validity

Tips for Managing Data as a Team

When leveraging a cross functional data management team, it's important to set a strong foundation so everyone feels informed and clear on the project goals.

This approach can serve as a framework for managing other data projects.

Key Steps

Steps 1, 2, and 3 should be drafted by the data team leader before the first call with the cross functional group. Steps 4, 5, and 6 will be defined by the larger team.

1. Explain the data problem/need and the impact it has on the business
2. Outline project goals
3. Familiarize the data team with best practices that relate to resolving the data problem
4. Define how the data issue is identified
5. Determine which records should be included in the data project
6. Create a consistent method for tracking each data point and how it is impacted by this data initiative

Below is an example of how this process would look for managing duplicates.

Description and Business Impact

We have constant complaints of duplicate records in Salesforce that should be merged. Duplicate records cause a multitude of problems and should be managed on an on-going basis, including putting a preventative barrier in place.

Business issues caused by duplicates:

- Decreases adoption and use of the CRM
    - Forcing sales and support to search for and reference multiple records to complete their job will result in workarounds outside of the CRM. This results in less data entering the CRM, which leads to data inaccuracies.
- Loss of the single customer view
    - A lack of a single customer view is perhaps the most detrimental aspect of duplicate data since sales, marketing, and support all rely on the same record throughout the customer's lifecycle. This also leads to using inaccurate or incomplete information when we serve our customers and prospects.
- Negative impact on email deliverability
    - When we deliver duplicate messages to the same person multiple times, we are more likely to have our messages flagged as spam.

- Skewed Campaign ROI
  - Including the same person in the same campaign multiple times ensures low Campaign ROI.
- Harms brand perception
  - Sending the same person the same campaign two or three times is annoying at best and perception-altering at worst.
- Multiple reps reaching out to the same Lead
  - When new leads come in and aren't deduped against existing Leads and Contacts, it results in multiple reps contacting the prospect. This causes confusion for the prospect and reveals our mismanagement of data.

Goals

- Stop duplicates from disrupting our users and our business
- Automate deduplication to run daily
- Implement a preventative duplicate blocking tool

Deduplication Best Practices

- Test deduplication scenarios in a sandbox environment before deploying to production. If a sandbox environment is not available, create a set of dummy records in the production environment to test the dedupe scenarios against.
- Create a backup of the data before executing a merge. Undoing a merge can be time consuming and uncertain without a recent backup readily available.
- Dedupe objects in the following order: Account; Contact; Lead; Lead to Contact; Lead to Account, Opportunity, Custom Objects.
- Per object, define the different ways records can match to surface all variations of duplicate records.
- Once the duplicate definitions are in place, organize them from rigid to loose. A rigid comparison typically matches on 4-5 fields and produces high-confidence match results. Very little spot checking is needed on these results since they match on so many different data points. Loose(er) scenarios match on less fields and may bring back false positives mixed with legitimate dupes, be mindful of which scenarios are automated and which are manually executed.
  - Ex: A rigid Account deduplication pass would match on Account Name, Street Address, City, Phone, and Website (Note: The matching algorithm used on each field also dictates how rigid or loose a scenario is.)
- Standardize address, phone, and name fields prior to deduplication. Standarization is a data best practice on its own and can help with identifying duplicates.

Duplicate Matching Criteria and Record Designation

Matching criteria will depend on the standard and custom fields used by each organization. Below are examples that can be used as a starting point and come with DemandTools. It's a good practice to also show the mapping type used to compare the data in each field.

Object	Match Level	Record Designation	Field	Mapping Type	Other Enabled Match Engines
Account	Rigid	All Accounts	Name	Cleaned Account Name	Transpose, AlphaClean
			Billing Street	Street Address Match	
			Billing City	Exact	Fuzzy
			Billing Postal Code	Zip 5 and 9	
			Phone	Numeric	
Account	Sem - Rigid	Accounts in North America	Name	Cleaned Account Name	Fuzzy, Transpose, AlphaClean
			Billing Street	Relaxed Address Match	
			Billing City	Exact	Fuzzy, AlphaClean
			Phone	Relaxed NA Phone Match	
Account	Sem - Rigid	Accounts in EMEA	Name	Cleaned Account Name	Fuzzy, Transpose, AlphaClean
			Billing Street	Relaxed Address Match	
			Billing City	Exact	Fuzzy, AlphaClean
			Phone	Numeric	
Account	Loose	All Accounts	Name	Cleaned Account Name	Fuzzy, Transpose, AlphaClean
			Billing Street	Relaxed Address Match	
			Billing City	Exact	Fuzzy, AlphaClean

Field Value Handling

- *Detail the rules that determine, by field, which values are retained when records are merged. Use a shared template that each team member will fill out to explain why certain field values need to be kept and how they are leveraged by their department. Capture things about the field like:*
    - Object
    - Field Name
    - Field Description (what data is held here)
    - Field Type
    - Key Uses
    - Departments used by
    - Related processes that impact the field
    - Values to keep during a merge and WHY
    - Other considerations

Object	Field Name	Field Description	Field Type	Key Uses	Departments Used By	Related Processes	Value to keep during a merge (list parameters)	Other considerations
Account	Account Name	The legal name of the business	Text	Reports	Sales, Service, Marketing, Finance	Updated nightly by ERP system integration	Keep the value from the record that was last modified by the ERP system integration user	
Account	DBA Name	(could match the Account Name)	Text	Campaigns	Sales, Service, Marketing, Finance	Updated by sales or service reps	Keep the value from the record that has the most closed won op	
Account	Parent Account	The legal name of the parent account	Lookup to Account	Sales Reports to identify expansion opportunities	Sales	Updated quarterly by the ERP system integration	Keep the value from the record that was last modified by the ERP system integration user	Merging will break the hierarchical structure if best practices are not followed and ParentFixer is not installed
				Expansion Campaigns	Marketing			
Account	ERP ID	The unique ID on the Account's financial record in <ERP System>	External ID	To check if there is a master service agreement from the parent entity that also applies to this subsidiary	Service			This data is pulled over to the Case page for the service team to reference

We hope this guide helps with managing your data and promoting teamwork. Please reach out to Validity if you have any questions!