

SOLUTION ENABLEMENT GUIDE

CUSTOMER MASTER DATA MANAGEMENT

The logo for StiboSystems, featuring the word "StiboSystems" in a white, sans-serif font. The letter "i" in "Stibo" has a small crown-like symbol above it. The logo is positioned on the left side of the page, within a large orange triangle that points to the right.

StiboSystems

STEP Trailblazer 8.3

Table of Contents

Table of Contents	2
Customer Master Data Management Solution Enablement	6
Customer Master Data Management Overview	7
Improving Customer Experience	7
Customer Data Management Challenges	7
Data Quality Challenges	7
Operational Inefficiencies	8
Reactive Strategy Decisions	8
Governance	8
Implementation Methodologies	8
Consolidation Implementation	8
Coexistence Implementation	9
Registry Implementation	10
Centralized Implementation	10
Customer Master Data Management Implementation Guidelines	12
Use Case Overview	13
ACME Company's Business to Customer Use Case	13
Considerations	13
Difficulties and Problems	13
ACME Company's Business to Business Use Case	14
Considerations	14
Difficulties and Problems	14

Data Management Improvement Plan	14
Use Case Appendix	16
Business to Customer Use Cases	16
Business to Business Use Cases	16
Use Cases - Merging Individual Records ...	17
Auto-Merge Individual Records	17
Merge Individual Records	17
Advanced Merge Individual Records	17
Use Cases - Marriage and Relocation	19
Marriage	19
Relocation	19
Use Cases - Trigger Clerical Review after Manually Editing Customer Record	20
Use Cases - Address Validation	21
Use Cases - Rejecting Potential Duplicates	22
Use Cases - Managing B2B Contacts	23
Identifying and Deduplicating Contacts Within the Same Organization	23
Handling Contacts Across Different Organizations	23
CMDM Data Flow	24
Data Flow Methodology	24
Consolidation Example - Use Case 1: A new customer via the website	27
Consolidation Example - Use Case 2: Updating a Customer over phone in ERP ..	28
Coexistence Example - Use Case 1: Updating the Source Systems	28
Matching	29
Matching Algorithm Configuration	29

- Match Codes Configuration29
- Example Match Codes 30
- Email30
- Phone30
- Person Name (First, Middle, and Last) ... 30
- Address30
- Name and Address31
- Starter Package Match Codes31
- Matching Algorithm Tuning31
- Algorithm Tuning Process32
- Other Considerations 33
- Working with Customer Data 33
- Algorithm & Match Codes - Individual Customer 35**
- Matching Algorithm 35
- Normalizers 35
- Person Name Normalizer 35
- Address Normalizer 35
- Email Normalizer 35
- Phone Normalizer 35
- Matchers35
- Person Name Matcher36
- Address Matcher36
- Email Matcher36
- Phone Matcher36
- Rules 36
- Match Codes37
- Email Match Code37

- Phone Number Match Code 37
- Address and Name Match Code37
- Configuration Considerations38
- Algorithm & Match Codes - Organizational Customer 39**
- Matching Algorithm 39
- Normalizers 39
- Words Normalizer 39
- Address Normalizer 39
- Matchers 39
- Words Matcher39
- Address Matcher39
- Rules 40
- Match Codes40
- Email Match Code40
- Phone Number Match Code 40
- Organization Address and Name Match Code40
- Algorithm & Match Codes - Contact Person 41**
- Matching Algorithm 41
- Normalizers 41
- Person Name Normalizer41
- Address Normalizer 41
- Email Normalizer 41
- Phone Normalizer 41
- Matchers41
- Person Name Matcher42
- Address Matcher42

