

UNPRECEDENTED
QUALITY CONTROL,
FOR YOUR
BEST DATA EVER.



www.surveycto.com



Electronic data collection with SurveyCTO

- ✓ Random audio audits
- ✓ Behavior-based audio audits
- ✓ Speed limits
- ✓ Photo audits
- ✓ Intelligent workflows
- ✓ Automated data monitoring
- ✓ Advanced skip logic, barcodes, and more

Your data is worth the effort, but we make it easy.



Audits Based on Enumerator Behavior to Collect Survey Metadata



Christopher Robert
Dobility, Inc. (SurveyCTO)

March 5, 2015



Collecting *quality* data

- Want to steadily **lower** the **cost** of collecting **high-quality data** (as opposed to garbage)
 - Random audio audits
 - Speed limits
 - Quietly track violations
 - Trigger audio audits on x violations
 - Automated quality checks




Collecting *quality* data

- Want to steadily **lower the cost** of collecting **high-quality data** (as opposed to **garbage**)
 - Random audio audits
 - Speed limits
 - Quietly track violations
 - Trigger audio audits on x violations
 - Automated quality checks



Collecting *quality* data

- Want to steadily **lower the cost** of collecting **high-quality data** (as opposed to **garbage**)
 - Random audio audits
 - Speed limits
 - Quietly track violations
 - Trigger audio audits on x violations
 - Automated quality checks



Select the quality check you would like to create:

- Value is too low
- Value is too high
- Value is an outlier
- Mean is too low
- Mean is too high
- Value is too frequent
- Value is too infrequent
- Group mean is different
- Group distribution is different



Collecting *quality* data

- Want to steadily **lower the cost** of collecting **high-quality data** (as opposed to garbage)
 - Random audio audits
 - Speed limits
 - Quietly track violations
 - Trigger audio audits on x violations
 - Automated quality checks

QUALITY CHECKS (Dataset: Sample data)

Value '1' is more frequent than 40%
Fields: gender

Group mean differs by group 'enumerator id' with p<0.05
Fields: gender, age

Group distribution differs by group 'enumerator id' with p<0.05
Fields: education

Mean is less than 0.4
Fields: gender

Value is less than 18
Fields: age

Value is greater than 55
Fields: age team related

Group mean differs by group
Fields: ?

Quality checks: overall options
The below options apply to the set of quality checks configured for this dataset. Individual quality checks, listed to the left, each have additional options.

Select the field that uniquely identifies each record:
household id

Run all checks nightly (unchecked to pause)
(Nightly checks will run only when data have changed since last run.)

Send email summary of quality-check reports?

Email address(es):
crobert@surveycto.com

Separate multiple email addresses with commas (e.g., "manager@surveycto.com, teamlead@surveycto.com")



Collecting *quality* data

- Want to steadily **lower the cost** of collecting **high-quality data** (as opposed to garbage)
 - Random audio audits
 - Speed limits
 - Quietly track violations
 - Trigger audio audits on x violations
 - Automated quality checks

QUALITY CHECKS (Dataset: Sample data)

Fields: gender, age

Group distribution differs by group 'enumerator id' with p<0.05
Fields: education

Mean is less than 0.4
Fields: gender

Value is less than 18
Fields: age

Value is greater than 55
Fields: age team related

Group mean differs by group 'enumerator id' with p<0.05
Fields: age team related

Group distribution differs by group '?' with p<?
Fields: ?

Group distribution is different
Use this check to test for equality of distributions across groups using a chi-squared test.

Select the fields to be checked:
 age
 age team related
 random1
 random2
 education

Trigger warning when chi-squared p-value <= this value:
0.05

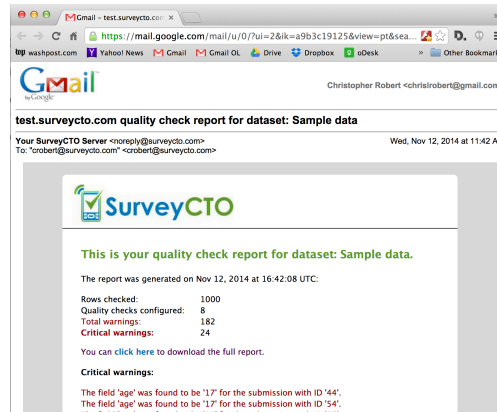
Field that identifies groups:
enumerator id

Critical quality check: flag warnings as high-priority



Collecting *quality* data

- Want to steadily **lower the cost of collecting high-quality data** (as opposed to **garbage**)
 - Random audio audits
 - Speed limits
 - Quietly track violations
 - Trigger audio audits on x violations
 - Automated quality checks

