

De-Identify Data

Back to the [Data Export Overview](#)



This feature requires a **CommCare Software Plan**

This feature (De-identified Data Exports) is only available to CommCare users with a **Pro Plan** or higher. For more details, see the [CommCare Software Plan page](#).

De-Identify Data

This export allows a user with secure access to CommCare HQ to download data that can be analyzed for specific outcomes, but makes the personal identity of the cases unknown to the data analyst. CommCare applications may contain both personal identifying information and sensitive information. This type of information includes:

- the individual's past, present, or future physical or mental health or condition
- the individual's financial background, ownership information
- the individual's sexual history

By combining sensitive information with common identifiers such as date of birth, geographic location, or sex, the information could be used to identify a single person. On the data collection side, CommCare requires all users to log in with a secure, unique identification. The data is then securely hosted and is encrypted using RSA 256-bit encryption. All interactions on the CommCareHQ website are conducted using industry standard transmission encryption. CommCareHQ reports are only made available to users with appropriate access to public health information. However, it is the responsibility of users with access to data in their project spaces to make sure that it is shared appropriately.

CommCare's De-identification Function

- Sensitive IDs - any field marked as a sensitive ID will be replaced with a random alphanumeric code. This code will be consistent within forms; that is if you are treating owner_id as a sensitive field and owner_id is the same in 10 form submissions, then it will be replaced with the same code in all of the form submissions.
- Sensitive dates - dates are shifted by up to one month, randomly. However the length of the shift is consistent within a given form or case. So if in one form you ask both mother's date of birth and child's date of birth, both dates will be shifted by the same number.

Downloading de-identified report is a three-step process:

1. Select a Form Export

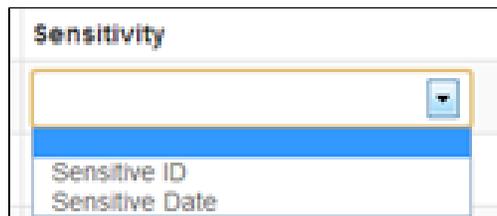
Create or select a form export: Information to create a form export is located [here](#). To select an existing form export go to:

CommCareHQ -> Data -> Export Data -> Export Forms -> Exports and select "Edit" for the export you want to download with de-identified data. Scroll to the bottom of the page to the Privacy Settings.

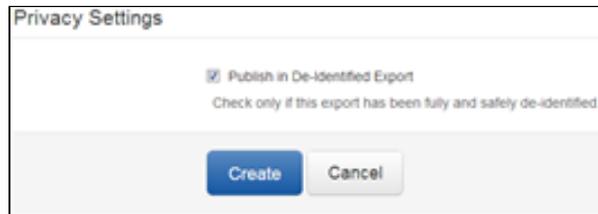
2. Configure Privacy Settings

This allows the user to select form data to be de-identified so when the data is exported to an excel sheet, the columns will still be in the data export but the data values will not contain personal information that can be tracked to a single beneficiary

1. Click "Allow me to mark sensitive data" and another column called "Sensitivity" will be added to the Form table.
2. A drop down box will appear next to each field name. A field can be marked as "Sensitive ID" which can be used for all text or numeric fields such as name or age. Alternatively, a field can be marked as "Sensitive Date" which would be used for date of birth. Finally, a field can be left blank and the data will export directly as it was input into the application



3. Once you have marked the sensitive fields, scroll all the way down to Privacy Settings and check the box "Publish in De-identified Export". Checking this box will make the export appear as a "De-identified Export" on the form export page. By checking this box, you are confirming that you have excluded or marked as sensitive all identifiable information, and users who only have access to de-identified data may access this export.



Privacy Settings

Publish in De-identified Export
Check only if this export has been fully and safely de-identified

Create Cancel

3. Download Your De-Identified Reports

- a. Go to: CommCareHQ -> Data -> Export Data -> De-Identified Export
- b. All the reports that have been published to De-Identified Reports will appear here to download
- c. All fields that have been marked as a sensitive ID will now be a de-identified ten digit number such as: 8K6Q5G4LCI
- d. All fields that have been marked as a sensitive Date will now be a new date -31 to 32 days from the actual date (within individual forms, all dates are shifted by the same amount).