# How to Create a Research Subject Map

## Introduction

When overseeing a study, one of the primary concerns for a research study coordinator is to protect the confidentiality of their patients by hiding potentially identifiable information such as medical record numbers (MRNs). MRNs act as identifiers for the patients in your local electronic health record (EHR). To hide these identifiers, you must use a research subject id for each patient that can be stored in REDCap and is not protected from the team running the REDI software.

This document will guide you in the manual procedure of creating a map between the MRNs for your patients and their research subject ID, as well as including the dates of interest for your study. This file will feed into the REDI process to help pull the data for your patients into REDCap.

Note that if you do not wish to use this manual procedure and have the ability to work with a Python program, you may use Research Subject Mapper software (<https://github.com/ctsit/research-subject-mapper>). This solution uses a specially-built REDCap project to generate a subject map file.

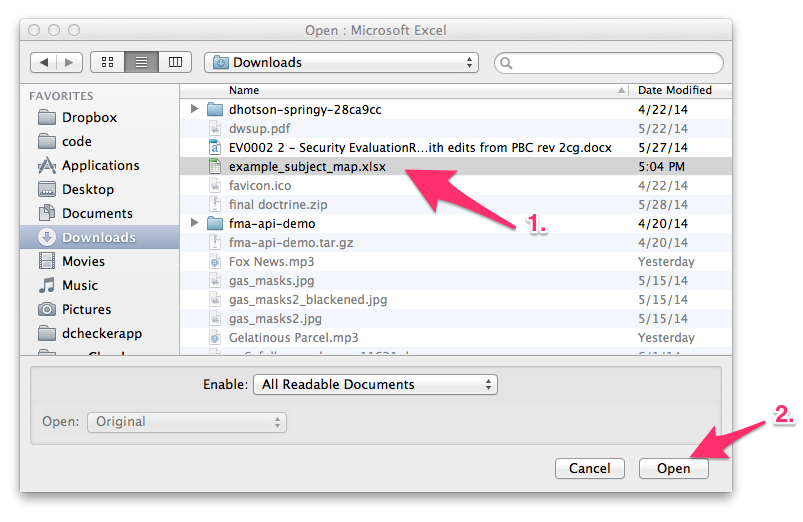
## Requirements

The subject mapping file is created in Microsoft Excel. You must be able to edit a spreadsheet to complete this procedure.

