

A Beginner Guide to Research Data collection using EpiData

Quality data collection for research
with EpiData

Myo Minn Oo

Copyright © 2020 Author Myo Minn Oo

All rights reserved.

ISBN: 9798673464779

Imprint: Independently published

DEDICATION

This book is dedicated
to
my family, teachers, and friends.

CONTENTS

CHAPTER 1 Introduction to EpiData.....	3
1.1 Its features.....	4
1.2 Installing EpiData Manager and EntryClient.....	5
1.3 Terminology.....	6
1.4 Help and Documentation	7
Chapter 2 Getting Started with EpiData Manager	8
2.1 The interface of Manager.....	8
2.2 Creating a New Project	9
2.3 Navigating a new project.....	9
2.4 Change the name of the project tree.....	12
2.5 Saving and closing current project.....	13
2.6 Opening existing projects	14
Chapter 3 Introducing a real-life project.....	15
Chapter 4 Creating a codebook.....	19
4.1 Characteristics of a codebook	19
4.2 Name.....	20
4.2 Label.....	21
4.3 Type.....	21
4.4 Range	22
4.5 Value labels	22
4.6 Codebook for Form 1	22
4.7 Solutions – Codebook for Form 1	23
4.8 Additional notes	30
Chapter 5 Designing a data form	31

- 5.1 Changing the name of a data form 31
- 5.2 Design tools 32
- 5.3 Adding headings 34
- 5.4 Adding string fields 35
- 5.5 Adding date fields..... 39
- 5.6 Adding Integer field 40
- 5.7 Using integer field for enter categorical input..... 41
- 5.8 Adding a Memo field 42
- 5.9 Finish the remaining variables 42
- 5.10 Alignment 43
- Chapter 6 Using calculated field 45
 - 6.1 Adding identifier field..... 45
 - 6.2 Creating a derived field..... 47
- Chapter 7 Advanced techniques 51
 - 7.1 Making unique index..... 51
 - 7.2 Jumping to next variables..... 52
- Chapter 8 Getting started with EntryClient..... 55
 - 8.1 Opening an existing project..... 55
 - 8.2 Entering values in fields 56
 - 8.3 Navigating records 58
 - 8.4 Printing records 58
 - 8.5 Deleting records 58
- Chapter 9 Double data entry and validation 60
 - 9.1 Preparing for double data entry 60
 - 9.2 Double data entry 61
 - 9.3 Validation 65

9.4 Generating data validation report.....	68
9.5 Revising and finalization	71
Chapter 10 Exporting data.....	72
STATA.....	74
CSV.....	74
SPSS.....	74
DDI.....	74
EPX.....	75
XML in EpiData.....	75
Chapter 11 Appending records.....	76
Chapter 12 File backup.....	80
Storage drives	80
Zipped EPX File	81
Chapter 13 Archiving files with encryption.....	82
Annex for Shortcut keys.....	86

ACKNOWLEDGMENTS

I sincerely thank Professor Edward B McNeil for taking the time to read the whole book and providing critical comments.

CHAPTER 1 INTRODUCTION TO EPIDATA

EpiData is a collection of freeware that are specifically designed and developed for quality data entry, data documentation, data management, and basic statistical analysis. Its main applications are in public health surveillance, outbreak investigations and scientific research. The development and distribution of the EpiData software is maintained by Denmark based EpiData Association.

The EpiData freeware succeeded its principles from the Epi Info software package was developed by the United States Center for Disease Control and Prevention (US CDC) during the 1980s. In 2000, the US CDC released a new Epi Info version 2000 that used Microsoft Access® database for data storage. Hence, to developing an independent and text-based system, Jens M Lauritsen took the initiative of the EpiData project which later grew into a fully developed data entry and documentation software known as the EpiData entry. It has several advantages in addition to a standalone freeware which include double entry verification, list of ID numbers in several files, codebook overview of data, date added to backup and encryption procedures.

