized database for a single data collection purpose has performance gains compared to the scenario where the database size and number of transactions cannot be predicted as in the single deployment model. When it's time to upgrade, however, you need to upgrade the system across all of its websites and databases.

3.1.2 Single Organization Vs Multiple Organizations

You can configure the Epi Info™ Cloud Data Capture system to have a single organization with multiple users, or multiple organizations, each with at least one or many users. If only one member per organization, that member would be, by default, the Org Admin for their organization. This is important to know when sharing forms across organizations.

With just one organization, you control who in the organization has access to the form by assigning the form to the individuals you want to have it. Assigning a form to a member in your organization gives that member full access to all records in the form. There is no restriction as to which records are accessible to the members to which you assign the form.

With multiple organizations, you are not only able to assign full access to members within your organization, but you can also share a form with other organizations in the system. When a form is shared with another organization, it can be done with one of three data access levels. By default, all Org Admins in the target organization have access to the form according to the selected data access level. They may then assign the form to other members, Analysts and fellow Org Admins, within their organization.

The three levels of data access are described later in section 3.3 Form Sharing and Data Access Rules Considerations.

3.1.3 Single Form vs Multiple Forms

Although the Epi Info™ Cloud Data Capture system is intended to host any number of forms, there is no obligation or requirement to publish more than one. Publishing just a single form focused on a specific purpose can help reduce potential confusion and necessary training when your data entry pool has high turn-over or there are many temporary data entry staff members. This scenario is ideally suited for “canned” systems you may want ready-to-go for rapid deployment in emergency situations.

Publishing more than one form, if the need occurs, offers you flexibility and diversity in the purposes with which you can use the system. Data entry into different forms related to different events or studies can be going on simultaneously. Only those members to which you have assigned the form, or organizations with which you have shared the form, will have access to enter and manage data. This access can be reversed at any time. As the publisher of the form, all records entered in the system, whether by your organization or a shared organization, will be in your control.

3.2 EPI INFO™ PROJECT TYPE CONSIDERATIONS

In Epi Info™, you can create your project using a Microsoft Access file as the project database, or, if you have a Microsoft SQL Server available, you can opt to use a SQL Server database for the project. When considering to develop a project to use the Epi Info™ Cloud Data Capture system, your decision about which project database to use will determine the features available to you and the level of effort required to set up the system.

3.2.1 Using Epi Info™ SQL Server Project in the System

The Epi Info™ Cloud Data capture system supports an enterprise implementation using the SQL Server database type. This enterprise implementation is recommended when you have a data collection effort that has or will have:

- a large volume of data;
- a data entry period spanning a long time;
- needs for high data security.

In these situations, it is highly recommended that you consider creating the Epi Info™ project using a Microsoft SQL Server database as the data repository for the project. Using a SQL Server database offers the following features not available with Access-based projects:

- Configure security controls and policies for regular data backups and data recovery.
- Data are immediately available to epidemiologists and data scientists for analysis using the Epi Info™ Visual Dashboard or any other statistical software without having to download data back into the Epi Info™ project.

- Integration with Epi Info™ Cloud Data Analytics to produce web-based analyses, dashboards, and reports online that you can share with others in the organization.
- Sorting and searching of records based on specified data columns.

An Epi Info™ SQL Server project typically requires the involvement of the IT department or a database administrator (DBA) to create the new SQL Server database on the SQL Server and configure the necessary security constraints. The DBA has to provide you with the name of the database and possibly security information in order for you to create the Epi Info™ project. Once the project is created, then you can design the form and eventually publish the form to the Epi Info™ Cloud Data Capture system using your user identity and the Organization Key.

3.2.2 Using Epi Info™ Access Project for the System

Using a Microsoft Access database as the project database is an option when you have:

- a relatively small volume of data in your data collection effort;
- a data collection effort over a short period of time;
- little or no need for extra data security;
- no access to a Microsoft SQL Server or the IT and DBA resources to create the necessary SQL Server database

If your Epi Info™ Cloud Data Capture system is deployed using an Access-based project, then you can quickly design the form and publish it to the web without extra assistance from IT staff. We recommend only considering to use an Access-based project for small data collection efforts or prototyping because Epi Info™ Cloud Data Capture does not allow searching and sorting of the data on the web. Also, while data can be managed on the web, you must download the data back into the Epi Info™ project before you can analyze or export the data. Since you have no built-in backup and recovery procedures, as with SQL Server projects, you will need to manually backup the data file. Finally, the Microsoft Access-based project has limited security if it is put on a shared network drive for use by other members of the organization.

3.3 FORM SHARING AND DATA ACCESS RULES CONSIDERATION

During the planning process of the data collection effort, you need to decide whether or not the data collection effort involves members of just one organization or multiple organizations. The system provides some built-in options to support these various use cases.