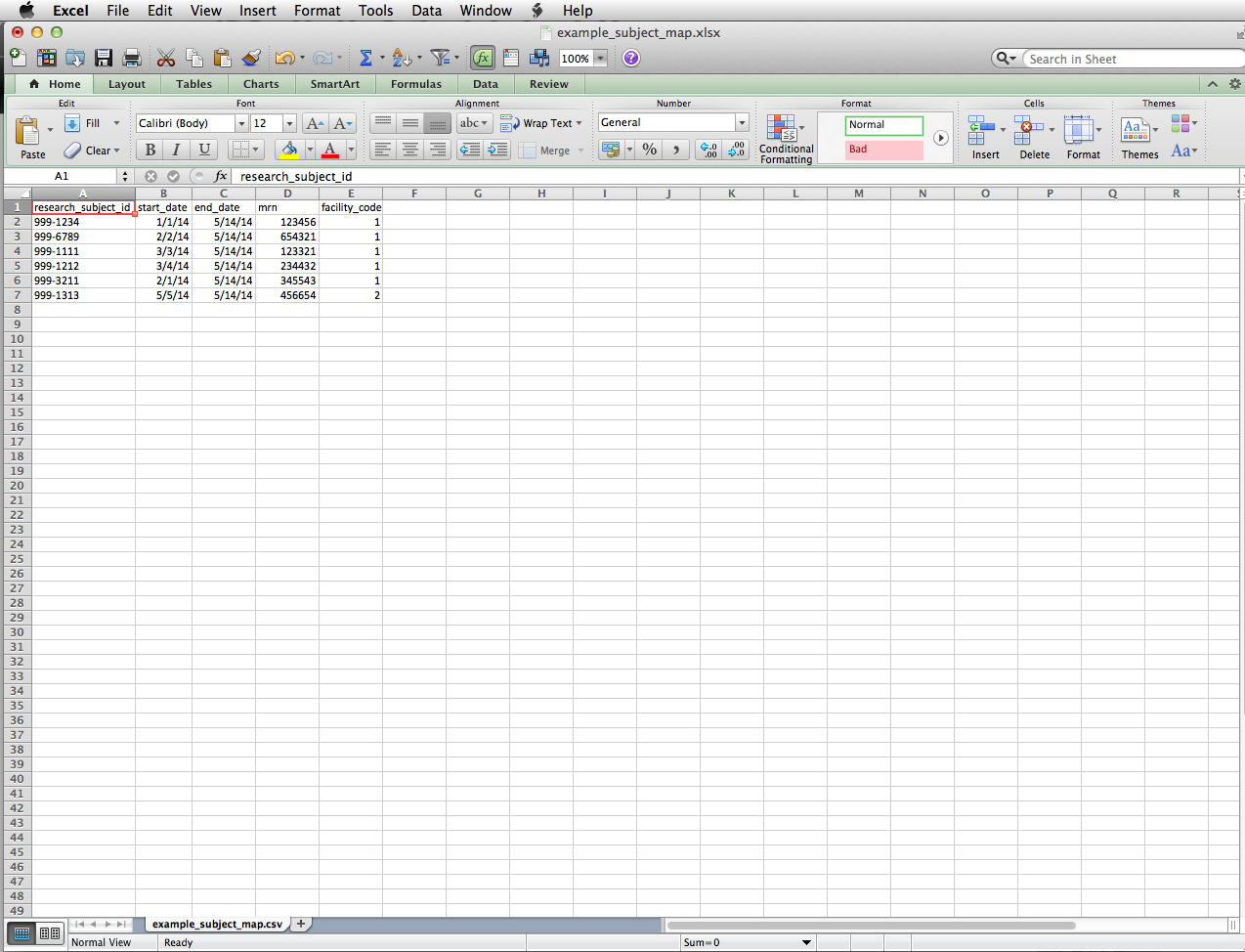
## How To

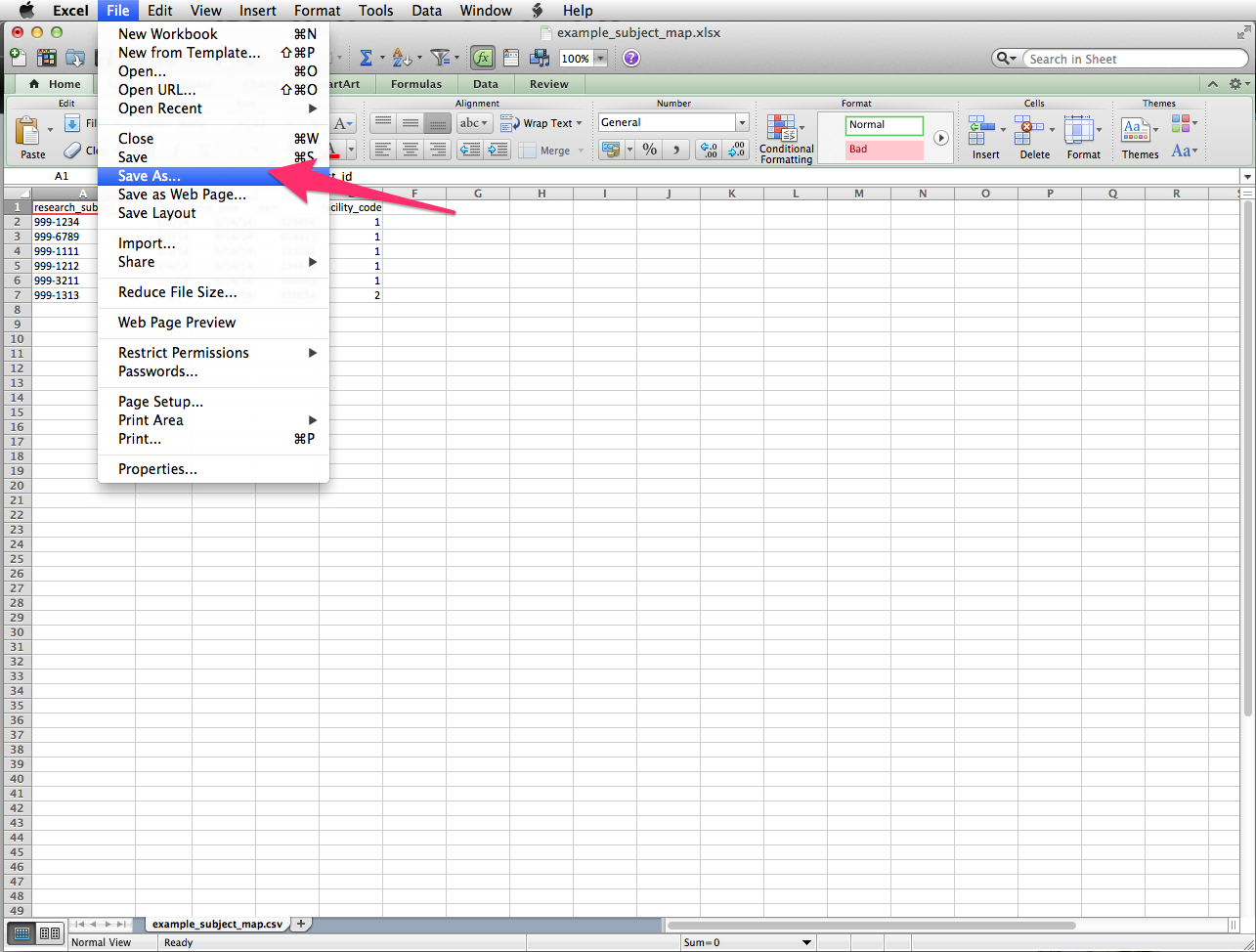
1. Download and open the example research subject map. (<https://github.com/ctsit/research-subject-mapper/raw/develop/doc/manual_research_subject_mapping/example_subject_map.xlsx>)

