Collecting PPI® data through TaroWorks™

A BEST PRACTICES GUIDE FOR IMPROVED DATA VALIDATION

This guide shares best practices for collecting data, specifically PPI data, through TaroWorks, a mobile technology platform that helps organizations to manage field operations and collect data. It is broken into the following sections: Survey Design, Enumerators, and Auditing.

To learn more about TaroWorks, please visit www.TaroWorks.org.

Survey Design

General tips when creating additional questions for PPI surveys

When collecting PPI data, it's common to collect additional data about clients, including name, age, and occupation. By creating the right kind of question and utilizing TaroWorks' features, you can ensure better data collection. Please consider the following tips:

▶ Avoid using the question type known as free text where possible. The free text question type is prone to leading to data entry errors and is difficult to report on. When possible, use another question type, such as single or multiple choice, which restrict an enumerator to selecting pre-defined responses.

For example, let's say you are collecting the client's occupation. If you use free text, you may run into issues in which you have multiple answers that are equivalent in meaning but treated separately in analysis, (i.e., "chef" and "cook"), or in which misspellings disrupt analysis ("cehf" would be treated differently from "chef"). Reducing the client occupations to a set of categories and using a multiple choice question would avoid this issue.

▶ When collecting numerical data or dates set an appropriate range. For example, you may want to only survey adult clients. In that case, you would set the minimum age to 18 (or some other minimum age). Adding a maximum value, such as 110, will also reduce data errors.

▶ It may make sense to ask the same question twice, but in different ways. Doing so can permit data validation for potentially confusing questions or important data points. For example, you might collect the person's age and their date of birth and compare the two as a quick way to check for data quality issues.

▶ The more data that you collect, the more variables you have to run analysis, but keep surveys as short as possible while still collecting the necessary information. Superfluous data collected goes unused, clutters your database, and requires more storage space. What's more, too many questions may lead to "survey fatigue." Survey fatigue occurs when respondents (or even enumerators) become restless due to a survey's length, either in terms of time or the number of questions asked. Respondents may ask to end the survey early or take less care when answering after survey fatigue has set in.

Build surveys by starting with the PPI template.

When you want to use the PPI in a survey, create the survey by selecting the desired PPI template in TaroWorks and then adding the other questions. TaroWorks does not currently allow PPI templates to be added to existing surveys.

Use the GPS collection feature.

Visiting a client's home is not a requirement to administer the PPI, but your organization may want to include GPS data collection as a question in the survey if your operations do require such in-home data collection as part of ongoing interaction with clients. Virtually all Android devices support GPS collection, so consider adding a step to gather the location where the survey is being conducted. This can then be used as a validation point to ensure the enumerator was in the correct location. (See documentation for using this feature here: https://taroworks.zendesk.com/hc/en-us/articles/202629960-Using-the-GPS-Location-Question-Type.)

Use the image collection feature.

TaroWorks supports the capture of images from mobile devices that have cameras via the photo question type. You can use this feature to validate that the enumerator interviewed the correct client. This may also be useful for your records because photos of clients may be used for other purposes. (See documentation for using this feature here: https://taroworks.zendesk.com/hc/en-us/articles/200146469-Using-the-Photo-Question-Type.)

Use the signature collection feature.

TaroWorks also supports signature collection. Using the same mobile device that enumerators use to enter responses to questions, respondents may use their fingers or a stylus pen to sign their names. You may want to add this to your survey at the end as an extra validation that the enumerator interviewed the correct client. It can also be used as a compliance step since the client could confirm his or her responses with a signature. Furthermore, it can be stored and used for signature verification in the future. (See documentation for using this feature here: https://taroworks.zendesk.com/hc/en-us/articles/200571074-Using-the-Signature-Question-Type.)

Of course, one issue with requiring a signature is that not all clients or beneficiaries may be able to sign their own name. If this is relevant to the context of your work, consider the image collection feature.

Prevent enumerators from reviewing a client's previously collected responses to the PPI.

Reviewing a client's previous PPI responses may cause the enumerator to skew the data. The previous responses may influence the enumerator's judgment of the actual responses at the time of re-survey. By knowing the previous answers, enumerators may also be tempted to repeat a response without actually asking the client.

For example, question 9 of the Afghanistan PPI asks "Does the household own any motorcycles or cars?" The enumerator, knowing the answer was "Motorcycle only" previously, may not see a car on her follow-up visit and assume the answer is the same.