3.3.1 Data collection effort involving users of a single organization

This is the most typical scenario of the system. After designing the form in Epi Info™ Form Designer, you would publish the form to the Cloud Data Capture

system. When you first publish the form, it is initially not shareable. If your system has only one organization, then this non-shareable mode is to be used. In this scenario, the system works as an open system. This means that everyone in your organization that you have assigned access to the form will be able to Add, Edit and Delete its data. The default state of not shareable can be found under settings for the form. Only the publisher of the form will have access to its settings.

Make Shareable/Not Shareable:

3.3.2 Data collection effort involving users of multiple organization

You may have use cases where you need data entry from several different organizations or teams at a time. An example of this scenario is a state health department having designed a study on heart disease patients discharged from hospitals. In this example, let's suppose several hospitals are involved. Data collected in each hospital needs immediate availability to the state health department for review and analysis.

In such a use case, there will be several organizations in the system. The host organization might represent the state health department and another organization is created for each hospital. It is better for all hospitals to use the same data collection form, rather than each hospital having their own form which might contain slight differences. To ensure all are using the same version, the form is designed and published to the Epi Info™ Cloud Data Capture system at the state health department and made shareable to the hospital organizations.

There are three data access rules that can be applied to the form depending on how data will be used by the organizations involved. The three options are:

1. Open system
2. Siloed system
3. Mixed system

3.3.2.1 Open system

With this data access rule, all the assigned users in all the organizations have access to add, edit, and delete any record entered into the system. In the example above, the state health department and each hospital can add, edit, and delete any record entered by their organization as well as any other hospital organization. Org admins at each shared hospital organization automatically have access to the form by default. Members with the Org Admin role can assign the form to other members that have the Analyst role. This mode is implemented by the state health department member that published the form by

opening the settings for the form and making the form “Shareable”. Then they would select the data access rule of “Enable data access to all” as shown below.

Make Shareable/Not Shareable:

Shareable | ▼

Select Data Access Rule:

Enable data access to all | ▼ ⓘ

3.3.2.2 Siloed System

With this data access rule, users within a given shared organization can only add, edit and delete records that were entered by their own organization. Records entered by one organization are not accessible to members of any other organization. Continuing with the example above, but in a Siloed system, each hospital can access only its own hospital’s records and not each other’s records. It has full control of their data, and only their data--no other hospital’s data. Even the host organization—the state health department—cannot access data entered by a hospital. This mode is implemented by the state health department that published the form accessing the settings for the form and making the form “Shareable”. Then they would select the data access rule of “Access within organization” as shown below.

Make Shareable/Not Shareable:

Shareable | ▼

Select Data Access Rule:

Access within organization | ▼ ⓘ

3.3.2.3 Mixed System

This data access rule is a compromise between the rules for open and siloed systems. The organization that published the form has access to all of the data entered by all of the shared organizations. However, the other shared organizations can only access their own data. Shared organizations cannot access each other’s data. Using our example from above, the state health department would be able to add and manage records entered by any hospital organization, but Hospital A can only access records entered by Hospital A. Likewise, Hospital B can only see Hospital B’s records. This mode is implemented by the state health department that has published the form accessing the settings for the form and making the form “Shareable”. Then they

would select the data access rule of “Enable data access for central organization” as shown below.

Make Shareable/Not Shareable:

Shareable | ▼

Select Data Access Rule:

Enable data access for central organization | ▼ ⓘ

3.4 EPI INFO™ SQL SERVER PROJECT IMPLEMENTATION GUIDELINE

As described above, we highly recommend using an Epi Info™ SQL Server project to make full use of the Epi Info™ Cloud Data Capture system. When the system is implemented as an Epi Info™ SQL Server project, the systems logical backbone that exists between the Epi Info™ Cloud Data Capture’s centralized database and the Epi Info™ SQL Server project communicate seamlessly to synchronize data between the two databases. For the synchronization to be seamless, the following configuration needs to be present in the SQL Server:

- a. The Epi Info™ SQL Server database should be on the same database server as the Epi Info™ Cloud Data Capture system’s centralized database.
- b. The application user account must have “db_datareader” and “db_datawriter” roles on the Epi Info™ SQL Server database.

Below are the steps you will need to create an Epi Info™ SQL Server project, publish the project, and begin using the Epi Info™ SQL Server project as part of the Epi Info™ Cloud Data Capture system:

- a. First, work with your DBA or IT administrator to get a SQL Server database for your Epi Info™ project. The Epi Info™ database should be on the same database server as the Epi Info™ Cloud Data Capture system.
- b. To create the database, your DBA configures the security to use SQL Server authentication or Windows authentication. The database name can be the same as the name of the project, but it may be different.
- c. The DBA must create an “application user” account in the Epi Info™ Cloud Data Capture’s centralized database. This “application user” account needs to have db_datareader and db_datawriter roles on Epi Info™ SQL Server database.
- d. After the database is created, you should have the following:
 - server name,
 - database name,
 - authentication information (Windows or SQL Server authentication)
 - If SQL Server authentication, then you will also need the