Email Matcher	42	Attribute Setup	49
Phone Matcher	42	Domain Folders	50
Rules	42	Entity Super Types	51
Match Codes	43	Important Considerations for CMDM Data Modeling	51
Email Match Code	43	Customer Hierarchies	52
Phone Number Match Code	43	Web UI Configurations for CMDM	53
Address and Name Match Code	43	Role of the Data Stewards	53
Configuration Considerations	44	Clerical Review	53
Algorithm & Match Codes - Household	45	Manual Maintenance	54
Matching Algorithm	45	Data Monitoring	55
Normalizers	45	KPI Widgets	56
Last Name Normalizer	45	Collections	58
Address Normalizer	45	STEP Work Flow	62
Matchers	45	Extra Considerations	62
Name Matcher	45	Best Practices and Considerations when Configuring Web UI	62
Address Matcher	45	CMDM Configuration Guide	63
Rules	46	STEP System Components	63
Match Codes	46	STEPXML Configuration Files	63
Loqate	47	STEPXML Import Sequence	64
Validate on Import	47	1 - Shared Config Properties File	65
Validate on Survivorship	47	2 - Link, Reference, and Object Types	65
Validate Updates to Existing Records	47	3 - List of Values	65
Data Modeling	48	4 - Attributes and Attribute Groups	65
Customer Object Types	48	5 - Link, Reference, and Object Types	66
What is a Customer?	48	6 - Setup Groups	66
What is a Household?	49	7 - Setup Entities	66
How are Addresses Structured?	49		

- 8 - Data Containers 66
- 9 - Entity Root Nodes 66
- 10 - Classifications 66
- 11 - Business Rules and Libraries 66
- 12 - Component Models 66
- 13 - Import Configurations 66
- 14 - Transformation Lookup Tables 66
- 15 - Match Codes 67
- 16 - Match Algorithms 67
- 17 - Integration Endpoints 67
- 18 - Users and User Groups 67
- 19 - Event Processors 67
- 20 - Collection Definitions 67
- 21 - Profile 67
 - All Individual Collection 67
 - All Objects Collection 68
 - Recently Updated Contact Persons
Collection 68
 - Recently Updated Households Collection 68
 - Recently Updated Organizations
Collection 69
 - Recently Updated Individuals Collection . 69
- 22 - Workflows 69
- 23 - Web UI 70
- 24 - Sample Data Load Files 70
- CMDM Starter Package Configuration
Delivery Mechanism 70

Customer Master Data Management Solution Enablement

Customer MDM (CMDM) is a technology enabled discipline which is required to create and maintain an accurate, timely, complete, and unified view of a company or customer. CMDM will consolidate, cleanse, enhance, and govern customer data from disparate siloed systems enabling organization to identify, consolidate, and link customer data across heterogeneous data sources in a single, accessible customer view.

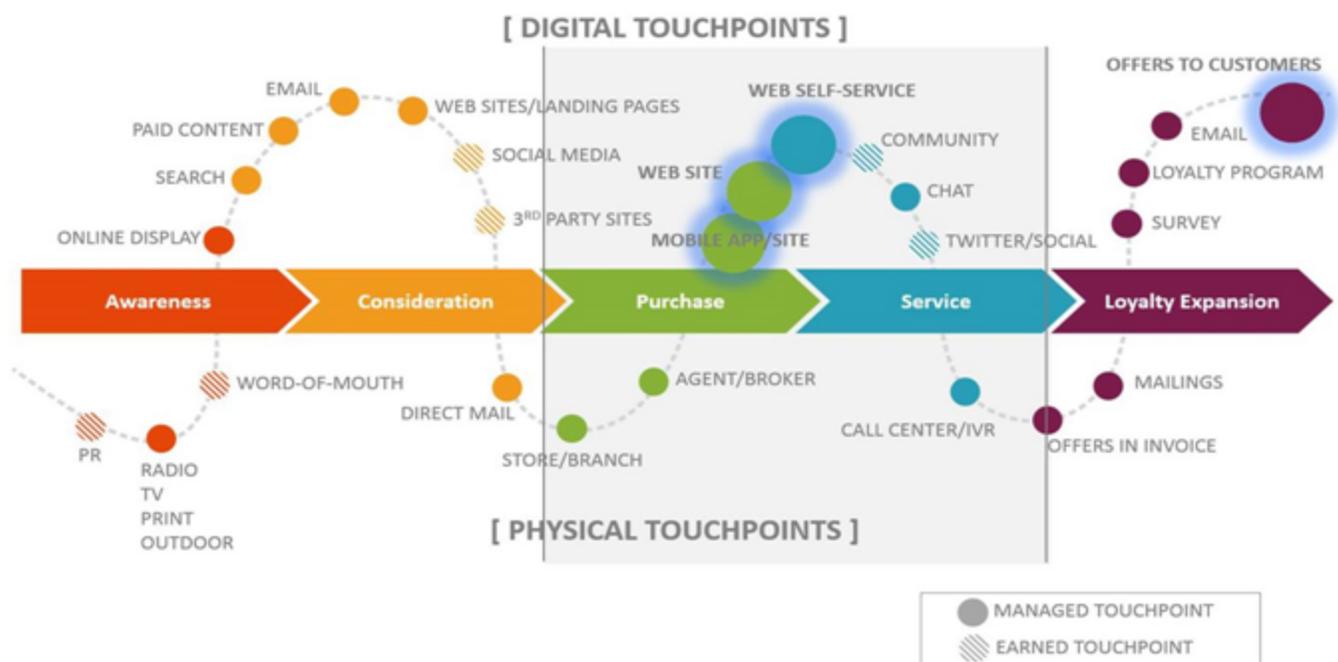
Customer Master Data Management Overview