Take care when asking questions similar to PPI questions.

By asking and collecting information that is closely related to a PPI question, the enumerator may be tempted to mark a response without properly asking the specific PPI question. Staff weariness can also contribute to this – if it's been a long day of survey taking or the survey itself is particularly long, staff may be tempted to mark responses without asking the questions in order to save time. Therefore, when drafting new questions, run through the following checks:

- ▶ Is the question clear in what it is asking?
- ▶ Can it be worded differently in order to lessen confusion?
- ▶ Is this data already being collected elsewhere in the survey? If so, is there a good reason to ask for this again (i.e., for validating answers)?

Reorder the PPI questions to better suit your survey.

The PPI questions may be asked in a different order and/or split apart in a survey. Moving PPI questions around to different sections of your survey can improve its flow and reduce the time required to complete the survey.

Test surveys before launching.

Before launching a survey, be sure to test it using sample data. A best practice is to use already existing data, if available. For example, if you were collecting PPI data before by using paper forms, collect a sample set from those previous surveys, and use it to test and validate your process.

This can be a good way to uncover issues with surveys, field mapping, and reports before beginning data collection. It is also a good way to test the flow of the survey and discover the time required for each survey. Finding errors and correcting them after launching a survey can create problems for data quality and reporting, and in extreme cases may require restarting the data collection initiative.

Enumerators

Motivate enumerators.

Recent research on motivation in the workplace has shown that much of our previous theory on how to inspire staff and get more work out of them is actually counter-productive. Here are some basic recommendations to help motivate your staff:

- ▶ Share overall results from the PPI surveys. Make sure staff can see the actual output from all of their work.
- ▶ Share the reasoning behind conducting these surveys. Knowing that this work is meant to improve the lives of others can help motivate staff.
- ▶ Set attainable survey goals that take into consideration workload, travel time, staff experience, and client availability. Review these targets regularly, and adjust as necessary. If your goals are set too high, data quality may decrease as enumerators focus more on completing surveys than collecting accurate data.

Use TaroWorks features to provide information to your enumerators.

TaroWorks includes the following options for sharing information with your enumerators through the mobile app that can be used to inform your enumerators:

- ▶ Questions have a "Supporting Text/Instructions" field that can be used to provide more information about each question. (Currently, this feature is not available for PPI questions.)
- ▶ Jobs have an "Instructions" field that can be used as a refresher for enumerators.
- ▶ You can add instructional videos to jobs as a "View Resource" task for jobs.

You can leverage these various features in order to:

- ▶ review the survey training, which can be of particular use to new enumerators;
- ▶ highlight new processes or changes to existing procedures;
- ▶ address current issues and problems your team may be having while conducting surveys; and
- ▶ share general tips with your team.

Repeat training at regular intervals.

Over time, enumerators may forget parts of their training or unintentionally develop poor practices. This can lead to reduced data quality. In order to avoid this deduction, provide refresher trainings on a semi-annual or annual basis. These sessions should generally be a review as opposed to a full training and should focus on covering the key processes and most common challenges.

Additionally, these sessions can be used to update enumerators on changes to processes, discuss common difficulties, and collect feedback. Enumerators should be encouraged to share their experiences with one another and discuss their work, including challenges they face.

It is also worth considering hosting a review session shortly after the initial training for new enumerators or new surveys. Typically, enumerators are more prone to making mistakes during the first few weeks of putting training into practice. By scheduling a review shortly data collection has begun, organizations can quickly reiterate the primary instruction, as well as address any potential problems early on.

Train enumerators to confirm answers with clients before continuing to the next question.

A general best practice when conducting surveys is to repeat a client's answer back to him or her to verify that the correct answer has been selected. It is better to do this before tapping "Next" so that the enumerator can verify they have selected the correct answer; however, after advancing to the next question, a tapping "Previous" will allow enumerators to change responses if they realize they have made a mistake.

Do not share login details.

It is common practice for employees to share login details with coworkers or for organizations to allow individuals to use a single login. While this is generally done with the best of intentions, it can cause many issues, including the following:

- ▶ Tracking individual enumerator's performance is more difficult, as surveys may be attached to the wrong staff member.
- ▶ Following up on complaints and/or issues is more difficult because identifying the correct staff member is less straightforward.
- ▶ There is an opportunity for questionable behavior.