- Application user name
 - Application password
- e. With the information above, you can create the Epi Info™ project.
 - f. At this point, you can design the data entry form in Epi Info™ 7 Form Designer.
 - g. Upon completion of the form, open the form in the Epi Info™ 7 Enter module. This allows Epi Info™ 7 to create the data tables needed by the project from the metadata created during the design phase.
 - h. Return to Epi Info™ 7 Form Designer to publish the form to the Epi Info™ Cloud Data Capture system.
 - i. At this point, you can log into the Epi Info™ Cloud Data Capture system and you should see the form listed in the left panel of the Cloud Data Capture portal.

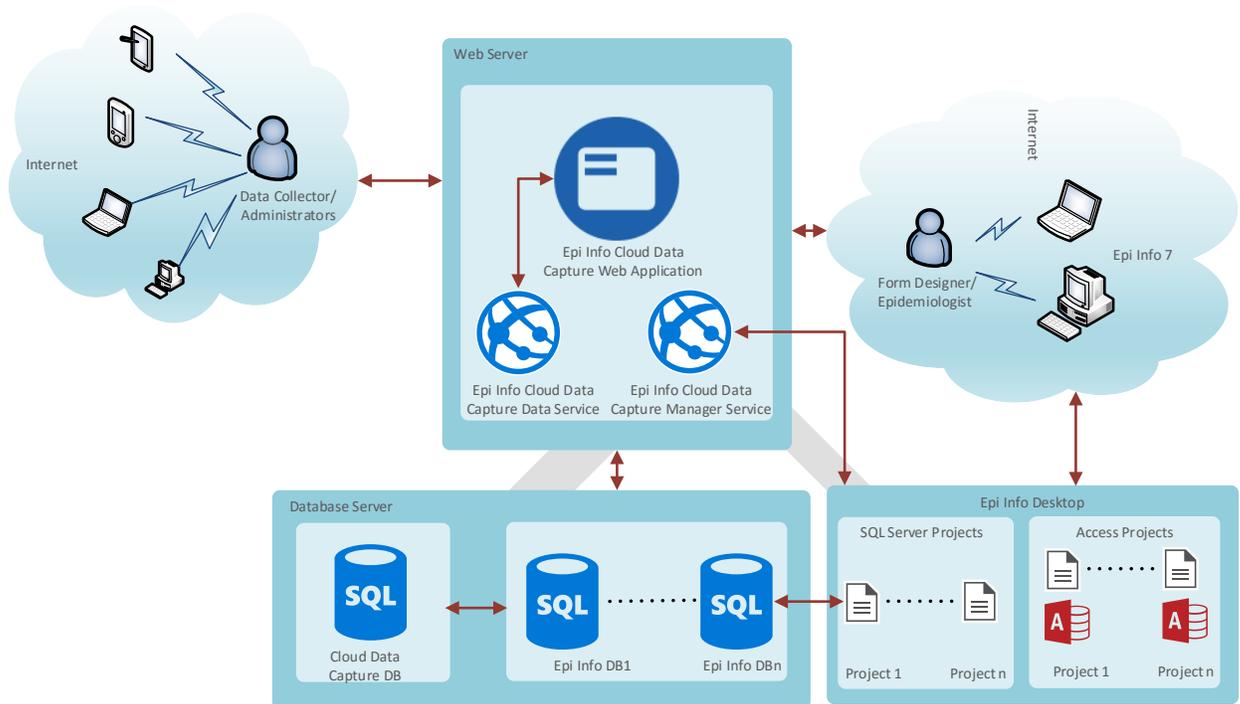
The process of creating an Epi Info™ SQL Server project results in the creation of Epi Info™ schema objects (required tables) in the database. Opening the data collection form in the Enter module results in the creation of data tables for the form where the collected data resides.

3.5 EPI INFO™ CLOUD DATA CAPTURE DEPLOYMENT GUIDELINE

Epi Info™ Cloud Data Capture can be deployed for different purposes as discussed in the section on [single deployment vs multiple deployments](#). The system can have organizations in it as discussed in the section on [single organization vs multiple organizations](#). The system can also support different kinds of Epi Info™ projects as discussed in the section on [Epi Info™ project type considerations](#). The projects are enabled on the web for data collection purposes through the process of publishing the form to the Cloud Data Capture system.