We will be editing this to create our file.

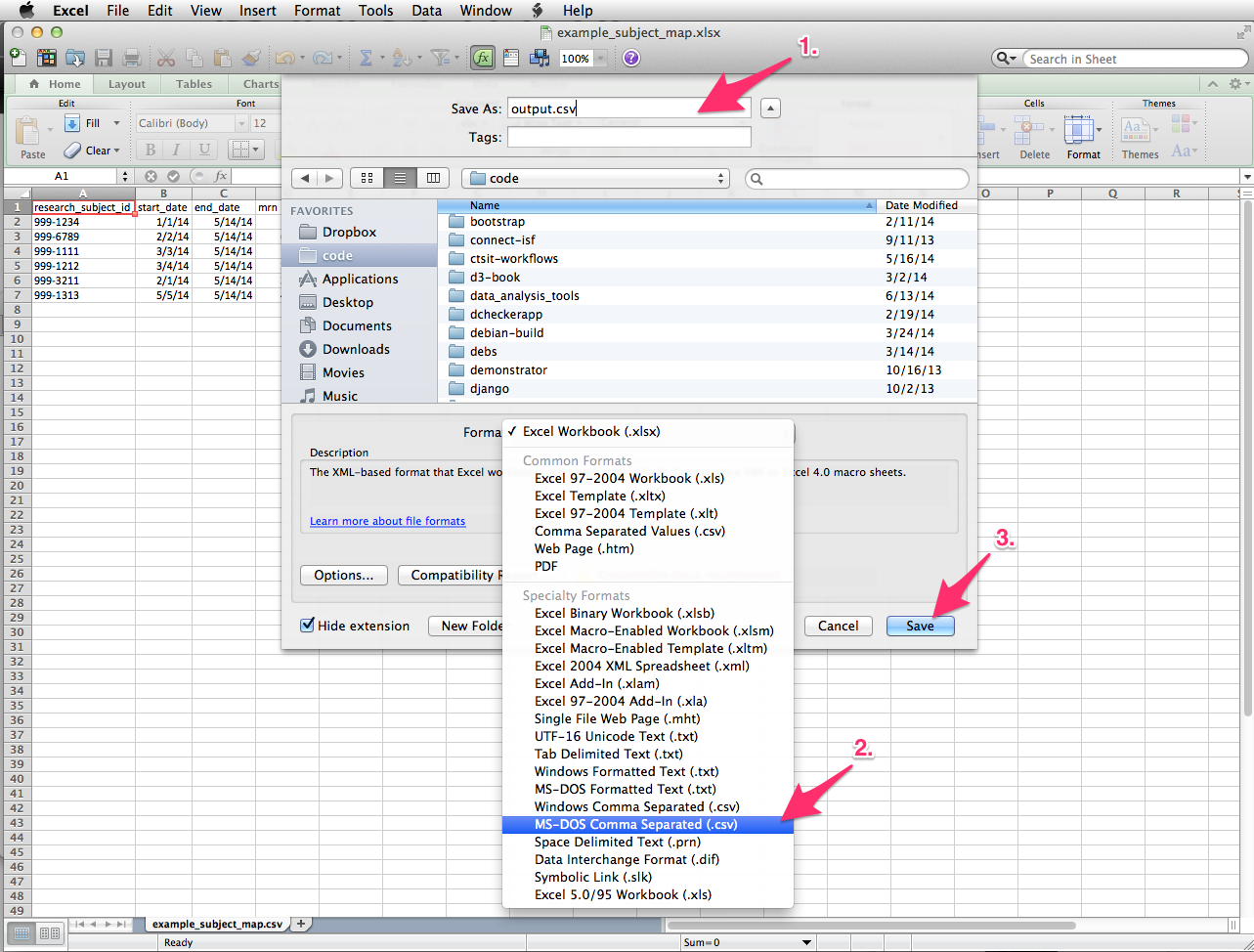


1. After opening the sample subject map file in Excel, you will see the following. There are six rows of sample data, and one header row that labels each column.