This freeware, EpiData Entry, uses three-file-type system, the so-called QES, REC and CHK triplet. To create a data entry project, users must manually type texts in EpiData's text editor (QES) and later convert into a record file (REC) where data are stored. If checks for data validation are desired, record file must be called in to create checks (CHK). However, most people are not good at manual typing and multiple-file-based system can lead to error if files are not in the same directory. Hence, these two have been major drawbacks of the freeware. After version 3.1 was released, the EpiData Association stopped its further development.

Since 2008, the EpiData Association started developing another similar freeware known as EpiData Manager and EntryClient which will be our focus in this book. As the names suggest, Manger allows users to create projects and develop data forms whereas EntryClient is solely for data entry and record management. The new system added

several features that old EpiData Entry lacked, such as single-file system, better user interface, click and drop function to create data entry fields, improved relational data system and extended user access control system. Yet it still maintains the principle of simplicity. The last one is its cross-platform compatibility meaning that this freeware can be used on Windows, Mac OS as well as Linux operating systems.

The intention of this book is to give readers a rather practical approach to the EpiData Manager and EntryClient for efficient data entry, documentation, and data management. To facilitate the learning process, the use of technical terms is minimized. Instructions are also illustrated using a real-world project. It is my hope that this will enable readers to get started using EpiData freeware with minimal problems.

One important thing to note is that this freeware collection was developed and maintained by volunteers with very limited funding. Without their dedication, EpiData would not be accessible for many people. As a result, documentation and how-to guides are somewhat limited. This is where I hope this book can fill the gap.

1.1 Its features

The first EpiData software was released in 1999. It has been around for more than 20 years now that many aspects have been changed. The new EpiData provides several advantages over the old entry version. Meta-data and records are stored in a single file with extension “.epx”, which abandons the previous triplet system. The file is basically a text file written in a special web-programming language called “eXtensible Markup Language” (XML) which is used to store data using simple text. It has become more graphically oriented. It also supports Unicode (UTF-8) system hence non-Latin texts can be displayed. Moreover, a lot of efforts were also put to implement good clinical practice (GCP) principle required for many medical data projects. This means data encryption, detailed logging of events and user access control of data.

The EpiData Manager is a tool for the project manager. Its role is to define data structures, add meta-data, document, and export data. Files created are also independent of operating system. Once created you can open the file on any computers that install the freeware. The EntryClient serves only data entry. The data entry personnel are not allowed to change rules or structure while doing data entry.

1.2 Installing EpiData Manager and EntryClient

Visit its official homepage, <http://www.epidata.dk/>. To download, the page can be accessed here.

<https://www.epidata.dk/download.php/>

The installers are available for Windows, Mac, and Linux systems. Windows users can download an all-in-one installer which includes Manager, EntryClient, and Analysis in one package. Mac and Linux users can choose either 32-bit or 64-bit operating system and download each installer separately.

When you open the Manager for the first time, check the version online. There are three types of versions by the EpiData: 1) Current version, 2) Public version for stable release, and 3) Test version. Current version is 4.6.0.2 (as of 7th August 2020).

EpiData Software - Download

Home | **About** | Get Help | Download | Documentation | Examples | Links | Credit | Site Map

To work with EpiData software and the EPX project file you must download and install the following:

- To define, document and export data [get EpiData Manager](#)
- To enter data [get EntryClient](#)
- To analyse and do data-management [get EpiData Analysis](#)

In special situations or when a given function is not yet developed in Manager, EntryClient and

