



TREADWELL

Better data. Greater impact.

Monthly Event Topic:

DATA CLEANING

Ben Greenberg and Brittany Willard
January 17th, 2024

Today's Agenda

- Presenter Introductions
- Why Data Cleaning is Important
- Common Data Quality Problems
- ETO: Data Cleaning Strategies
- Apricot: Data Cleaning Strategies
- Best Practices in Data Cleaning
- Questions & Comments
- February 2024 Monthly Topic

Your Presenters

Ben Greenberg
Managing Consultant
Salt Lake City, UT



Brittany Willard
Managing Consultant
New Castle, DE



Why Data Cleaning is Important

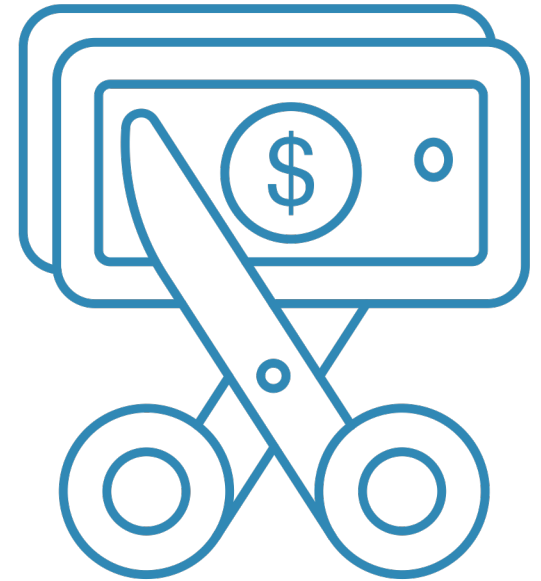
Reliable Insights from
Accurate Data



Avoiding Staff Frustration
due to Errors in Data



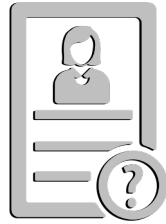
Minimizing Errors that can
Lead to Financial Losses



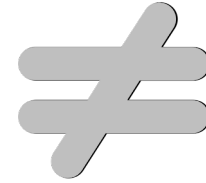
Common Data Quality Issues



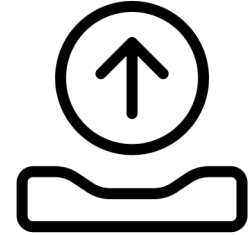
Duplicates



Incomplete/
Inaccurate
Records



Inconsistencies
in Formatting



Importing /
Batch Uploads

Breakout Rooms

DATA CLEANING STRATEGIES

Data Cleaning Strategies

eto^o

DATA CLEANING STRATEGIES



- 1) Duplicates – duplicate check settings & duplicated participants
- 2) Incomplete/Inaccurate forms: Touchpoints vs. Demographics
- 3) Fakes
- 4) Inconsistencies in formatting
- 5) Batch Uploads

Data Cleaning Strategies



Duplicates

- Duplicate Check Settings
 - Why use “like”?
- Duplicated Participants
 - Gives you a list of participants who meet duplicate check setting criteria
- Merge Duplicates
 - Allows you to search for suspected/potential duplicates by Last Name, SSN or Case Number
- Both tools allow for merging duplicated records
- Once merged, this process can NOT be undone

Data Cleaning Strategies



Incomplete/ Inaccurate Records

- Demographics vs. Touchpoints
 - Demographics = Data that ISN'T expected to change
 - Touchpoints = Data that IS expected to change
- Incomplete Demographics
 - What's required? Should it be?
 - What's disabled? Should it be?
- Incomplete or Unused Touchpoints
 - Empty fields – is this field consistently empty across records?
 - Are all currently active touchpoints being used?
- Conditional Rules
 - No two conditional rules can create a conflicting action
- Fakes
 - Should they still be in my system? How do I get rid of them?
 - What if I need them in my system, but not counted in reporting?