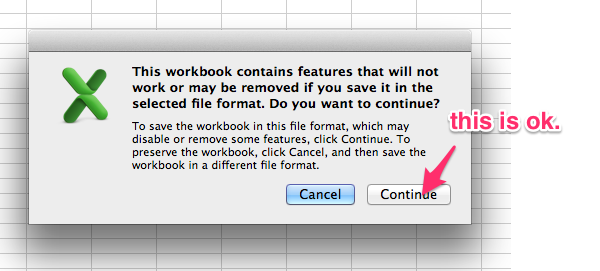


1. Next, select Save As from the File menu. You will need to re-save the file so that you are no longer working on the sample file. You will be re-saving the file in a CSV (comma-separated-value) format, a generic format that can be read by many programs.

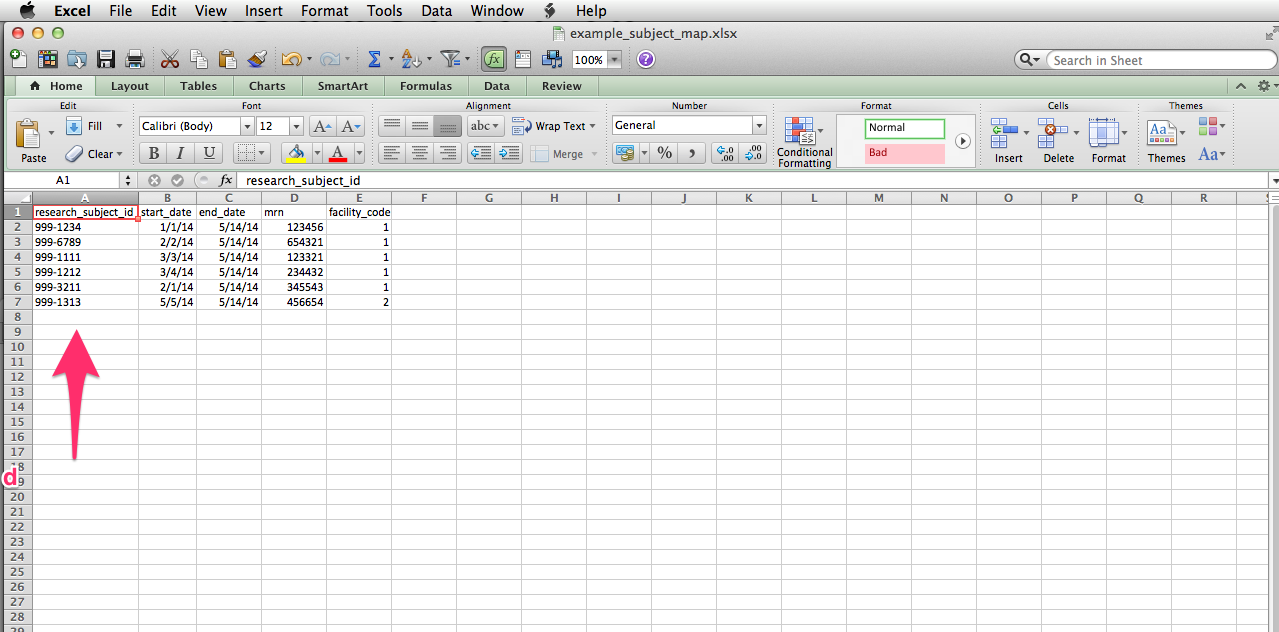
Enter “output.csv” for the file name, then select “MS-DOS Comma Separated (csv)” from the dropdown, and finally click save.



