Data Quality Cheat Sheet

By: Andrew Cauldwell

Index

01

Refresh

02

Property Validation

03

Sensitive Data

04

Data Quality Command Center

01 - Refresh

Data Quality Definition

Best Practices

What is *Data Quality*?

High-quality data in HubSpot is characterized by valid, consistent, and relevant properties and object records. This involves you taking proactive measures to eliminate duplicate records or properties and ensuring that the data is relevant to your business operations. Consistency is achieved across different objects (contacts, companies, deals, etc.) by standardizing the data collection process within your CRM.

Data quality is crucial because it directly impacts the accuracy and reliability of business insights, which in turn affect decision-making and strategy development. Inadequate data quality can lead to inefficiencies, wasted time, and missed opportunities due to inaccurate information and duplicate records. By ensuring high-quality data, your business can optimize operations, enhance customer relationships, and *drive growth* through informed and strategic actions.

Refresh

Best Practices

Review

Data Cleansing

- Validate email addresses
- Standardize data entry
- Remove duplicate contacts

Email Marketing Consent

Criteria*

- Submitting a form
- Purchase with company in last two years
- Filled out paper or electronic signup sheet

Questions to consider

What tools are primarily used to create records in the CRM?

Are valid email addresses used when creating contacts?

Are import spreadsheets cleansed before importing?

What pain-points need addressed regarding data collection?

HubSpot Defaults

Deduplication of records

- Contacts = Email (Primary factor) & Record ID
- Companies = Company Domain Name ((Primary factor) & Record ID
- Deals = Record ID
- Tickets = Record ID

*More here

02 - Property Validation

[Overview](#)

[Use-cases](#)

Settings > Properties > Create property > Last step of property creation screen

What is *Property Validation*?

Overview

Property rules that you as an admin can implement with the goal of maintaining consistent data input across your CRM. Users must adhere to the rules, resulting in alignment across your team.

Key points to note

- Rules not enforced for workflow or form submission input*
- Available for some HubSpot standard properties
- Available for custom text, number, and date properties
- Available rules depend on the property's field type.

Property Validation

Overview - 2

Field Type	Available Rules
Single & Multi-line text	Require unique values Set min & max character limit Restrict to numeric values Don't' allow special characters
Number	Require unique values Set min & max value – (same as character limit above but numbers only) Limit number of decimal places
Date & Datetime picker	Any date Future dates only Past dates only Specific date range Allow Monday - Friday only

Overview - 3

You can implement validation rules for some standard HubSpot properties.

The notable exception is 'Require unique values' because that rule can only be implemented during property creation.

Object	Property List
Contact	Street Address, Annual Revenue, City, Company Name, Country/Region, Email, First Name, Industry, Job Title, Last Name, Message, Salutation, State/Region, Postal Code
Company	Street Address, Street Address 2, Annual Revenue, City, Country/Region, Country/Region Code, Employee range, Company name, Number of Employees, State/Region, Postal Code
Deal	Amount, Closed Lost Reason, Closed Won Reason, Deal Name, Next step
Ticket	Ticket name

Overview - 3

You can implement validation rules for some standard HubSpot properties. The notable exception is 'Require unique values' because that rule can only be implemented during property creation.

Object	Property List
Contact	Street Address, Annual Revenue, City, Company Name, Country/Region, Email, First Name, Industry, Job Title, Last Name, Message, Salutation, State/Region, Postal Code
Company	Street Address, Street Address 2, Annual Revenue, City, Country/Region, Country/Region Code, Employee range, Company name, Number of Employees, State/Region, Postal Code
Deal	Amount, Closed Lost Reason, Closed Won Reason, Deal Name, Next step
Ticket	Ticket name

Use cases

Goals	Description
Unique ID deduplication criteria	You need an unique identifier as an alternative to HubSpot default properties (Email & Company Domain Name). You can create a text property and require unique values for use in importing and deduplicating records that may not have an email address or other identifying property.
Keep name properties tidy	You need to ensure first name, last name, and other name text properties are text only without special characters. People on your team type fast and mispress keys resulting in names like ")luto"(Pluto). You can validate your properties by not allowing special characters.
Email automations sent during business days only	Your marketing team tracks outreach dates during business days on the contact record. They want to limit user-error and make sure date values are only during Monday - Friday. You create a custom date property that only allows Monday - Friday dates. The marketing team then uses that property for their drip campaign automation and prevents emails from going out on weekends.

03 - Sensitive Data

Overview

Use-cases

What is *Sensitive Data*?

Overview

It is confidential personal information that requires an extra layer of protection by law. Areas include personal sensitive data, financial data, and health/medical data. Enabling sensitive data options in your account will allow you to create properties to store and use this information in the CRM.

Key points to note

- Sensitive Data is in Public Beta
- Highly Sensitive Data (HIPAA & Financial) is in Private Beta
- Sensitive data is available in these areas*:
- CRM properties, activities, and attachments (including API)
- Lists
- Workflows
- Search
- Reports
- Integrations
- Forms

*Not an exhaustive list. More details here

Overview - 2

Navigate to **Settings > Privacy & Consent**
Click **Configure sensitive data settings**
In the setup wizard choose the types of data you will be working with.

Use cases

Organization	Business Need
Education	You need to store faculty and student information that includes sensitive information such as demographic data, gender and citizenship/immigration status.
Finance	You are a bank and need to store financial information, income, demographic, and employment history data to process credit or bank account applications.
Healthcare	You are non-for-profit clinic and need to store HIPAA data in addition to non-sensitive data you already collect from your patients. Now you can store both sensitive and non-sensitive data in the same platform.

