# idealware

# Data Quality Best Practices for Nonprofits

**Data should be an integral part of every nonprofit's decision-making. But for decisions to be informed by data, the quality of the data needs to be consistent and reliable. So how do you make sure your organization's data is good?**

When informed by data, decisions can affect every area of an organization, getting volunteers moving, activists petitioning, and donors giving. It can also lead to a better understanding about who an organization interacts with, and how.

With data, a nonprofit can determine the right time and tone for asking for more donations. It can figure out when volunteers are most needed and allocate resources accordingly. In short, a data-informed nonprofit uses its resources efficiently to better help its constituents.

But for decisions to be informed by data, the quality of the data needs to be consistent and reliable. In the 1950s, at the dawn of the modern computer age, someone coined a phrase to express the idea: "Garbage in, garbage out." Said more eloquently, faulty inputs result in faulty outputs. A decision made with bad data is a bad decision.

So how do you make sure your organization's data is good? First, let's define what we're talking about.

Ask a dozen people what "data quality" means and you might get a dozen answers, all varia-

tions on the same theme: good, clean, accurate, consistent, reliable, complete. But for data to be considered quality, it must also be useful. For example, your organization may have a spreadsheet of accurate phone numbers for all of its constituents, but it's not useful if you want to email the people on that list.

Or, as another example, you might have three separate records for the same donor: one as Robert Williams, his actual name; one as Bob Williams; and one as William Roberts. He's still going to get your emails, but he's going to get them three times. And

one of those emails will have a wrong name. He may get frustrated and not donate. And your understanding of how many donors you have is going to be inaccurate.

## CLEANING EXISTING DATA

If you're setting out to make sure your organization has access to quality data, start with existing data before establishing processes to ensure that new data also meets your needs. Adding new data into a messy database is like piling in junk in a overcrowded house—the new things won't be useful, and the house only gets messier.

Start by prioritizing—ask yourself what are the top five things you want your data to tell you and make a list. Then, think of all the campaigns, programs, and efforts you would kick off if your data was reliable. Break those down to the data points you most need to move those efforts forward, and start there.

Let's look at one example—say, a nonprofit that relies on individual donations for the majority of its program funding and wants to win back lapsed donors. To meet this goal, staff first has to determine what kind of data will help target lapsed donors best. They begin by listing relevant data fields from the database:

*Sources of Donations*. Rather than limiting a lapsed donor campaign to online donations, they consider all the ways people might donate and how those donations find their way into the database.

*Date of Last Donation*. Does every known donor have a value in this field? Do they include all sources (Online, Offline, Event-Based)?

*Relationships and Households*. Can they determine if someone who is considered a lapsed donor is the spouse or partner of someone who remains an active donor?

*Channel of Previous Donations*. Are all possible channels properly recorded for each donation? Using the appropriate channel can help meet donors where they are comfortable, removing friction.

*Campaign of Previous Donations.* If the donor's previous donations come from campaigns, what were the themes of those campaigns? This can help in developing targeted language to get the donation.

Once staff has identified those data points, they can analyze the data to see if it's accurate and complete. To do so, they ask the following questions:

- Do all donors have data in the 'last known donation' field?

- Does the summary data pass a "gut check?" For example, if they know that the annual gala is always a big revenue generator, where does it show up in a revenue report?

- Do all donations have a source so your team can do analysis and know how to talk to lapsed donors?

- Is all of the contact information accurate and functional?

- Does the database enforce address formatting to make it easier to find households? If so, do they have a full sense of household relationships among the constituents? Can they tell the difference between a married couple or roommates at the same address?

*Note*: Household relationships can be a challenge for databases—especially when it comes to donor data. For example, you may have multiple records for different people at the same mailing address, but how do you interpret that data? Are they married or in a relationship? Are they roommates? Parent and child? Are they unrelated, living in different

apartments in the same building? There is no easy solution, but being consistent with how you collect the relevant data can mitigate problems.

If you're looking to mount a similar effort to the one in the example at your organization, how do you go about finding and fixing data problems?

There are a variety of approaches. Which method you choose will depend on what your database can do. Some of the more common options are covered below.

## Scan the Database Records

A problem with missing data may be easily detecting by running a simple report in your database and sorting by the field you're trying to find information about. In the example, the organization might sort for Date of Last Donation and look for missing fields.