In cases in which sharing an Android device becomes necessary, such as when a device breaks or runs out of charge in the field, establish proper procedures for staff to log in and log out of their devices before sharing. Be sure to include this in your standard operating procedures.

Auditing

Investigate unusual survey results.

When reviewing survey results, you may encounter unusual data. When you have found unusual or unexpected data, it is best to conduct a follow-up investigation to discover whether the data are valid and whether the enumerator is following protocol; however, don't assume that this data are invalid. The unexpected responses may be valid. Additionally, you may make initial assumptions about expected results that do not reflect the reality of data collected.

Create reports in Salesforce that help to spot unusual or incorrect data.

Build Salesforce reports and dashboards that quickly highlight unexpected data. For important numerical or date fields, decide on expected ranges that make sense. For example, if you are collecting information on farm hectarage, you may decide that appropriate responses are above 0 hectares and equal to or below 500 hectares. In Salesforce, you could create a report that shows any unexpected responses. The criteria for such a report would look as follows:

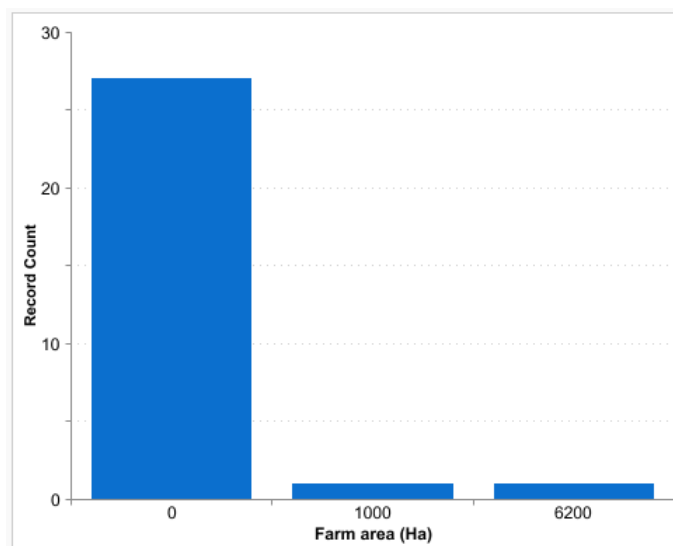Such criteria would display a chart similar to the following and would show the records that should be reviewed:

Another example of using Salesforce reporting to highlight potentially poor data is setting an expected range for poverty likelihoods (such as between 10% and 90%) for specific poverty lines. The criteria using the Kenya PPI and setting the probabilities relative to the national poverty line would be as follows:

Filter Logic: **1 AND 2 AND (3 OR 4)**
1. **PPI Template: Survey: PPI Identifier** equals **"KENYA_2005"**
2. **Poverty Line Identifier** equals **"National Poverty Line_2005"**
3. **Probability** less than **"10"**
4. **Probability** greater than **"90"**

Similarly, reports can be built for multi-select question types or custom text question types by using filter logic such as "contains," "does not contain," "is equal to," or "does not equal." This is a way to highlight responses to the PPI that don't make sense given the context of your working environment (e.g., you work with farmers but the selected occupation is day laborer).

Conduct random sample audits.

When validating results or following up on questionable data, conduct a random sample audit. You should assign these on a semi-regular basis to continue to validate data, processes, and staff – or potentially flush out problems that don't necessarily show up in the results. (Read the PPI Standards of Use for more information on validating data: http://www.povertyindex.org/standards.)

Follow-up surveys should consist of only a sample set, be conducted by enumerators other than the original, and the auditing enumerators should not see the previous data. In fact, the less the auditing surveyors know, the better the data validation practice will be.

To select a sample size, refer to the Sample Size Calculator on each country's page on the PPI website: http://www.povertyindex.org/ppi-country.

Clean survey data.

When poor data is spotted, either through the methods described above or manual data checking, ensure that it is not used in analysis or reporting. If you are able, simply correct the data. If you don't know what the correct information is, then, depending on your need for such information, remove the record or re-survey the client.

In some cases, you may find that a cause for poor data is systematic, such as one enumerator misunderstanding multiple questions on a survey. In such cases, all records that would have been impacted by such systematic cause will either need to be corrected or discarded.