The Customer Master Data Management solution provides organizations with the means to address complex customer data challenges and improve their overall customer experience.

Improving Customer Experience

Customer Experience is the holistic journey of a customer that begins the moment they first become aware of a product and continues until the customer is building their own loyalty platform with a product.

This documentation is focused on providing guidance to organizations in how to improve their customer data so that they can make informed decisions, and thus improve their customer experience.



Customer Data Management Challenges

Organizations are faced with four primary challenges when managing customer data:

- Data Quality
- Operational Inefficiencies
- Reactive Strategic Decisions
- Governance

Data Quality Challenges

Data quality issues typically arise when attempting to house and manage customer data across multiple platforms. Doing so often times leads to inaccurate, duplicated, and incomplete data.

Operational Inefficiencies

Operational inefficiencies can be characterized by an organization's inability to share accurate and up-to-date data across the business, severely impacting downstream systems, processes, and initiatives. For example, some of these inefficiencies can lead to: returned mail, increased costs, payment delays, and unacceptable response time.

Reactive Strategy Decisions

Many organizations struggle to achieve their Customer Relationship Management (CRM) vision. Considering its impact on an organization's success, a concrete CRM strategy is a high concern for business and IT leaders.

Bad data is the most common reason many CRM initiatives fail.

Governance

The most important part of data governance is to consider the availability, usability, integrity, and security of data within an organization. Poor data governance can lead to financial, reputational, and legal risks.

Implementation Methodologies

CMDM comprises four paths to implementation. These paths are:

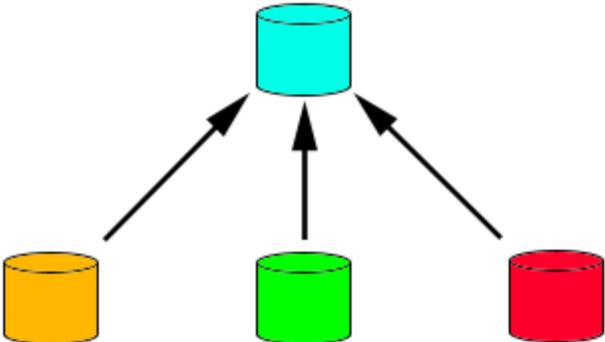
- Consolidation Implementation
- Coexistence Implementation
- Registry Implementation
- Centralized Implementation

Consolidation Implementation

This methodology refers to an implementation where STEP is a hub for externally provided data that is funneled into single Golden Records, which are then made available to external systems. These records are created via matching and merging operations with validation checks throughout the process. The STEP data created from this process is read-only, which is often a necessity as the data originates outside of STEP in an upstream system. A consolidation style implementation does not synchronize data back to the contributing systems.

For example, a company wishes to combine all of their sales channel data into one record bank so that data analytics may provide meaningful information. Once all these records are consolidated, marketing initiatives can use this information to target different channels as well as different regions.