## Use Built-In Database Features

Many databases come with built-in features for cleaning data, such as "deduping" tools—which search for duplicate records and merge or eliminate them—and "insight"-style functions that range from checking to make sure email addresses aren't the same to advanced AI features that notify users of potential problems.

## Export to Spreadsheets

Spreadsheet tools like Google Sheets or Microsoft Excel include a number of powerful features you can use to find bad data. For example, conditional formatting can let you find duplicate data. Sorting and Filtering can help you analyze. And COUNTIF and SUMIF functions let you how many of your records may have aberrant data. (For more information, see the *Transforming Data* sidebar.)

## TRANSFORMING DATA

In data terminology, "data transformation" simply means "make the data look different." You do this to ensure consistency across data sources. You can transform data in many ways—including buying sophisticated ETL software which is designed to "Extract, Transform, and Load" data—but for many uses, a spreadsheet can do the trick.

Here are some neat spreadsheet tricks, from the simple to the sophisticated:

*Highlight Duplicates.* Excel has a simple Conditional Formatting option to highlight duplicate values, which is very handy when deduping.

*Tables/Sorts/Filters.* In Excel, the Format as Table function can make any dataset sortable, filterable, and easy-to-read. Google Sheets offers similar, but less robust, functionality with its filter views.

*Text to Columns.* If you think you have multiple data values in a single field (like first name and last name entered in the Last Name field), you can use Excel's Text to Columns feature to split them (in Google Sheets, it's the Split function).

*String Functions.* Both Excel and Google Sheets offer standard string functions to substitute and replace values. Let's say your phone number format should be (555)5555555, but you have values set at 555-555-5555—use the CONCAT function to combine the text from multiple strings, and SUBSTITUTE function to replace it and standardize the formatting.

One thing to remember when it comes to transforming data is to always keep the raw data handy to allow you to undo any mistakes.

## Use Third-Party Tools

There are several data integrity tools on the market that may integrate with your database, enabling data cleaning. You can do a one-time "scrubbing" to clean all your existing data, or use them in an ongoing capacity to keep data clean moving forward. There are also tools which can help verify mailing addresses and telephone numbers for your constituent records.

### INTRODUCING NEW DATA

Once you've cleaned your existing data, ensuring its quality, how do you move forward introducing new data from other sources? Let's take a look at a few different types of data sources and integrations to better understand the path forward: Integrations between your own internal software platforms, integrations between your database and an external platform, and introducing data from a list you purchase or rent.

### Integrations Between Platforms

Nonprofits often use more than one tool to communicate with constituents—for example, a Constituent Relationship Management database, a broadcast email platform, and an event registration tool to allow people to sign up for events online. Rather than keeping all this data "siloed," or kept separately in different databases, it's best to integrate all the different components so that the same data is shared across the platforms. (For more information on CRM integration, download our free report, *CRM Integration for Nonprofits*.)

Remember, garbage in, garbage out—to keep the data high quality, it needs to be in good shape before it goes into the database. When integrating multiple platforms, this means cleaning the data in each before integration. If there are differences or discrepancies, transform them (see sidebar, below) first.

For example, your broadcast email database may use a 5+4 Zip code for constituents, but your CRM uses a simple five digit-Zip. To remedy this, you might export the email database into a spreadsheet and convert the Zip codes to fit the five-digit format before integrating it with your CRM data.

### Integrations From External Platforms

Integrating data from outside tools, such as form managers, can be more complicated than integrations from among your own databases because they may not enforce required fields and field formats the same way yours does, which means you'll need to take a few steps to mitigate the risk of introducing low-quality data.

If the volume of data you're important is small, or the stakes are low, you can get away with performing regular data checks and updates (either manually or in bulk) within your database. But if the quantity of data is large, or the stakes are too high to risk mistakes, then consider using an Application Programmer's Interface (API) or other third party ETL tool designed to extract, transform, and load the data into your database. When going this route, make sure you also follow a de-duping protocol to find duplicates and reconcile conflicting data (for example, a user may be "Robert" in your database but "Bob" when he takes his online survey).