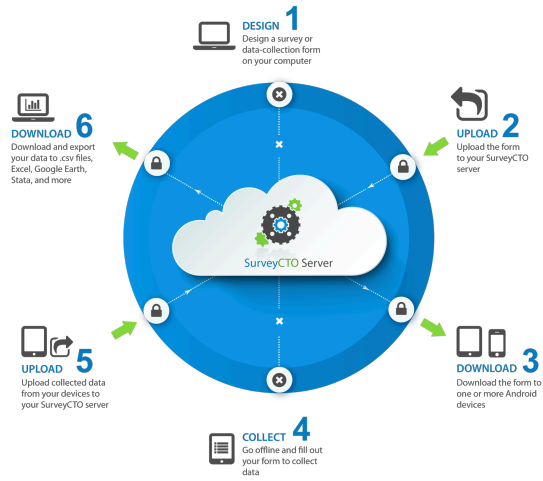


Quality-control options

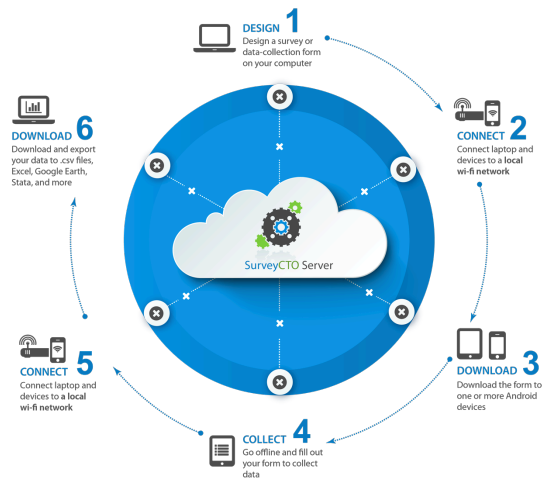
- Entry constraints, field and group relevance
- Barcode-scanning, double-entry, etc.
- Auditing time spent on questions
- Random and non-random audio audits
- Photo audits
- High-frequency checks
- SurveyCTO: supports (and encourages!) all of these



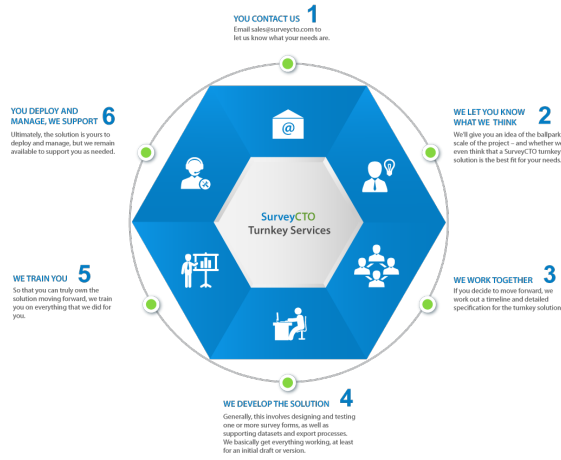
Online workflow



Offline workflow



Turnkey Services



How SurveyCTO Works



Server options

- Host your own server in the cloud
- Use a cloud-hosting service
- Host your own server in-house
- Avoid a server altogether

- SurveyCTO: primarily a cloud service, but self-hosted and fully-offline options available



Survey-design options

- Online WYSIWYG
- Offline WYSIWYG
- Offline XML
- Offline spreadsheet

- SurveyCTO: primarily offline spreadsheet, but supports XML, has builders, templates, samples



A hybrid solution

- **Our solution:** online and offline tools to systematically make common tasks easier

CONSTRAINT BUILDER

To require that a field be non-blank (that something be entered or selected), enter **yes** into the **required** column for that field. When something is entered or selected, only values that satisfy this constraint will be allowed. When a value does not satisfy the constraint, the user will not be able to proceed to the next field; you may specify a specific error message below.

Constraint type:

Allow decimals? No Yes

Minimum: (or blank for none)

Maximum: (or blank for none)

Allowable exceptions: (comma-separated list)

Please enter the error message to display when the constraint is violated.



A hybrid solution

- **Our solution:** online and offline tools to systematically make common tasks easier

EXPRESSION AND ERROR MESSAGE

Copy the highlighted constraint and error message to your clipboard by pressing ctrl+c or cmd+c, then right-click on the **constraint** column in your form definition, click Paste Special, then click to paste as Text.