04 - Data Quality Command Center

[Overview](#)

[Duplicate Manager](#)

[Use-cases](#)

CRM > Contacts > Data Quality (button located top right of page)
Data Management > Data Quality

What is Data Quality Command Center?

Overview

It is an overview tool that displays how data is being used in your CRM and highlights potential issues. These include properties, records, data sync integrations, and workflows. You can see things such as duplicate records or properties, formatting suggestions, and unused properties to name a few. You can click into more detail for most cards*.

Key points to note

- Data Quality Command Center requires Operations Hub Professional+ for some features
- The Duplicate Manager tool is available with all plans
- Duplicate suggestion criteria cannot be modified
- Duplicate scans occur daily
- *Workflows data is only viewable from the Command Center
- Workflow data includes: Unused and Needs review suggestions

Overview - 2

Tools	Abilities
Property Insights	Property scans occur daily at 8pm in your account time zone. View total number of properties across standard objects. Property suggestions: no data, unused, and potential duplicates. View update create and update dates, field type, source of property population (e.g. CRM user), assets using the property, and fill rate (how many records in the CRM have a value)
Data Sync Insights	Only applicable if you have a data sync app installed (e.g. Shopify). View apps with zero active syncs or with sync failures.
Data Monitoring	Notifications for duplicate record thresholds and property anomalies. Duplicate records apply for the Contact and Company object, and you can set the limit threshold before a notification is triggered to the selected users. Property anomalies apply to all standard objects and is determined by an AI model of the last 30 days of your CRM data. If a property is filled outside of the trend for a given day, then it will be marked as an anomaly (e.g. Trend = 5-10 fills per day – suddenly 15 fills on a day = anomaly).

Data Quality Command Center

Overview - 3

Tool	Abilities	Key Points to Note
Formatting Issues	Check property values for potential formatting issues. Categories include: Name, Email, and Location. Location properties are: Country, Timezone, State/Region, and City. Evaluation criteria includes*: value contains all capital letters, value contains unexpected capitalization (e.g. jAnE), value is not a valid email address, value contains common placeholder text (e.g. n/a). Automations available: capitalize first name/last name, make email lower case, remove last name from first name, spacing and punctuation (first/last name & email)	● Only available for Contacts and Companies objects ● The suggestions are generated by AI (may not be 100% accurate) ● Requires Operations Hub Starter ● Automation tools within requires Operations Hub Professional ● Format preferences for Country and State/Region properties: Settings > Objects > Contacts

Data Quality Command Center

Duplicate Manager

CRM > Contacts > Actions > Manage Duplicates

Abilities	Key Points to Note
Available to all Review duplicate contacts and companies in a table Display up to 2,000 duplicate pairs at one time Contact criteria: Uses First & Last Name, Email, IP country, Phone number, Zip Code, and Company Name Company criteria: Company Domain Name, Company Name, Country/Region, Phone Number, and Industry Merge criteria options: Most recent / Oldest engagement, Created first /last, Most recently updated **Operations Professional** Display up to 5,000 duplicate pairs Bulk manage duplicate suggestions **Operations Enterprise** Display up to 10,000 duplicate pairs	● Scans occur <u>once</u> daily ● Merging duplicate companies <u>is not</u> available when the Salesforce Integration is installed ● **Public Beta***: Search and Filter for duplicates (e.g. Owner, Create date, Last activity date)

Use-cases

Goals	Description
Automate property format	You want to ensure names and email values are capitalized and do not contain unrelated characters. You decide to enable the Formatting Issues automations to save time building workflows or manual proofreading work.
Audit property usage rate	You are running a test with your sales reps. You have created a few custom properties for detailed reporting. You want to know the fill rate, where they are being used, and how often the reps are inputting values. You decide to use the Property Insights tool to audit and make necessary changes.
Notifications on data quality	Yourself and another admin want to be notified when 10 contact and 5 company duplicates are surpassed in a 24hr period. You use the Data Monitoring tool to specify the duplicate threshold and add yourself and the user to the notification settings.
Manage duplicates	You have several contact duplicates and want to review them in one area. You may also need to bulk merge or reject some of the suggestions. You use the Duplicate Manager tool to view in a table format possible duplicates, select a few in the table and bulk merge them based on 'first created' criteria. You know that the oldest one is the correct record because it was created by you.

Additional References

[Opt-in consent for marketing email](#)

[Deduplication of records](#)

[Data cleansing](#)

[Data quality resources](#)

[Operations hub introduction](#)

Property Validation

Forms validation beta

Additional Use-cases

Your company tracks a certain number down to the hundredth decimal place for accuracy purposes. Secondly, the number should not have more than 4 numbers before the decimal (e.g. 1000). In order to maintain data quality, you implement the limit number of decimal places and the max value rules.

Your marketing team is running a campaign for a Q4 holiday promotion. The team wants to track form submissions during the Q4 period only. You create a custom date property and set the specific date range rule to the team's desired Q4 range. This prevents any dates from being added outside of the range, resulting in clean data.

Sensitive Data

Sensitive Data Legal Terms

Trust Center

- Table of covered services for sensitive and highly sensitive data listed in legal document above

- Requires any HubSpot Enterprise subscription

- Only Super Admins can setup and view Sensitive data

- Sensitive Data is not used to train AI models

Appendix

Data Quality Command Center

[Formatting issues criteria linked here](#)

Additional use-case

You want to ensure personalization tokens in emails and content are formatted correctly, so you implement formatting rule automations. This results in fewer marketing emails being addressed with names starting with a lowercase letter.

Duplicate Manager

Quick tip: Enroll your account in the [duplicate search beta](#) by clicking your Account Name (in the top right corner of HubSpot) > Product updates > Betas tab.

[Further reading](#) about the tool