Manager, EntryClient & Analysis			
Introduction	Manager	EntryClient	Analysis
Manual:	11.Jul 2014 (3.2 Mb) [171411]	11.Jul 2014 (0.3 Mb) [156682]	08.May 2018 (0.5 Mb) [36024]
Installation packages			
OS:	Manager v4.6.0.2	EntryClient v4.6.0.2	Analysis v3.0.0.1
	Standard 32 bit	Standard 32 bit	Standard 32 bit
Linux:	01.Jul 2020 (4.5 Mb) [4831]	01.Jul 2020 (2 Mb) [4728]	01.Jul 2020 (2.5 Mb) [737]
	Linux 64 bit	Linux 64 bit	Linux 64 bit
	01.Jul 2020 (4.6 Mb) [4575]	01.Jul 2020 (2.1 Mb) [4317]	01.Jul 2020 (2.7 Mb) [849]
Mac OS X:	Standard Intel	Standard Intel	Standard Intel
	01.Jul 2020 (10.3 Mb) [6574]	01.Jul 2020 (6.2 Mb) [5725]	01.Jul 2020 (8.6 Mb) [896]
Windows:	All-in-one Installer		
	01.Jul 2020 (12.9 Mb) [3463]		

Figure 1.2.1 Download Page on EpiData webpage

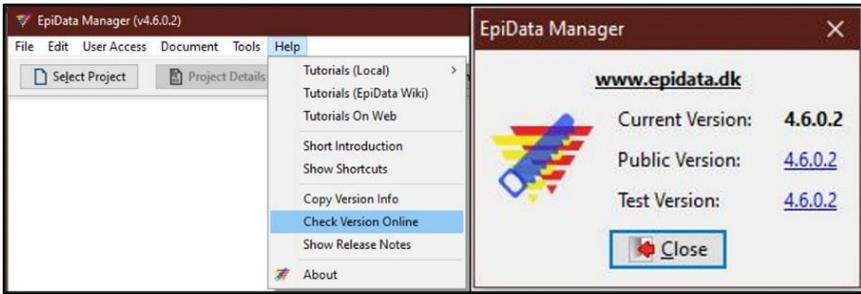


Figure 1.2.2. Check EpiData’s version online

1.3 Terminology

Field refers to variables with certain characteristics such as numeric, decimals, text, or date. During data entry, values will be put into these fields corresponding to their pre-specified data types.

Record refers to the combination of fields or variables in a subject or participant.

Dataset refers to a compilation of such records. In EpiData, it also refers to a dataform which holds a number or such records.

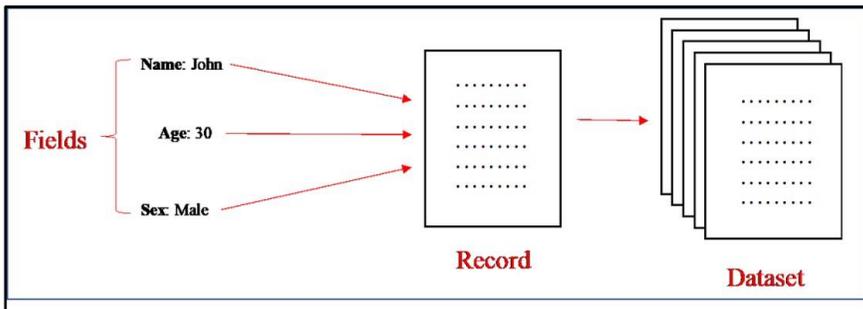


Figure 1.3 Visual representation of field, record, and dataset

1.4 Help and Documentation

There are many ways to seek help.

First, there are introduction manuals and examples from the website, <http://epidata.dk/downloads.php>. Second, you can get help from an online forum called “EpiData-list -- EpiData development and support” which lists number of online subscribers.

<http://lists.umanitoba.ca/mailman/listinfo/epidata-list/>

After your subscription to the list, you can access the forum. The forum is run by administrators and experts as volunteer basis and can only respond to queries in free time. The response may not be instantaneous.

Third, there is another web-archive where you can ask questions and look for the queries and responses that already address your problems. The archive can be accessed here:

<http://lists.umanitoba.ca/pipermail/epidata-list/>

Finally, I would also recommend reading the following manuals.

