Help - Exporting and publishing data - Exporting data with SurveyCTO Sync

Below, please find a series of topics relating to how you can use *SurveyCTO Sync* to export and process your incoming data.

Installing and using SurveyCTO Sync

SurveyCTO Sync is software you can use to download, process, and export data collected with SurveyCTO. Get started by downloading and installing the appropriate version for your computer's operating system:

- SurveyCTO Sync for Windows
- SurveyCTO Sync for Mac (OSX)
- SurveyCTO Sync for other platforms

The first time you run Sync, any or all of the following may happen: (a) it may need to install Oracle's Java environment (on which it relies); (b) it may need to upgrade your Java environment in order to support strong data security (if so, it will walk you through the process); and (c) it will ask you where on your local hard drive to store downloaded form data.

(If you are running Windows and have trouble upgrading your Java encryption support, you can close Sync, right-click on its .exe file, and choose *Run as Administrator*. Upgrading generally works fine as long as your current login has sufficient permissions to access the Java system files. If your organization does not allow you permission to upgrade your own system, tell your IT department that you need them to install the "Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files.")

Once Sync is installed, you can use it to download, transport, and export data. Simply choose the source of the data (typically "Server", in which case you enter **survey.wb** as the server name, plus a valid username and password), a destination (typically "Local CSVs", in which case you select a local directory into which exported data should be saved), and then press *GO!*. Next, choose the forms for which you want to download, transport, and/or export data. Finally, press the second *GO!* button. Data will be transported and/or exported to the specified location.

Form data is exported in .csv format by default, then Sync offers a range of options for exporting additional files in other formats as well. For more about the default .csv format, see *Understanding the format of exported data*. (You should review that help topic even if you plan to export in some other format, as the overall structure of the data is discussed there.)

Note that exported form data is cumulative in that it includes all data stored on the server and all data stored locally by Sync. However, this is not something upon which you should rely. Old data may eventually be purged from both the server and Sync, so exported .csv files should not be relied upon to always include all old data.

It is best to think of each export as containing all new data as well as some possibly-incomplete subset of old data. Therefore, your data-processing tasks should include:

1. Merging newly-exported data into a master dataset (probably ignoring all old data in the .csv exports).
2. Maintaining backups of both your master dataset as well as all raw .csv files exported by Sync. While it is redundant to back up both the raw .csv files and the master dataset, it is better safe than sorry. You should also rotate your backups so that some backups are maintained off-site, in case of theft, fire, or other office mishap.

You can configure Sync to execute processes that partially or completely automate both of these tasks. You can also export subsets of your data to Google Earth or to other .csv files; mail-merge data with Microsoft Word templates; or track the status of your back-office-data processing. See the following help topics for details.

Within Sync's preferences, you will also find various settings to control how your data is stored and exported. On Windows, this will be an option under the *Tools* menu in the upper-left of the *SurveyCTO Sync* window; on OSX, it will be an option under the *SurveyCTO Sync* menu in the upper-left of your screen.

Data export options

SurveyCTO Sync offers a wide range of options to control how data is processed and exported. You can find all of these options within the preferences (on Windows, find *Preferences* under the *Tools* menu in the upper-left of the Sync window; on OSX, find it under the *SurveyCTO Sync* menu in the upper-left of your screen). There, preference options are organized in a series of different tabs. Some of the less obvious options are discussed below.

Treatment of enclosing groups in exports

Internally, SurveyCTO stores fields according to a "full name" that includes the field's name as well as the names of all enclosing groups (as in "module1/field1"). One implication of this is that a field named "field1" in one group is treated as effectively different from another field named "field1" that is located in another group. In part as a reflection of this, Sync's default behavior is to include group names in the column headers of exported .csv files (as in "group1-field1" or even "outergroup-innergroup-field1").

Since most people almost always give unique names to all of the fields in their forms, regardless of which groups they're in, including the group names in the export headers is often unnecessary – and unwanted, because it makes the headers longer and more unwieldy. So, there are two options to exclude those group names from export headers:

1. *Exclude groups, but still use them internally*. This option simply drops the group names from the export headers, but it still considers fields in different groups to be different fields. So, for example, if you did have a field named "field1" in two different groups, you would end up with two "field1" columns in your exports (because Sync would be exporting them as different fields, even though they share the same name).
2. *Ignore groups so fields with same name export together (slowest)*. This option merges data for different fields (in different groups) that share the same name, so that there will never be more than one export column for a single field name. This option slows down the export process, but it can be a life-saver if you have edited your form over time and moved a field from one group to another: Sync will export old and new data for a field even if it has changed groups. (See *Updating an existing form* for more.)