The system deployment comprises of a web server and a database server. The web server hosts all the components of the web application. The database server hosts the systems centralized database. The web application interacts with the centralized database. The centralized database is the repository that manages all the data collection forms and all the data collected on the web. The form designers/epidemiologists interact with the systems manager service hosted on the web server to publish their form. Data collectors/administrators interact with the system's web application hosted on the web server to configure the data collection form and the collect data. Both Epi Info™ SQL Server projects and Epi Info™ Access projects can be published on the web using the systems manager service. For Epi Info™ Access projects, the project's database resides on the form designer's computer and all the collected data for the project in system's centralized database has to be downloaded back to the local Access database before any data analyses. For Epi Info™ SQL Server projects, the database is hosted on the same database server where the systems centralized database is hosted. The system sends the data from the application's centralized database to the Epi Info™ SQL Server database automatically through an integration enabled between the two databases. The data collected by the web application is available

in real time for data analysis. There is no extra step of downloading data back into the project as with Access-based projects. The diagram below summarizes the overview of the system described above.



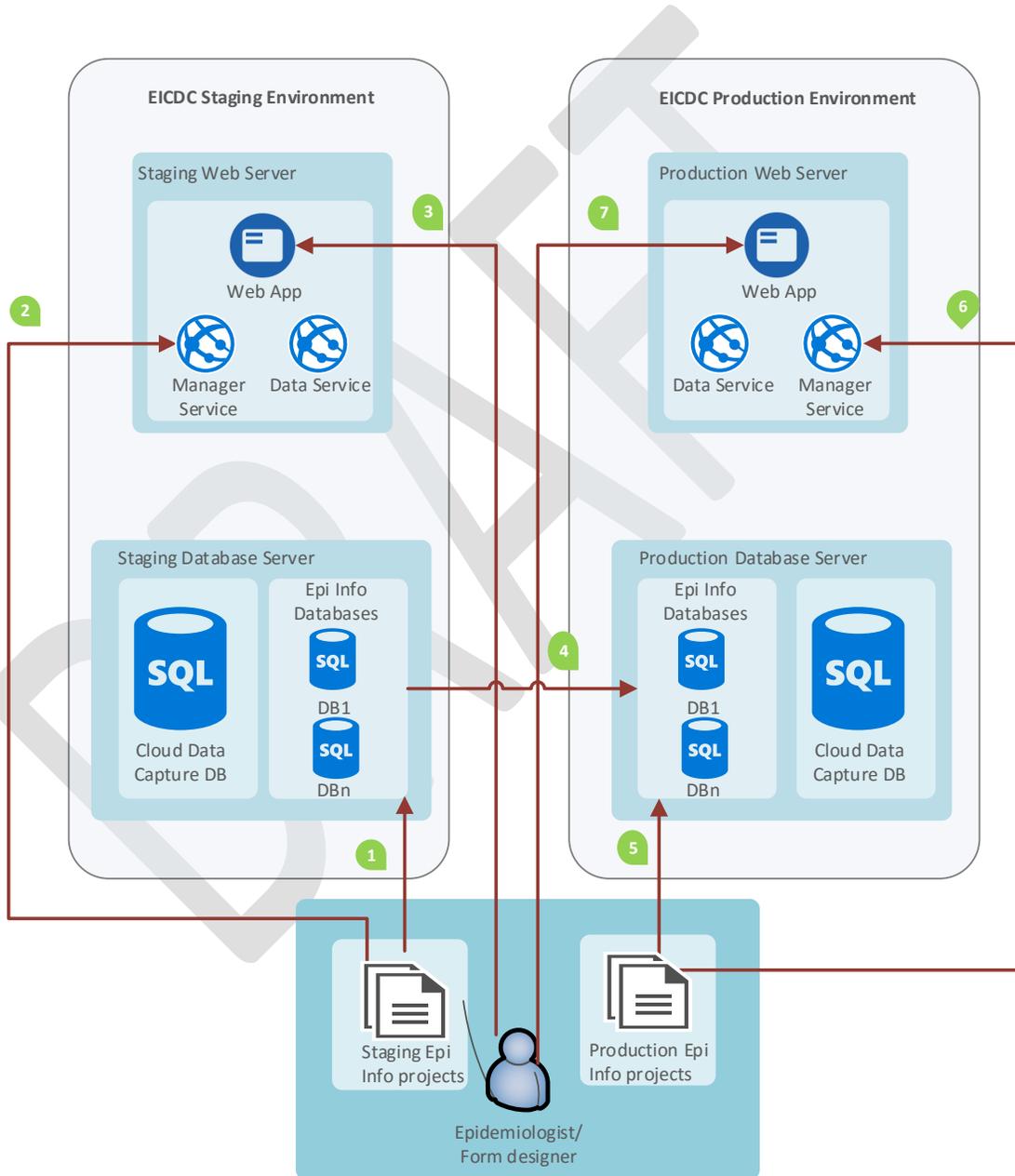
With this information in context, below are details on various deployment scenarios for the Epi Info Cloud Data Capture system.

