# How to Create a Research Subject Map File for RED-I

# An instructional document

## Purpose

This document describes the need for creating a research subject map to use with the RED-I software project, and details a method for creating this manually. It is intended to be used along with the example\_subject\_map.csv file.

## Need for a research subject map

When transforming and loading patient data from the electronic health record (EHR) to a REDCap system it is important to protect the patient’s confidentiality by hiding potentially identifiable information from the RED-I software and anyone running it. MRN (medical record numbers) uniquely identify patients and therefore must be obscured by an intermediary process.

## Software to map automatically

Research subject mapper (<https://github.com/ctsit/research-subject-mapper>) software has been developed by CTS-IT that can automatically create the mapping file. It uses a REDCap project that site coordinators can enter MRNs, date of birth, and research subject IDs which the software then uses to create a mapping file for use by the local EMR data curators. This obscures the MRNs from the team running RED-I to protect patients.

## Manual mapping file creation

Site coordinators may instead choose to manually create a mapping file, using the example csv file attached. This method is useful in the event there are a small amount of patients with data that will be gathered, or if the coordinator prefers not to use Python software.

## Formatting of mapping file

See the example CSV file for a demonstration of what data to enter. It is recommended to do this in a text editor such as TextEdit on Mac or Notepad on Windows, instead of using Excel or Word to prevent formatting problems. Formatting of the subject mapping file must be followed strictly to generate a conforming file. Use these guidelines:

1. All fields must be surrounded by double quotes
2. All fields must be separated by commas, except at the end of a line.

In addition, use these guidelines for entering the data:

1. “research\_subject\_id” is the subject number for the patient (i.e., for the HCV Target project, it is the Target subject number)
2. “start\_date” and “end\_date” must be formatted YYYY-MM-DD. For the HCV Target project, the start date is the consent date minus 180 days. This can be calculated easily in Excel by entering the date in this format:

=”YYYY-MM-DD”-180

where YYYY-MM-DD substitutes for a valid date.

You can also use an online calculator like <http://www.calculatorsoup.com/calculators/time/date-day.php>.

1. “mrn” is the medical record number in the local EHR.
2. “facility\_code” is an optional code that corresponds to a hospital in the EHR. This may be required for institutions that have patients at multiple hospitals who may be assigned different MRNs. In this case, using facility\_code prevents erroneously mismatching patients with the same MRNs.