1. Short Introduction to EpiData Manager Version 2.01 J. Lauritsen & T. Christiansen,
<http://www.epidata.dk/downloads/epidatamanagerintro.pdf>
2. EpiData EntryClient Short Introduction, Documentation and help file. Version 2.0 J. Lauritsen & T. Christiansen,
<http://www.epidata.dk/downloads/epidataentryclientintro.pdf>
3. EpiData Software for Operations Research in Tuberculosis Control: A course developed by the EpiData Association. Hans L. Reider and J. Lauritsen,
<https://tbrieder.org/epidata/epidata.html>

CHAPTER 2 GETTING STARTED WITH EPIDATA MANAGER

EpiData Manager is the first part of the freeware where you create and manage data collection projects. This is intended for data manager so that the final version can be freely distributed to data staffs without worrying them modifying the projects.

2.1 The interface of Manager

One major advantage of the EpiData Manager is its simple and intuitive graphical user interface which include two types of menus: 1) dropdown menu, and 2) button-based menu, also called **work process toolbar**. The toolbar provides a general idea of data collection process in EpiData, including project creation, checking dataset, documentation, data entry and finally data export.



Figure 2.1 Interface of EpiData Manager

2.2 Creating a New Project

To create a new blank project, click “**Select Project**” on toolbar or “**File**” on menu bar, and then choose “**New Project**”. Alternatively, you can use the keyboard shortcut “Ctrl + N” on Windows or “⌘ + N” for Mac OS.

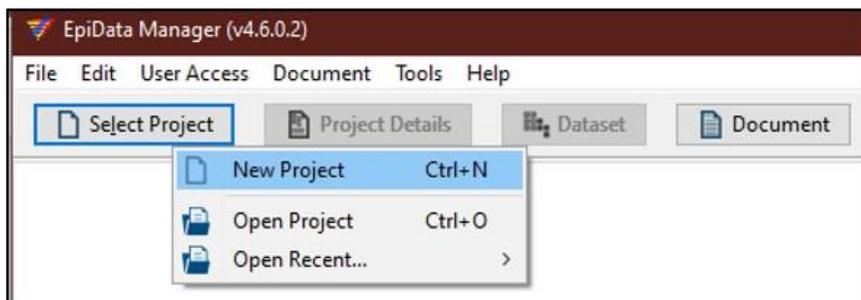


Figure 2.2 Creating a new project from toolbar

2.3 Navigating a new project

When a project is created, the white screen below the toolbar is split into two parts: a small one on the left side, called “**Project Tree**” and the big one on the right side for filling details and creating data entry fields. Under the project tree, we can create data forms or sub-forms as many as we want. There is a “**status bar**” at the bottom.

As you create a new project, you will see **Study Information** page on the big screen on the right side. As the name suggests, you just fill relevant information regarding the study. It is essentially study protocol.

The Welcome tab provides you with a brief instruction on how to create single and relational data forms. Now close this page and let us fill the study information.