### Buying/Renting Lists

It's possible to buy or rent lists from vendors, but without careful planning, these lists can wreak havoc on your data quality. Here are some tips when getting data from third-parties.

- Make sure these names are from a reputable vendor. Ask around.

- Follow a legal and responsible opt-in process before contacting or soliciting acquired prospects.

- Plan for what should happen if someone on the list is already on your list, and make sure to run whatever deduping tools are at your disposal when importing the list.

- Keep an accurate "no contact" list and run any third-party data additions against it to make sure you honor the wishes of anyone who used to be on your own list but asked to no longer be contacted.

- Transform the data in a way that is consistent with the database before importing it.

## REGULAR CARE AND MAINTENANCE

Cleaning your existing and imported data are one-time, standalone actions. Data hygiene is the process of keeping your data clean and ensuring that new data entered into the system retains the integrity. The following practices can help ensure good data hygiene.

**1. Dedupe regularly.** Most CRM databases come with deduping tools built-in. If deduping isn't enforced in real-time, make sure it occurs on a regular schedule. Note that it may be more than just making sure there aren't multiple records with the same email address. Some users may use more than one email address, or a single household may contain more than one person. Consider your goals for the information when creating a deduping protocol. (For more information, read our free _Nonprofit's Guide to Data Migration_.)

**2. Set up automated reporting.** The most time-saving methods let the database do the work for you. Many CRMs have "exception reports," or automated reporting tools built in that can highlight problems while they are still small. Use them to find oddities in data—when run regularly, exception reports can help keep data clean and reliable.

**3. Eliminate blank fields, or N/A field entries.** Make sure required fields have a useful value. During integrations, requirements on these fields may be bypassed—or someone may enter "N/A" or a similar term to bypass a required value during manual entry.

**4. Watch for unusual activity.** Take past data quality issues into account, and then set up regular reports to ensure that they are not repeated. For example, if email deliverability crashed because of a high number of bad email addresses, consider running a regular deliverability report. Or, if a donor who regularly donates $1,000 a year donated only $10, consider running a report to find other such discrepancies to help determine why they are happening. Note that not every exception will be the result of "bad" data—sometimes the data is accurate, and the donor just decided to give less money—but exploring the anomalies can help you learn.

**5. Segment data properly.** Have you ever gotten a call from an donor angry about being solicited after they had already donated? Consider running a report to see how long it takes for donations to appear in constituent records, and then think about how to improve the data flow so you don't annoy constituents with too many asks too soon.

**6. Enforce consistent data formatting (except when intentional).** There may be good reasons not to enforce consistent formats on certain fields—for example, you may not enforce a five-digit Zip code if you have constituents outside of the U.S.—but you may still want to make sure the data is relatively consistent and accurate. If you are having difficulty enforcing formats for certain fields, run a report that catches common but invalid values.

## CONCLUSION

Data hygiene is an ongoing process. As already mentioned, you can and should set automated reports to help catch as many issues as possible. But sometimes manual involvement—a hands on process known as "touching" the data by looking at it directly—can uncover issues as well.

An easy place to start is by scanning the larger reports and tables and looking for a high number of "N/A" entries, strange characters, or inconsistent formatting. Once you've fixed those, look for the weak spot in your data entry process that's causing the errors and make changes to avoid them. Don't forget to set up automated reports to catch them on a regular basis.

Make this a regular part of your routine by scheduling it—weekly, bi-weekly, or monthly, for example. Look at 10 random records from the database, for example, and keep an eye out for anything unusual, such as missing fields or records in segments where they don't. If you find something, run a report to determine how widespread the problem is. Regular maintenance routines go a long way toward helping to ensure that your data is consistent, reliable, and useful, and that it remains that way.

## ABOUT THE AUTHOR

*Marissa Goldsmith is a digital analytics specialist. She works with nonprofits, government agencies, and small businesses, implementing digital data collection strategies, integrations, and helping them to use the data gleaned in a privacy-friendly way.*

## ABOUT IDEALWARE

Idealware, a program of Tech Impact, is the authoritative source for independent, thoroughly-researched technology resources for the social sector. Our publications, assessments, and training can save you time and money by providing impartial guidance that gives you the knowledge and confidence you need to decide what's best for your organization. To be notified about new publications, sign up for our mailing list at www.idealware.org.