

Surviving the Duplicate Epidemic:

A Guide to Stopping Duplicate Data in your CRM and MAP Before it Enters

It's no secret that poor data quality is among the most frustrating and costly issues facing the modern data-driven business. In fact, organizations surveyed in Experian's 2018 global data management benchmark report believe 33% of their customer and prospect data is inaccurate in some way.

What is data quality?

Data quality is all about the accuracy, completeness, and trustworthiness of the data throughout your organizational systems. The ramifications of bad data are not limited to one department. Sales, marketing, finance, and IT alike are all negatively impacted by dirty data. Tracking campaign success, ROI, and revenue attribution become impossible when the data behind the campaign is faulty from the start.

That is why data-driven organizations across the world have made it a priority to improve data quality through proactive data governance strategies: namely, **database safeguards**.

The digital economy has changed the way that businesses operate, putting great importance on leveraging data insights to drive business decisions. And despite the undeniable value of data, many businesses struggle to put their data to good use. To survive in this digital economy, organizations need to lay a foundation of high-quality data; and that begins with examining the entry points of your database to stop the entrance of bad data before it rears its ugly head in a CRM and MAP system.

The Duplicate Epidemic

The biggest indicator of poor data quality is the existence of duplicates. Simply put, a duplicate is the same information strewn across two or more customer (person) records, yielding redundant data and fragmented single customer views. **Duplicate records cause:**

- Storage overages
- Improper communications
- Ineffective nurture campaigns
- Skewed analytics
- Lost revenue.

Silos between databases and duplicated information can make it nearly impossible to have a single source of truth, giving new meaning to the phrase 'untrustworthy data.'



CRM and MAP duplicates can skew campaign tracking and results, often making it difficult to accurately attribute revenue. Duplicate leads and accounts can also result in account ownership conflicts, false leads, and inaccurate revenue forecasts.

Imagine if you are trying to track a single lead throughout the stages of the funnel, only to learn that they are already a customer. This is an unfortunate reality faced all too often by sales/marketing teams working inside databases rife with duplicates. The presence of duplicates is not only a harbinger of a poorly managed database, but it can paint a good picture of the areas in which data collection and acquisition could be improved.

Where does data come from?

To discover the source of duplicates, we must examine the entry points of your CRM and MAP.

Collecting data is only half the battle within the larger scope of data operations. The other half, the much more labor-intensive half, involves ensuring the integrity and trustworthiness of the data coming in. This starts with monitoring the entry points and optimizing the process to ensure as few duplicates enter as possible -whether through manual intervention or automated means.

The best way to prevent duplicates is to implement rules, protocols, and governance practices throughout your organization to combat the issue. Simply informing CRM and MAP users of the harmful effects of duplicates may be enough to encourage healthy data entry practices. These initiatives, however, are nearly impossible to uphold for extended periods of time, as monitoring each user's CRM usage is not practical. Safeguarding the entry points to your CRM and MAP requires careful thought and planning, especially with high volume data collection at larger organizations.

Manual Entry

When a member of a sales team -whether it be an account executive or business/sales development representative- generates a new lead, they must enter that into the CRM. But with high quotas and otherwise demanding job responsibilities, it is all too easy to forget to check if this Lead/Contact exists in the database already and whether or not it should be linked to an existing Account. **Thus, we have the first source of duplicates.**

Ideally, a CRM user searches for a record before creating it; and if it exists, they append the existing record with new data and/or notes. But in an effort to maintain speed and efficiency in daily sales workflows, this practice is impractical and unsustainable.



Manual entry can be among the toughest duplicate type to stop because it means enforcing rules amongst CRM users that are hard to adopt. As stated earlier, most sales teams simply do not have enough time to check incoming records against existing data, especially in organizations where this data may live in multiple Orgs.

Not to mention if the user manually locates an existing record, it is up to their discretion whether to override existing information with new data. The potential loss of valuable data is precisely why manual entry duplicates are so prevalent in most databases.

How to Prevent Manual Entry Duplicates

Unfortunately, Salesforce lacks native functionality to effectively prevent manual entry duplicates. Salesforce's native Dupe Alerts allows admins to define logic, but requires users to type in an entire record, then click save, only to find out they have created a duplicate. This is opposed to the smarter solution- trigger-based duplicate protection.