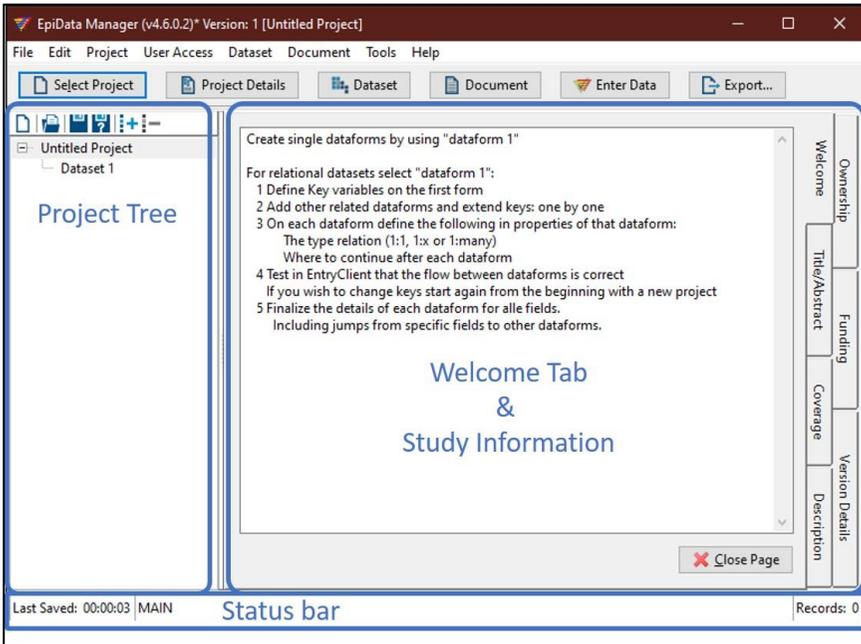


Figure 2.3.1 Project tree, Study Information, and status bar

Project Tree lets you navigate through all data forms under the tree structure. You can easily switch the main window from the project's **Study Information** to other data forms by pointing and clicking there on the tree structure. Above the tree, there are certain tools to manage projects such as creating or opening new project, saving project, or saving project as another file and finally creating or deleting data forms.

The study information is basically *metadata* of EpiData. The concept is adapted from the full set of study information, known as the *Dublin Core Collection*. You can read more about it here. <https://www.dublincore.org/>

There are six categories of metadata, displayed in six tabs.

1. Title/Abstract – title of the study. The abstract should contain a short summary of the project. As soon as you change the title, the name of the project tree will change, but the file name will not be affected unless you save the file.
2. Coverage
 - i. Geographical – indicate study location
 - ii. Language – specify the current language used in the

- project. This is particularly useful as some projects can be conducted through multicenter collaboration.
- iii. Date time coverage - Study period in *dd/mm/yyyy* format.
 - iv. Population – study population, samples and sampling procedures can be described
 - v. Units of observation – study units, tools, and measurements
3. Description
 - i. Keywords – Any specific keywords.
 - ii. Purpose – Specify rationale, research questions, aims & objectives and implications of the study.
 - iii. Citations
 - iv. Design
 4. Ownership
 - i. Organization/Institute – Names of organizations, institutes, and other partnerships
 - ii. Agency (Short acronym ...) – Name of the place where the data is stored, e.g. headquarters
 - iii. Authors/Contributors – Authorships and any acknowledgements of contribution to the project
 - iv. Rights - Any statements of copyrights, disclaimers, and credits.
 5. Funding: Any funding statements.
 6. Version Details: Information about version is particularly important when you update the project as it is implemented. You can use simple incremental method, starting from 1, 2, 3, etc. To be more precise, you can add date on which the version is created, such as 1.07082020 which denotes version 1 created on August 7, 2020.

Status bar at the bottom of the screen is handy at times. It currently provides three pieces of information: 1) **Last Saved**, 2) **MAIN**, and 3) **Records**. “**Last Saved**” shows the time in hours, minutes, and seconds since you last saved the project. “**MAIN**” indicates that the **canvas** is currently selected. This text will be changed if you select a data entry field on a form. “**Records**” shows the number of records in current project. Currently we have no records since this is a new project.

If you click on the data form “**Dataset 1**” under the project tree, the big screen with study information changes to a white screen with small grid layout. We call this as **a canvas**. On this blank canvas, we will add data entry fields and relevant information to create data forms. Above the canvas located a series of design tools to add fields and texts on data forms.

When you maximize the application window, you may notice a red dashed line on the right edge of the blank canvas. This line indicates the margin of the form when you print. So, it is recommended to add fields and texts within the red margin.