Data Cleaning Strategies



Query Filters

> <

AND

Family Name Different from pattern %fake%

Name Different from pattern %fake%

Icons: [Print] [Trophy] [Close] [Trash] [Up] [Down] [Share]

Data Cleaning Strategies



Inconsistencies in Formatting

- Dates
 - Is everyone using the same format/field for dates?
- Touchpoint building and formatting
 - Are there too many fields in my touchpoint?
 - Can this be broken down into smaller touchpoints?
 - Have I used touchpoint settings to my advantage?
- Use of “Other”
 - this value allows multiple variations of the same option
- Field changes – how will this affect reporting?
 - “yes” to “Yes”
- Importing – Batch Uploads

Data Cleaning Strategies



Batch Uploads

- Clean data in, Clean data out
- Are you set up for success?
 - Securities
 - New template every time
 - Data double-checked
- Utilize Results
 - Build extract reports that follow the template
 - File format – CSV is generally your friend
- Special characters – ', &, ñ
- Test Runs – Always run a short test (10 records) before any major import
- Break into smaller imports if you have a large number of records

Data Cleaning Strategies

apricot

Data Cleaning Strategies



Duplicates

- Tier 1 (Ex. Client)
 - Duplicate Check Settings
 - Merge Tool ([turn on here](#))
 - Reports
- Tier 2
 - Duplicate Checks
 - Archive Permissions
 - Reports

Field Properties (dropdown)

Display Name

What is your Favorite Color

Standard Properties

<input checked="" type="checkbox"/> Required	<input type="checkbox"/> Duplicate Check	<input type="checkbox"/> Locked
<input checked="" type="checkbox"/> Quick View	<input type="checkbox"/> Searchable	<input type="checkbox"/> Hidden
<input type="checkbox"/> Clear On Copy		

Data Cleaning Strategies



Incomplete/ Inaccurate Records

- Incomplete Forms
 - Workflows
 - Linking (especially Wizard Linking)
- Incomplete Records
 - Required Fields
 - Form Logic
- Inaccurate Records
 - Form Logic
 - Lookup Lists
- Reports

Data Cleaning Strategies



Inconsistencies in Formatting

- Beware of the "Allow Other" option in Apricot
 - For opportunities to expand options and remove other
- The Import Tool is Your Friend
 1. Report for Concerning Fields/Forms
 2. Assess Formatting Issues
 3. Export Report
 4. Transform Data
 5. Re-Import (as an Update Import) into Apricot

*Tip: be sure to build the form record ID into the report
- Be Thoughtful about Field Changes – what impact might this have on reporting?

Inconsistencies in Formatting

Harry Potter -
Quick View Information ▶

Main ▼

***What is your Favorite Color**

--Please Select-- ▼

- Please Select--
- Red
- Orange
- Yellow
- Green
- Blue
- Violet
- Other

Harry Potter -
Quick View Information ▶

Main ▼

***What is your Favorite Color**

Other ▼

Other Value

This field is required.

- ▶ Pink
- ▶ Pink/Purple
- ▶ pur
- ▶ purpl
- ▶ Purple

Data Cleaning Strategies

- Clean Your Data Prior to Importing into Apricot
- Use the Import Instructions when Preparing for Import
- Be Mindful of “Allow Other”



Importing

Import Instructions for Inserting Participant Using Duplicate Checks ▾

* Participant - Name (field_2057)

Any value (including special characters).
The first name and last name fields are required, but the middle name field is not.

* Participant - Date of Birth (field_2059)

Accepts the following date formats
YYYYMMDD
YYYY-MM-DD
January 1st, 2009
Jan 1 2009
mm/dd/yyyy

Email (field_2058)

Any number of characters followed by an '@' followed by any number of characters

Primary Phone (field_2060)

Accepts numerals, spaces and the following symbols: . () - ext x #
Examples
555-555-5555
555.555.5555
555 555 5555 ext.55
(555) 555-5555

DATA CLEANING BEST PRACTICES

Best Practices

Establishing Data Quality Standards

- Assess Processes and Define Quality and Completeness
 - Consider grant and funder requirements

Staff Training and Awareness

- End User Training Sessions
- Training Manuals
 - We love Tango for this!
- Encourage Staff to report any potential issues they notice

Continuous Monitoring and Improvement

- Identify High Need Areas
- Establish a Schedule
- Work with program staff

QUESTIONS OR
COMMENTS?

UP NEXT IN FEBRUARY