3.5.1 Dedicated enterprise system for critical systems

In certain cases, there is a need to collect long term data, the size of the database is large (>1000 and could run into several hundred thousand records), the need for database security is high, and data may be sensitive in nature. In such a scenario, the system should be implemented as a dedicated system for the enterprise to support the needed data collection. There would be a dedicated website for the data collection effort, and SQL Server will be used for both the Epi Info™ Cloud Data Capture centralized database and the Epi Info™ project database. The system's centralized database will have only one Epi Info™ project in it. That single project collects data over the web and its database is automatically synchronized with the Epi Info™ SQL Server database.

The process of creating an Epi Info™ SQL Server project results in creation of the Epi Info™ schema objects (required tables) in the database. Opening of the data collection form in the Epi Info™ 7 Enter module results in the creation of the data tables for the form where all the collected data will reside. With the Epi Info™ SQL Server project, an assumption is being made that enterprises do not allow the epidemiologist to create and manipulate database objects directly in production

databases. In such cases, the Epi Info™ SQL Server project is first created in a staging environment and tested using the Epi Info™ Cloud Data Capture system. Once the form is final, the database team migrates the Epi Info™ database to the production environment. Even if the epidemiologist is authorized to create the Epi Info project on the production SQL Server database, after the form goes live for production data collection, it is not recommended to make changes to the Epi Info project on the production SQL Server database. The diagram below shows the various parts of the workflow involving the staging and production environments.



- 1 Epi Info project is created as a SQL Server project with project database on Staging SQL Server
- 2 The Epi Info SQL Server project is published to EICDC system in Staging environment using the systems Manager Service
- 3 The Epidemiologist/form designer tests the published project (form) is in Staging environment
- 4 Epi Info SQL Server project database is created on production database server using scripts, sql tools or using staging template depending on organizations policies.
- 5 Epi Info project file points to Epi Info SQL Server project database in production
- 6 The project is published/republished to EICDC system in production environment after SQL Server database is synced.
- 7 The project is validated in production environment and made available to the data collectors

The most important step in the workflow is handling the deployment of the Epi Info™ project database to the production environment. This can be done as follows when deploying the database to the production environment for the first time:

- a) If the organization is willing, then, only for the first time, the production database can be created using a template created from an Epi Info project in a staging environment. A new Epi Info SQL Server project is created on the production database server using the Epi Info project template from the staging environment. This process assures that everything needed for the Epi Info database is created in the production environment.
- b) If creating the database with a template is not an option, then the DBA team must create the needed scripts to migrate the database from the staging environment to the production environment. During the migration process, all of the tables in the schema have to be migrated to the production environment. The data present in all of the tables that have the prefix starting with “meta” and “code” must be inserted into the database in the production environment. Data for the tables having names matching the form name followed by a number do not need to be migrated to production environment.

After the Epi Info SQL Server project database is created in the production environment, the data collection form needs to be enabled in the Epi Info Cloud Data Capture system for data collection. This can be done as follows:

- a) Create the Epi Info project file using Epi Info by pointing to the Epi Info SQL Server database in the production environment. If IT or the DBA is concerned with sharing the production environment credentials, then they can create the project file.

- b) The Administrator for the Epi Info™ Cloud Data Capture system will create an account for the organization in the production site. The Form designer/Epidemiologist is added to the organization as an Org Admin or an Analyst. The Form designer/Epidemiologist will open the Epi Info project using the Epi Info project file created in previous step.
- c) The Form designer/Epidemiologist will publish the form in the production environment using their identity and organization key.

If there is a need to make changes to the data collection form in the production environment, it is recommended that changes are first made in the Epi Info SQL Server project in the staging environment and tested using Epi Info Cloud Data Capture system in the staging environment. Once the changes are final in staging, the Epi Info SQL Server database needs to be synced with the production version. This can be achieved as follows:

- a) First, compare the Epi Info SQL Server database in the staging and production environments for schema differences. This will identify any new tables that are added with “code” prefix. All new tables whose names match the name of the form followed by a number can be ignored for migration purposes. The delta can be synced using database tools or scripts.
- b) After the new schema objects are created, the staging and production environments are compared for differences in data. The comparison should look for data differences in all the tables having the prefix of “meta” and “code”. Ignore the differences noted for the columns “EWEOrganizationKey” and “EWEFormId” in the table “metaViews”. Tables whose names match the name of the form followed by a number can also be ignored. This delta can be applied to the production environment.