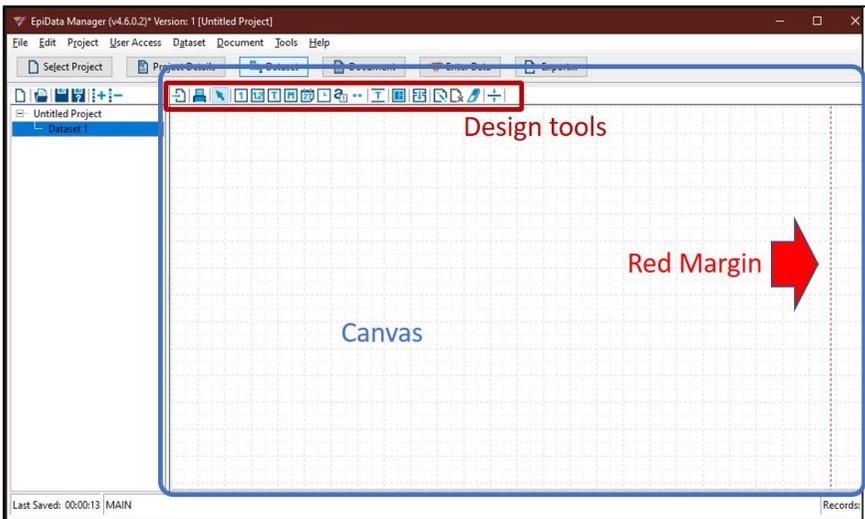


Figure 2.3.2 Blank canvas of the data form

2.4 Change the name of the project tree

There are two ways to change the name of the project tree:

- 1) Select the project tree by clicking on “Untitled Project”, then go to Title/Abstract, type in the name “**Form 1**” and press enter.
- 2) Double click on the project tree, type in the text and press enter.

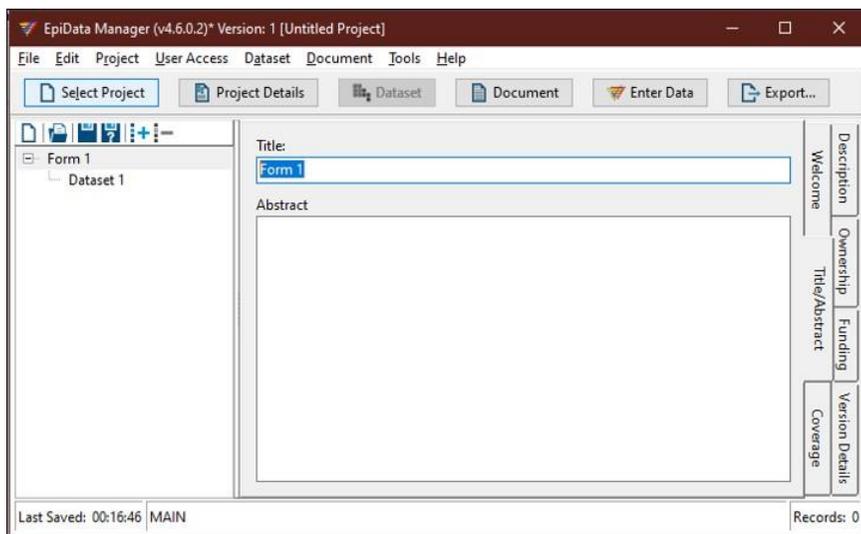


Figure 2.4 Changing the name of the project tree

2.5 Saving and closing current project

To save the project file, click “**File**” and then choose “**Save Project As**”. Navigate to the folder of your choice. Give file name “**form1**” and then click “**Save**”. Alternately, you can use keyboard shortcut “**Shift + Ctrl + S**”.

Always remember to save your project periodically as you might never know when your computer will crash!

After you save the project, the status bar will show “**Cycle No.: 1**”. This means that you have saved current project once. This is the number of times you save your project. “**Cycle No.**” and “**Last Saved**” in the status bar goes hand in hand. Every time the cycle number increases, “**Last Saved**” resets the time.

To close the project, click “**File**” in the menu bar and choose “**Close Project**”. Shortcut key is “**Ctrl + F4**”.

2.6 Opening existing projects

To open an existing project, click “**File**” in the menu bar, choose “**Open Project**”, navigate to the folder and select the file. Alternatively, you can use “**Open Recent**” for any previously opened project. Press “**Ctrl + Shift + 1**” for the most recently opened project.