Note that, because Stata and mail-merge templates reference field names as they appear in your .csv headers, you will need to re-generate any Stata .do file templates or Microsoft Word mail-merge templates after changing this option.

Exclude note fields from exports

Forms typically include a lot of *note* fields, which display something to the user but don't actually record any data. For completeness, Sync includes these fields in exports by default – but their columns are always blank. You can use this option to simply exclude those fields from your exports.

Field value to include for missing values

By default, "missing" values are exported as blanks. This includes all cases where a field was never displayed and so the user could not have given a response (the field might not have been *relevant*, or the field might not have existed in the version of the form filled out by the user). To distinguish these missing cases from cases where the field did display but the user simply left it blank, you can ask Sync to replace all missing values with some character or string. For example, if you wanted to follow the standard set by Stata, you could configure Sync to replace missing values with a single period (".").

Show test forms

By default, Sync won't show forms with titles that begin with "TEST - ", but you can override that by enabling this option. See *Hiding forms during testing* for more.

Apply corrections

By default, Sync will automatically apply all corrections made to data using the review and correction workflow. To export the originally-collected data, without any corrections, un-check the *Apply corrections* option.

Override the default value separator in .csv files

By default, SurveyCTO separates columns in .csv files using commas (after all, CSV stands for "comma-separated values"). Depending on your region, however, your system may prefer to use semi-colons (;'s) or some other character to separate columns in .csv files.

You can override this default and specify some other separator to use. For most cases, you will want to use either a comma (,) or a semi-colon (;).

Replace line-breaks in exported data

Depending on what software you use to open .csv files exported by Sync, line-breaks within field values might cause confusion. For example, Microsoft Excel can get confused and start a new row when a field includes a line-break. Therefore, you can have Sync auto-replace all line-breaks with some other character or string of your choosing. (A single space character is the most popular replacement for line-breaks.)

Export repeat groups in "wide" format

Sync defaults to exporting data from repeat groups into a series of separate long-format .csv files. However, if you wish, you can choose this option to also export a single wide-format .csv file that includes extra columns to accommodate repeated data (the long-format .csv files will still be exported as well). See *Understanding the format of exported data* for more on this issue of long vs. wide export formats.

Export select_multiple responses as series of 1/0 columns

By default, Sync will export each *select_multiple* field as a single column that includes a space-separated list of selected choice values (e.g., "1 2 5" or "Boston NYC"). If you prefer a series of 1/0 "dummy" variables, Sync can also export an additional 1/0 column for each possible selection (though it will slow down the export process somewhat). See *Understanding the format of exported data* for more on the structure of exported data.

Looking at data in Microsoft Excel

The default approach to exporting your data with *SurveyCTO Sync* is to export it to "Local CSVs", in which case one or more .csv files will be exported into the destination directory of your choosing. (The default behavior is to export all data in "long format", in which case data from repeat groups is exported into separate .csv files. You can also export a unified wide-format .csv file that includes all data, repeated or not. See *Data export options* above for details.)

The first thing many people do is open their data in Microsoft Excel. You can certainly review, clean, or even analyze your data in Excel. However, please note the following:

1. If you double-click on a .csv file, most systems will default to opening the file in Excel. In this case, Excel will import the .csv file based on a series of assumptions, most of which will not alter your data. But it will assume that the characters in the file are Windows or Mac characters as opposed to Unicode characters, so it will mess up accents or foreign scripts.
2. The safer way to import a .csv file is to first open Excel and then choose either *File...Import* or *Data...Get External Data...Import Text File*. You will want to specify that it is a .csv file, that it is comma-delimited, and that its "file origin" or "encoding" is Unicode/UTF-8. You will then get the opportunity to preview the columns and override the cell formats used for each column. Finally, when it asks you where to put the data, there will be a *Properties...* button that you can click to see options for "refresh control": whether to re-import the data whenever you open the Excel workbook and whether to re-prompt for the source .csv file every time. Once you import this way, you can essentially re-import at any time by clicking *Refresh* on the *Data* tab.
3. Sadly, when you import data using Excel's import wizard, it gets confused when it sees line-breaks inside cells – which will be a problem for you if anybody ever presses Enter when entering a text response into one of your survey forms. Excel will end the row at the line-break and essentially break one row of data into multiple rows. To avoid this, there is an option in Sync's preferences to replace

line-breaks in exported .csv files with some other character; the default is to replace all line-breaks with spaces.