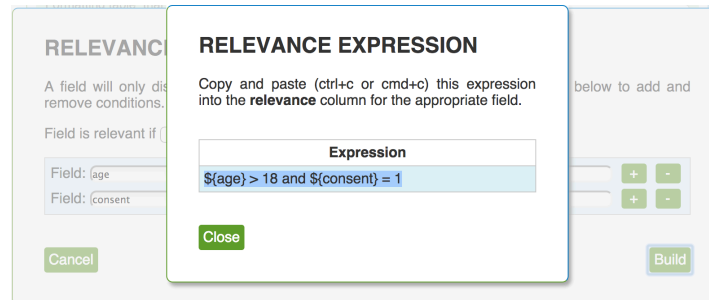
Expression	Error Message
(>=3 and <=120) or (=888 or =999)	Please enter a valid age to continue.

Note: To restrict users to only enter whole numbers, use the **integer** field type in your survey definition.



A hybrid solution

- **Our solution:** online and offline tools to systematically make common tasks easier



Data-pre-loading options

- Just static details like locations, enumerator names and ID's, etc.
- Additional details about targeted households, perhaps from a prior baseline or listing survey
- Information about "cases," team or individual assignments
- SurveyCTO: supports static pre-loading, plus pre-loading of any .csv data as survey field defaults, plus case management



A hybrid solution

- **Our solution:** online and offline tools to systematically make common tasks easier

DOWNLOAD FORM TEMPLATE

Form title*:
(Mock Midline Survey)

Form ID*:
(mockmidline)

Auto-generate fields necessary for pre-loading data? Yes No

Upload a .csv file here:

| No file chosen

Or, select a server dataset here:

Please attach a .csv file or select a dataset that demonstrates the format of the data you will pre-load. The data does not have to be final, but the name of the file or dataset does need to be correct, and the column names need to be correct as well. Additional options will appear after you click Download below.

Do you want this form's data to be encrypted? Yes No



A hybrid solution

- **Our solution:** online and offline tools to systematically make common tasks easier

PRE-LOADING TEMPLATE WIZARD

We recognized 11 columns in your data: household id, gender, gender string, enumerator team id, enumerator id, age, age team related, random1, random2, education, education team related

Here is a preview of how your data appears:

household id	gender	gender string	enumerator team id	enumerator id	age
1	1	Male	3	11	21
2	0	Female	2	7	55
3	0	Male	3	12	34

Please review the first few rows of data before continuing. If some characters are not displaying correctly, you may need to use the SurveyCTO Client to re-encode the .csv file as UTF-8 format before uploading it.



A hybrid solution

- **Our solution:** online and offline tools to systematically make common tasks easier

PRE-LOADING TEMPLATE WIZARD

Which data column uniquely identifies each row?

Data is pre-loaded using the pulldata() function, which requires such a column.

How should your form determine which row to load?

To identify the row to pre-load, the template can include one, two, or three select_one fields, or a single text field. Examples of multi-step cases: (a) enumerator chooses team, then enumerator, then household, (b) enumerator chooses district, then village, then household, or (c) enumerator chooses district, then clinic.



Open source technologies

- **Are great...**
 - avoid reinvention of wheel
 - flexibility and interoperability
 - seemingly can't beat the cost (nominally free)
- **Are not so great...**
 - lack of adequate support
 - often designed by techies for techies
 - often require server admins, programmers, other technical consultants
 - *can turn out to be extremely costly in practice!*
- **Our solution:** Open Data Kit (ODK) – plus hosting, plus hundreds of improvements, plus help and support



Example: data security

- **Ethical and/or IRB requirement:** don't have sensitive respondent data vulnerable on public servers
 - Don't even want a wide circle of "authorized" users to have access to data
- **Our solution:** make ODK's support for public/private key pairs **easy to use**, build workflow and design product around it

The image shows two parts of the ODK software interface. The top part is a 'Tools' menu with the following items:

Tools	
Create a public/private key pair...	⌘K
Re-encode .csv file...	⌘E
Schedule sync...	⌘Y
Form tools	▶

The bottom part is a 'DOWNLOAD FORM TEMPLATE' dialog box with the following fields and options:

- Form title*: (Mock Baseline Survey)
- Form ID*: (mockbaseline)
- Auto-generate fields necessary for pre-loading data? Yes No
- Do you want this form's data to be encrypted? Yes No
- Upload public key file here: No file chosen
- Or, paste public key text here:
- Buttons: Cancel, Download to Google Drive..., Download to computer...