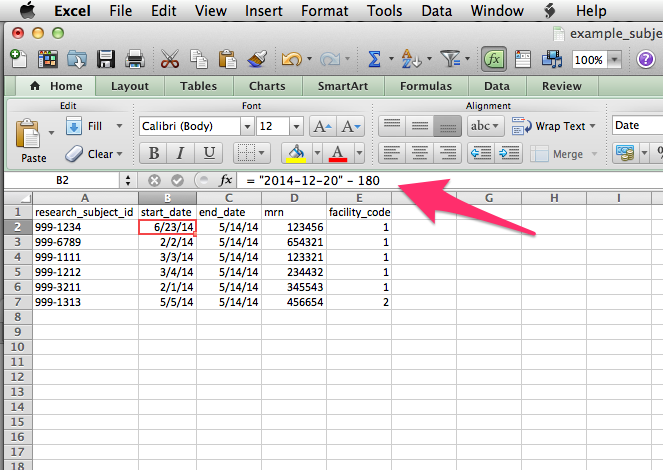
You will see the following alert dialog box. This is OK. Click “Continue”.



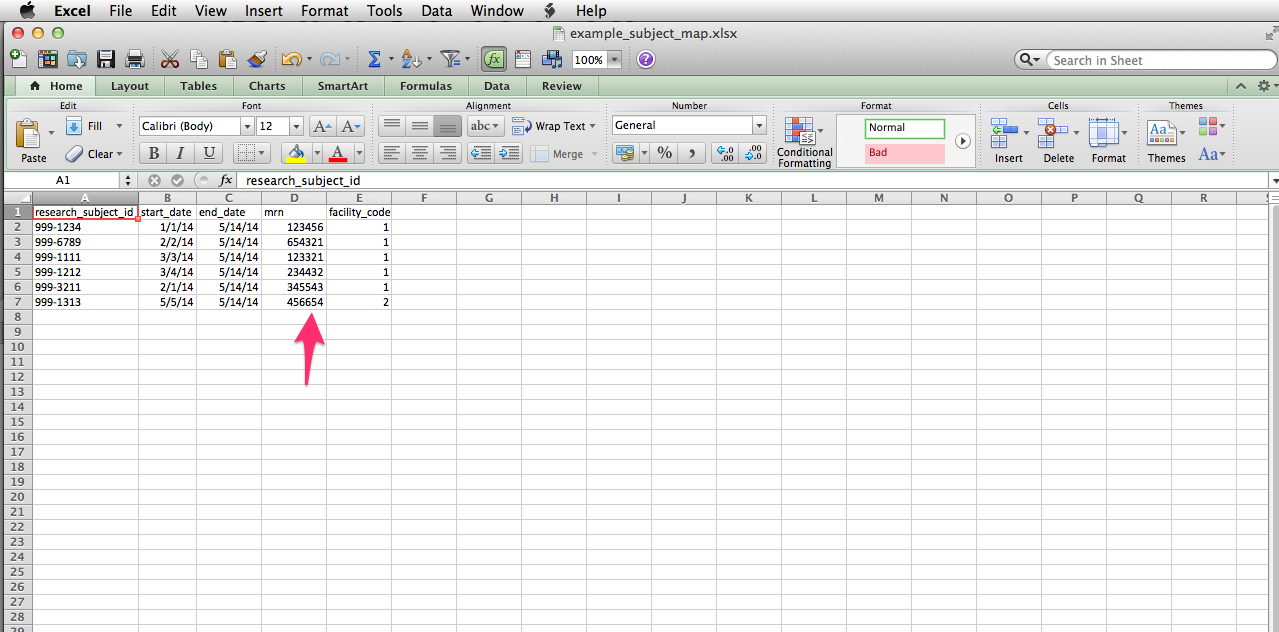
4) Now, you will begin entering data, one row at time. Each row corresponds to one patient. The first column (shown with the arrow pointing to it below) contains the HCV-TARGET research subject IDs for each patient in your study. Erase the sample data by clicking in each box and deleting, and then enter the subject ID for your first patient.



4) Next, you need to enter the corresponding start and end dates for the patients in your study. These are in the columns labeled below as 1 and 2. Start date is the patient’s consent date, minus 180 days.

To have Excel automatically calculate minus 180 days, enter the date as “ = “YYYY-12, 20” -180” in the cell, as shown in the below screen shot. For example, if your consent date is May 5, 2014, enter the date as =”2014-05-05” – 180.

5) Next, enter the patient’s MRN. These are the identifiers that must be protected at your site.



6) Finally, you may have different medical facilities that feed in patients to your EMR, with different facility codes for each. If these exist in your system, enter them in the final column. If they don’t exist, you can leave the column blank.

7) After filling out each column for the patient, you have finished the data entry for that patient. Return to step 4, and repeat filling out each column for each patient in your study. When you have finished with all the patients, your data entry is complete. Save the file again and you are finished.