4. Excel will sometimes convert things that aren't dates into dates. For example, say you have an ID number like "10-11-12": Excel will convert it into a date, encoding it internally (and saving it!) as the number of days between Jan. 1, 1980 and Oct. 11, 2012. Obviously, that is totally wrong. To prevent Excel from corrupting your data in this way, you need to take care when assigning the formatting of your columns on import: for ID numbers and other things that might be mistaken for dates, "General" is not a safe choice.
5. If you want to edit your data or make notes in new columns, you are better off configuring Sync to directly merge incoming data into an existing Excel workbook, rather than exporting to .csv format and then re-importing into Excel yourself. See this help topic for more on merging directly into Excel.

Using Google Earth

Whenever you export data to local .csv files, you can choose to also export a subset of that data to Google Earth (click here to learn more about Google Earth).

Start by beginning the *SurveyCTO Sync* data-export process the way you usually do: press the first *GO!* button and select the forms that you would like to export (including any forms that contain data that you would like to export to Google Earth). Before clicking the second *GO!* button, however, check the *Show export options after download* checkbox down at the bottom. Then, click the second *GO!* button, and when the export options window pops up, check the *Export Google Earth .kml files for some forms* option, click *Manage .kml exports...*, and select the forms for which you would like to view data in Google Earth. (Note that only forms containing at least one *geopoint* field will be listed since Google Earth requires GPS locations for all data points.)

For each form that you select, you will be prompted to configure some export options. First, specify which field to use for the GPS location. Next, specify which field to use when labeling your data points in Google Earth. Finally, choose which subset of fields to include in Google Earth; the data for these fields will appear when you click on a data point.

By default, all chosen data for a given form exports to a single .kml file. However, you can also organize your exported data into multiple .kml files by choosing the fields by which you want exports organized. For example, if you choose to organize first by district, then by block, and then by village, your .kml files will be named "Form title - district - block - village.kml", and each file will only include the appropriate subset of data. This makes it easy to, for example, monitor survey operations in different areas or for different teams.

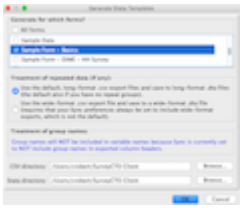
Once configured, your chosen subset of data will export to .kml file(s) whenever new .csv data is exported (to the same .csv export location).

Using Stata

SurveyCTO Sync can automatically generate Stata .do files that import, merge, and partially process your exported data. You can use these auto-generated .do files as they are, or you can use them as a starting-point for your own back-end processing code.

Start by beginning the *SurveyCTO Sync* data-export process the way you usually do: press the first *GO!* button and select the forms that you would like to export (including any forms for which you would like to generate Stata templates). Before clicking the second *GO!* button, check the *Show export options after download* checkbox down at the bottom. Then click the second *GO!* button, and when the export options window pops up, click the *Generate Stata templates...* button.

Next, you will be given a list of all survey forms that have been downloaded from your server. Check off those forms for which you would like to generate Stata templates. Also specify your .csv export location, the location to which you would like to output your Stata files, and whether you're using long- or wide-format exports. (The Stata template will follow your export format. See the help topic on export formats for more.)



Once you press *OK* to continue, Sync will output your Stata templates. If any of your selected forms include multiple languages, Sync will prompt you to choose a language for each. Stata variables will then be labeled according to the languages you select.

For most forms, Sync will output two files: import_FORMID.do in your chosen Stata directory, and FORMTITLE_corrections.csv in your .csv export location (where FORMID and FORMTITLE are replaced by the form's ID and title, respectively). If your form has repeat groups and you chose the long export format, then additional .do files will also be output and called from the main import_FORMID.do file.

Once you have exported form data into your chosen .csv export location, you can run any auto-generated template as-is. You can also copy auto-generated .do files into the *thenrun* subdirectory of your .csv export directory, to have them run automatically every time new data is exported (see *How do I automatically execute processes when new data becomes available?*).

For more details on Stata .do file templates, see the main help topic on the subject, which includes additional information on what the .do file templates do, how to use the corrections file, etc.