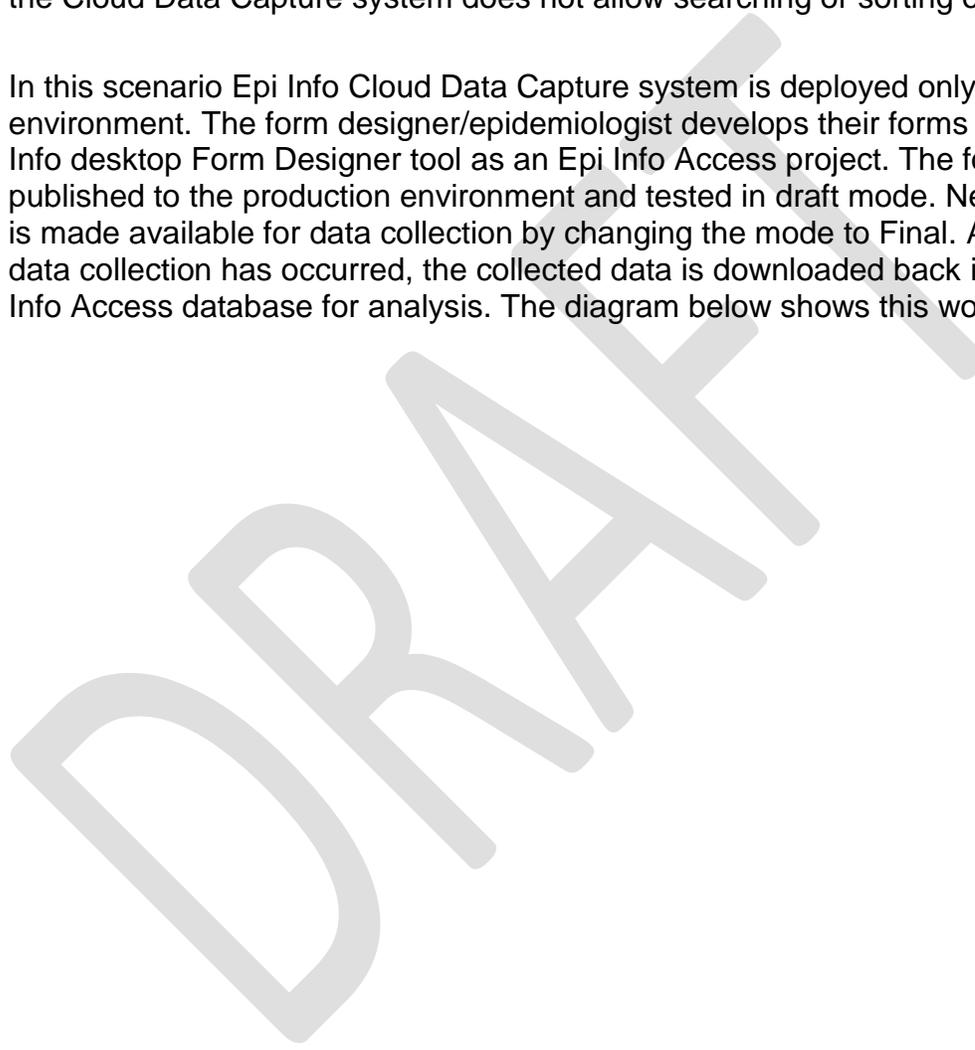
Once the changes to the Epi Info SQL Server database have been applied to the production environment, the data collection form can be updated in the Epi Info Cloud Data Capture system as follows:

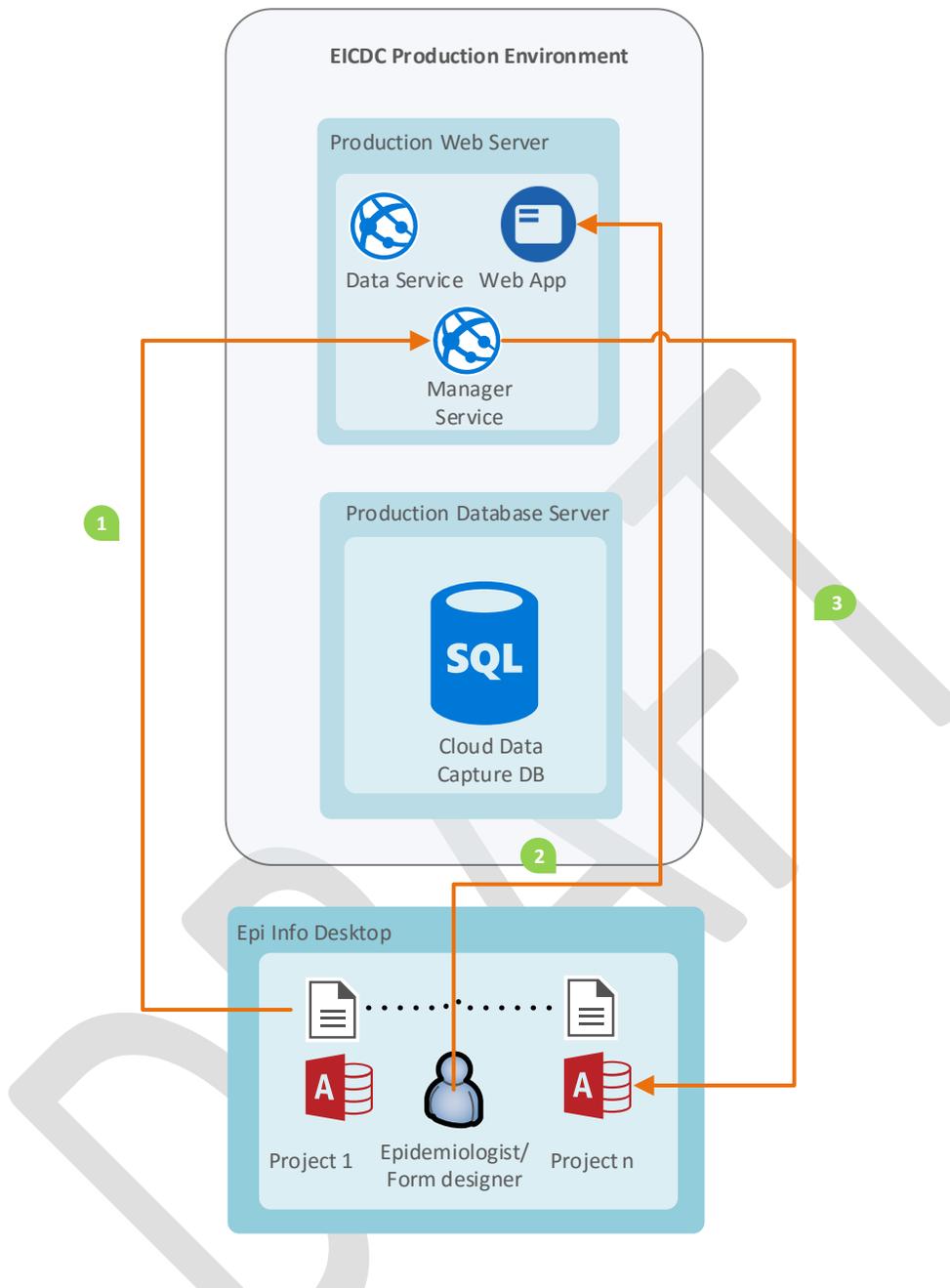
- a) Open the Epi Info project that points to the Epi Info SQL Server database in the Epi Info desktop. The form will now show all the changes that were done in the staging environment.
- b) Perform the process of republishing the form to the production environment using the identity and organization key.
- c) Log into the Epi Info Cloud Data Capture system in the production environment to confirm the changes.

3.5.2 Cloud Data Capture System as a service for small projects

The system can be made available in Software as Service mode for short term studies where the amount of data collected is small. While longitudinal data can be collected, it would not be a good candidate for this scenario. In this mode, users of the system can design a form on their machine as an Epi Info Access project. The form can then be published to the production environment quickly and made available for data collection after testing the behavior of the form in Cloud Data Capture System. The collected data can be downloaded to Epi Info for analysis by the epidemiologist/form designer. In this situation, using an Access-based project, the Cloud Data Capture system does not allow searching or sorting of the data.

In this scenario Epi Info Cloud Data Capture system is deployed only in production environment. The form designer/epidemiologist develops their forms using the Epi Info desktop Form Designer tool as an Epi Info Access project. The form is published to the production environment and tested in draft mode. Next, the form is made available for data collection by changing the mode to Final. After some data collection has occurred, the collected data is downloaded back into the Epi Info Access database for analysis. The diagram below shows this workflow:



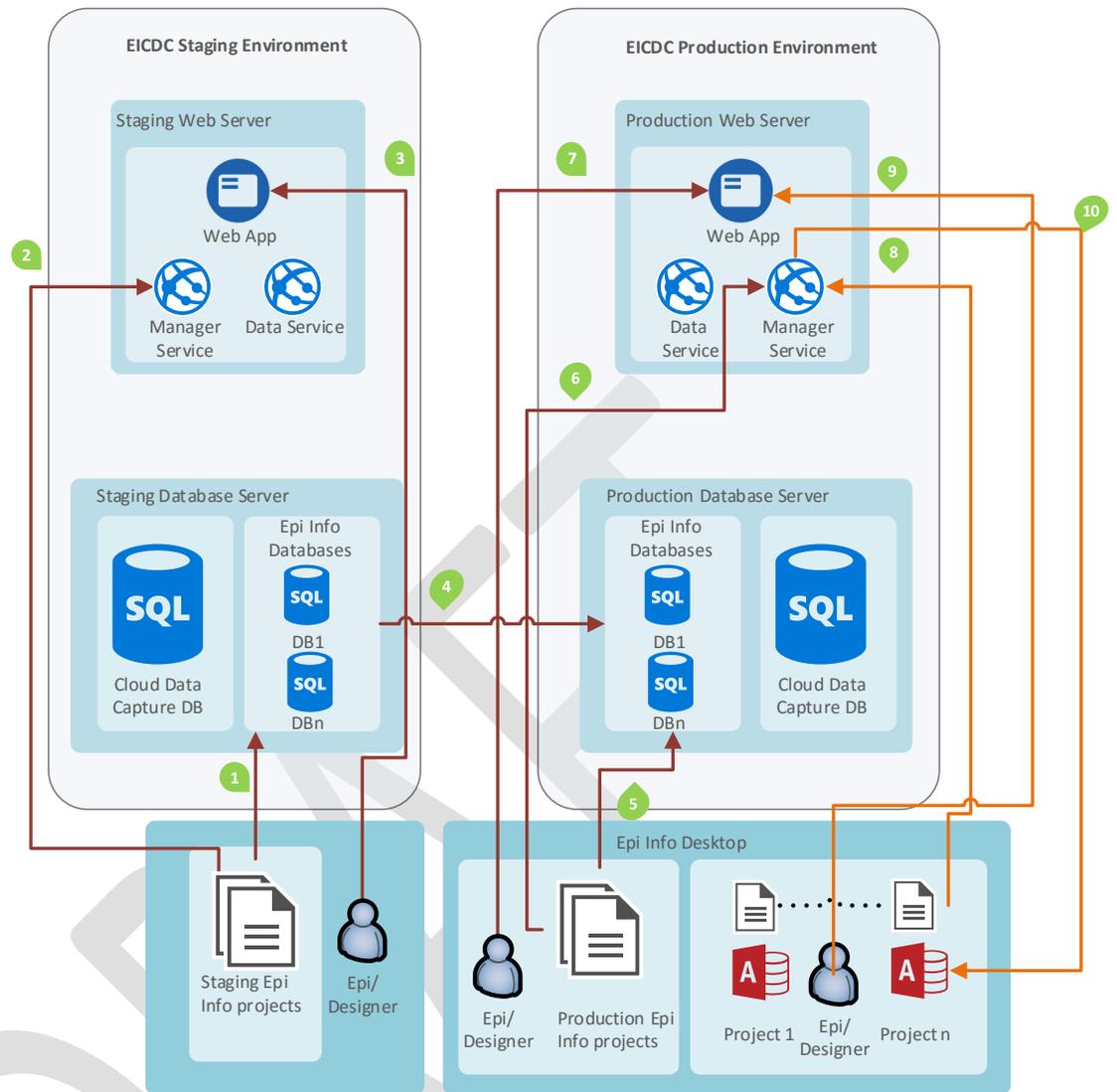


- 1 Epidemiologist/form designer develops the form using Epi Info desktop as an Epi Info Access project and published the project to Cloud Data Capture system in production environment.
- 2 Epidemiologist/form designer tests the published form in draft state in production environment and changes the state to final when form passes all testing
- 3 The epidemiologist/form designer downloads the data to Epi Info Access project on local machine on a regular basis.

3.5.3 Cloud Data Capture System for large and small projects

When the system is deployed as a single instance for an entire enterprise to use, the system can be used for short term projects with low data volumes as well as longer duration projects having large amounts of data. For long term projects or those with large amounts of data or for projects where the need for data security is higher, an Epi Info SQL Server-based project should be used. The system will depend on the IT department for deploying changes from staging to production. As with an Epi Info SQL Server project, the schema of the project cannot be directly manipulated in the production environment. The system can also be used for any number of short term projects or projects having small amounts of data. For such projects, the Epi Info project can be a Microsoft Access project. As compared to an Epi Info SQL Server project, the process of deploying such a project is much simpler. The data collection form is directly published to the production environment where it is tested and made available for data collection by changing the mode of the form to Final. For Epi Info Access-based projects, we DO NOT recommend any changes to the project after testing is complete and the mode is changed to Final.

To support Epi Info SQL Server projects, it is recommended that a staging environment be created where the SQL Server projects can be first tested before deploying them in the production environment. The process to manage changes in the Epi Info SQL Server project has been discussed in [Section 3.5.1](#). Epi Info Access-based projects can be directly managed in the production environment as discussed in [Section 3.5.2](#). The diagram below shows the workflow involved with managing both Epi Info SQL Server-based and Access-based projects:



- 1 Epi Info project is created as a SQL Server project with project database on Staging SQL Server
- 2 The Epi Info SQL Server project is published to EICDC system in Staging environment using the systems Manager Service
- 3 The Epidemiologist/form designer tests the published project (form) in Staging environment
- 4 Epi Info SQL Server project database is created on production database server using scripts, sql tools or using staging template depending on organizations policies.
- 5 Epi Info project file points to Epi Info SQL Server project database in production
- 6 The project is published/republished to EICDC system in production environment after SQL Server database is synced.
- 7 The project is validated in production environment and made available to the data collectors
- 8 Epidemiologist/form designer develops the form as an Epi Info Access project. The project is published to EICDC system in production environment.
- 9 Epidemiologist/form designer tests the form in draft state and changes to state to final when form passes all testing
- 10 The epidemiologist/form designer downloads the collected data to Microsoft Access database on the local machine.