There are 3rd party solutions for Salesforce that provide as-you-type duplicate alerts like [Unique Entry](#). Many organizations opt for a real-time duplicate checker that is trigger-based.

Before you even click "save" on a new record, a strong duplicate prevention system will go and check your new entry against existing ones to find the duplicate. The detection point should be once a few fields are entered versus after you've hit save.

In any case, manual entry duplicates are easy to stop in theory. But enforcing hygienic data entry practices amongst CRM users in an organization is easier said than done. If budget allows, external applications are available to stop manual entry duplicates.

Web Forms & Web-to-Lead

Gathering inbound web leads is often one of the most profitable avenues of demand generation. From PPC advertising to organic traffic via SEO and social sharing, web generated leads are lucrative because the interest is already there, saving a lot of nurture time and effort.

But as with manual entry duplicates, the automated systems in place to record and store inbound web-to-lead information are often flawed in the way that they handle the presence of duplicates.

Take for example a situation where someone fills out a form to download a gated resource from your website but this lead is already a customer. Automated systems may enter this individual's email address into a nurture campaign aimed at the beginning stages of the funnel, disregarding their status as an existing customer.



And because this customer record is fragmented between a customer record and a lead, the duplicate causes brand reputation issues and inflated storage costs. This is just one example of a web-to-lead duplicate, but it is clear that they are often the most disastrous to sales and marketing activities.

How to Prevent Web-to-Lead Duplicates

The first thing to do when attempting to clean up the web-to-lead data collection process is to examine the volume of information coming in and their sources. If you receive a high volume of leads from a specific campaign, tracking this data is critical to attributing revenue.

Duplicates can skew these numbers and must be prevented as a high priority. Today's marketing landscape includes multiple data entry points which are tied to various prospect touch-points, including white-paper downloads, free trial sign-up pages, contest entry forms and more. Marketing automation systems like Marketo have exact email matching rules that users can set to merge incoming data with existing records, but most people find that this is not robust enough.

That is where an external solution to link leads to accounts and merge incoming duplicates is necessary for a greater volume of leads. These third-party applications often connect to your MAP through API or javascript integrations, facilitating the web-to-lead process through automation. If budget allows, these solutions are commonly agreed upon to be the best solution to prevent web-to-lead duplicates.

List Imports

So you've purchased a list of contacts or perhaps gathered them from an event. Logically, you need to enter these new leads into your CRM as efficiently as possible to maximize ROI of all the money you've spent acquiring these leads.

Unfortunately, importing this lead list into Salesforce without protection leaves databases highly susceptible to duplicates. Oftentimes, there are even duplicate leads or contacts in the lists themselves, creating even more chaos.

Your list upload strategy should include your CRM from the start. These cannot be independent. Keeping the two connected as much as possible is crucial for experiencing cleaner data from the get-go. Because of the nature of list files, often times a simple excel duplicate formula is enough, but this is as basic as it gets. And this reduces the chance of finding fuzzy duplicates (e.g matt@ringlead.com & matthew.johnson@ringlead.com).



How to Prevent List Import Duplicates

Most Salesforce users employ the data loader to upload files from trade shows, list buys, acquisitions, etc. But the issue with the data loader is that it does not allow users to set up advanced duplicate prevention logic like fuzzy matching nor flexible duplicate field actions.

There are applications that allow you to control what happens when duplicates or related records are found. Creating tasks or appending data to fields on existing records yields cleaner data with little to no data loss. A strong list vendor and/or integration with your CRM will detect the duplicate, and give you the option to view both records -- old and new -- side by side. Then, you can merge the existing record and new record resulting in a single, updated record.

While list uploads may not be the most pressing duplicate causer, it certainly should be examined as an entry point of dirty data simply because of the time value of acquisition.

In Conclusion

While there's no cure-all for preventing duplicates and other bad data from entering your CRM and MAP databases, there are certain practices and -with the proper budget- 3rd party applications that can help prevent duplicates from entering your CRM and MAP before their negative effects are felt.

The best approach is having strong knowledge of data quality, as well as a reputable data quality vendor to get your duplicate prevention solutions in place.