Sharing subsets of data

Whenever you use *SurveyCTO Sync* to export data to local .csv files, you can choose to also export a subset of that data to other .csv files. You might do this, for example, if you wanted to share a subset of non-PII (non-personally identifiable information) fields with another person, program, or website.

Start by beginning the *SurveyCTO Sync* data-export process the way you usually do: press the first *GO!* button and select the forms that you would like to export (including any forms for which you would like to export subsets of data). Before clicking the second *GO!* button, check the *Show export options after download* checkbox down at the bottom. Then, click the second *GO!* button, and when the export options window pops up: check the *Export subsets of .csv data for some forms* option, click *Manage .csv subsets...*, and select the forms for which you would like to export subsets of data.

For each form that you select, you will be prompted to choose which subset of fields you would like to export.

By default, all chosen data for a given form exports to a single .csv file. However, you can also organize your exported data into multiple .csv files. Simply choose the fields by which you want exports organized. For example, if you choose to organize first by district, then by block, and then by village, then your .csv files will be named like "Form title - SUBSET - district - block - village.csv", and each file will only include the appropriate subset of data. This makes it easy to, for example, monitor survey operations in different areas or for different teams.

Once configured, your chosen subset of data will export to .csv file(s) whenever new .csv data is exported (to the same .csv export location). Primary .csv export files will contain all data, and subset .csv export files (named with "SUBSET" in their file names) will contain only the fields you selected.

Using Microsoft Word and mail merge to output and view data

By default, *SurveyCTO Sync* exports data as .csv or comma-separated-values format. This is ideal for opening the data in Excel, Stata, or any other program that treats data as columns and rows. Unfortunately, this format may not be ideal for users who wish to review the data. For example, you may have a data scrutiny process that partly involves people who manually look over incoming data (to catch mistakes that can't be easily caught by automated computer review). For human review, .csv format is not ideal.

The solution is to use Microsoft Word's "mail merge" feature to merge the computer-readable data with a more user-friendly template. SurveyCTO assists in two ways:

1. You can download a starter template for any existing form. On the Design tab of your server console, select *Download* and then *Mail merge template* for any form in the *Your forms and datasets* section, then specify how you want to label your fields and whether you have *SurveyCTO Sync* configured to include group names in .csv file headers. You can then customize this template or use it as-is.
2. Sync will auto-merge exported data for you. Go into Sync's preferences, click the *Export options* tab, and check the *Auto-run mail merge after export* checkbox. Then, copy the mail-merge template you

downloaded from the server into the destination directory into which you export your .csv files, and name it *formid_template.docx* (where formid is the unique ID of the form, like "sampleform"). Whenever a new .csv file is exported for that form, a *formid_merged.docx* file will also be output; that file will contain the combination of the latest .csv data with the mail merge template.

Note that fields within repeat groups (i.e., fields enclosed by *begin repeat* and *end repeat* rows) are not included by default in the mail-merge output. This is because such fields are exported into separate .csv files, and the mail-merge process only works with a single, primary .csv file. To manually add repeat-group data to a mail merge template, you would need to do the following:

1. Enable the Sync preference to export repeated fields in "wide" format.
2. Open the template and manually merge it with the *(title)_WIDE.csv* file now exported by Sync.
3. Add any of the repeated-field columns from the *(title)_WIDE.csv* file into your template to merge the data from those columns.
4. Re-merge with the updated *(title)_WIDE.csv* file whenever you want to update the mail-merge output with new data.

Also note that mail-merging large amounts of data will require large amounts of memory. If you receive an error about "Java heap space," then it means that you ran out of memory. By default, *SurveyCTO Sync* will use up to 1GB of memory; to use up to 4GB, you can download SurveyCTO_Sync.l4j.ini into the same directory from which you run *SurveyCTO Sync*, then close and re-run Sync. (The name of the .ini file must match the name of your *SurveyCTO Sync* application file. So, e.g., if your application is named "SurveyCTO Sync.app", change the underscore to a space so that the .ini file is named "SurveyCTO Sync.l4j.ini".)

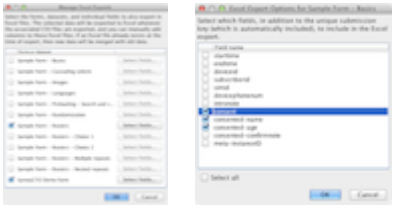
Using Microsoft Excel for back-office operations

Your back-office work is likely far from over once you have exported your data, imported it into a data-analysis package, and/or mail-merged it into Word documents. Chances are, you have teams scrutinizing and correcting data, performing back-checks, or otherwise working to assure the quality of your data. Managing that process can be a challenge, particularly when new data is continually flowing in from the field.

To help make your job easier, SurveyCTO can automatically export form or dataset data to Excel documents, merging exported data with other back-office details that you record in that spreadsheet. Here's how it works:

1. Start by beginning the *SurveyCTO Sync* data-export process the way you usually do: press the first *GO!* button and select the forms and/or datasets that you would like to export (including any for which you would like to export to Excel). Before clicking the second *GO!* button, check the *Show export options after download* checkbox down at the bottom. Then, click the second *GO!* button, and when the export options window pops up, check the *Export and merge to Excel files for some forms and datasets* checkbox.

2. Next, click the *Manage Excel exports...* button and check off any forms and/or datasets for which you would like Excel worksheets to be maintained. For each one that you check, click *Select fields...* to choose which fields you would like included in Excel. (For datasets, you will also need to specify a dataset field to use as the unique identifier. This is because each exported row must have a column that serves to uniquely identify it; for forms, this is automatic because each form submission has a unique key – but for datasets you have to specify an identifier column.)



3. Once the Excel export has been enabled for a form or dataset, SurveyCTO will maintain a *TITLE_status.xls* file for that form or dataset, located in the .csv output directory (where *TITLE* is the title of your form or dataset). Whenever new data is exported, SurveyCTO will add new rows to that spreadsheet. Each row will have columns filled in with the unique row identifier as well as the data associated with any fields you checked when configuring in Step 2 above.
4. You can edit the exported spreadsheet and add new columns to contain whatever additional, back-office details you wish.
5. Whenever you export new data, SurveyCTO will add rows to the spreadsheet and/or update data in the columns being exported; it will never delete rows and it will never change anything entered into any of the new columns that you added yourself. In those additional columns you can maintain whatever other details are useful for managing your back-office operations.

You have many options for how you manage your back-office operations, so you may not need these Excel exports. However, they are available if you need them.

Automatically executing outside processes

After *SurveyCTO Sync* has downloaded and exported new data, you can have it automatically execute outside processes (e.g., a Stata .do file to process and review the new data).

To do this, go into Sync's preferences, click the *Export options* tab, and check off the *Auto-run other processors after export* checkbox.

Then, within the destination directory into which you will export your .csv files, simply create a *thenrun* subdirectory (named "thenrun" exactly) that contains any .cmd, .do, or other executable files that you wish to run. Sync will automatically execute each of these files after exporting new .csv files to that destination. (When there are multiple files to execute, it will execute them in alphabetical order.)

If you will be analyzing your data in Stata, see *How do I create templates for importing into Stata?* for how to get started with auto-generated .do files.

Running SurveyCTO Sync on a schedule

You might want to schedule *SurveyCTO Sync* to download and export new data automatically, in the background. To do so, run Sync and configure your options like normal, press *GO!*, and select which forms to download/export. Then, rather than pressing the second *GO!* button to execute once, choose *Schedule sync...* from the *Tools* menu. The download/export will happen right away, then again every *x* minutes, hours, or days, depending on the options you specify. You can leave Sync running in the background, then close it whenever you want it to stop.

Working with cold-room computers

Typically, you use *SurveyCTO Sync* to sync from *Server* to *Local CSVs* in order to download the latest data from the server and export it to .csv format. However, there are other possible uses.

When dealing with highly sensitive encrypted data, for example, you may want to decrypt and export data only on a non-Internet-connected "cold-room" computer. In this scenario, you would sync from *Server* to *External drive* on an Internet-connected computer, switch the external drive to your cold-room computer, then sync from *External drive* to *Local CSVs* on the cold-room computer. That way, you maintain your private key and decrypted data only on the cold-room computer, where they are safe.

Note that whenever you sync from *Server* or from *External drive*, all new data is always stored in the local Sync storage directory. This is true even when you select a destination other than *Local*. For example, a sync from *Server* to *Local CSVs* actually goes *Server -> Local -> Local CSVs*. At any time, then, you can choose to sync from your local storage (*Local*) to some destination, or sync from somewhere to your local storage. You might do this, for example, if you wanted to re-export .csv files without downloading any new data from the server.

And don't worry: your encrypted data is stored safely (as still encrypted) in Sync's local storage. It is only decrypted when you choose to export it to *Local CSVs*, using the private encryption key that you specify.