Consolidation

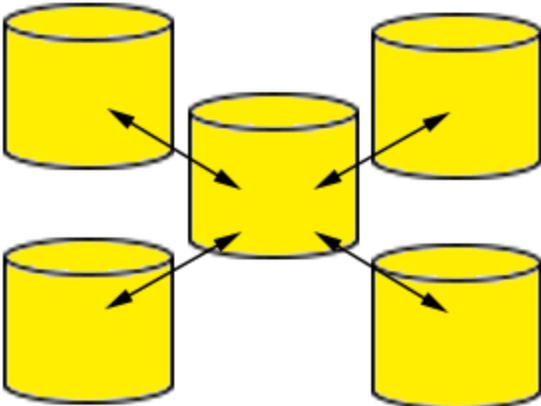


Coexistence Implementation

This methodology refers to a scenario where multiple databases containing the master data must operate at the same time. The customer data is authored and stored in various external sources while being synchronized to STEP. This process involves deduplication, conflict resolution, and validation operations, resulting in the creation of golden records. Finally, updates to the data can be done in the source system and any external systems that master customer data. STEP synchronizes the content of golden records back to the source systems, keeping every source up to date with data contributed from any source. This allows for a hybrid where DTA authoring may happen in the source system as well as in the MDM system.

A common situation for coexistence is in companies where each department has their own customer database. For example, a service department that has data on customers is stored in a separate database than where customer data is stored for the sales and marketing departments. However, using a CMDM solution with the Coexistence model, if information is changed in the marketing department, it is changed in all other departments.

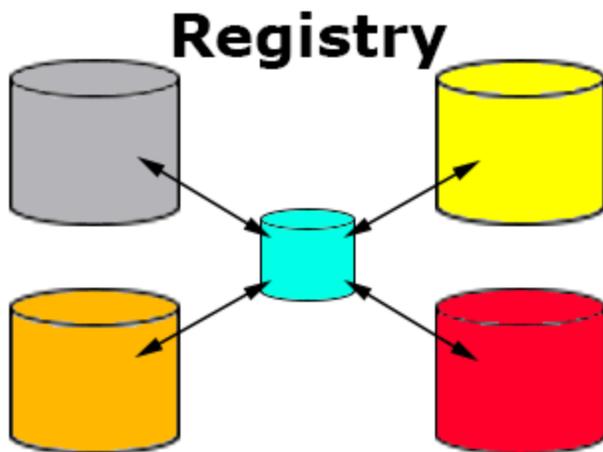
Coexistence



Registry Implementation

The registry method uses a simple database to reconcile identifiers for all customers across the enterprise. This method allows STEP to serve as a referenced, read-only source of mastered customer data for external systems with minimal data redundancy. STEP will only match on the bare minimum of needed data to confirm a unique record. All externally maintained data quality is the responsibility of the source systems.

For example, a company maintains different types of customer information across various external systems. A registry is used to identify and associate these different types of customer data with external systems. STEP serves as this registry by storing unique records with identifiers that are different from one external system to the next. STEP will ONLY store these records and unique record identifiers, and deduplicate their contact information.

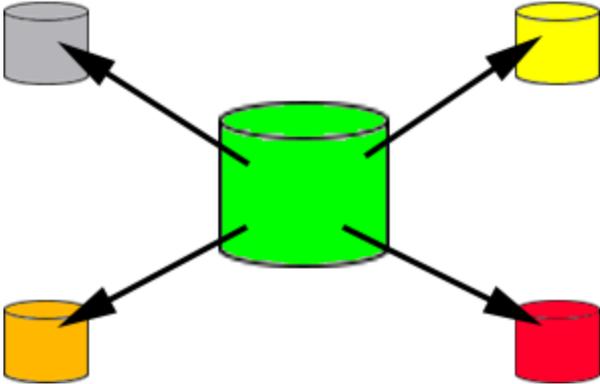


Centralized Implementation

The centralized implementation is based around a well-managed and governed central repository for all master data. This repository holds a set of Golden Records that are read-only by the operational and analytical systems throughout the enterprise. STEP is the centralized owner of customer data, serving as the system of record. This system of record is an organization's single version of truth for customer data. Data is not only centralized in STEP but also created and stored there. The most important aspect of a centralized implementation is that it is a system of record instead of a system of reference.

For example, a company wants to use STEP as the system of record for onboarding, deduplicating, and enriching new suppliers. Over time, STEP could remove duplicate records of these suppliers, and then the golden supplier records could be provided to external systems.

Centralized



This implementation guidelines focuses with a case that implements data with the **consolidation** and the **coexistence** methodologies.

Customer Master Data Management Implementation Guidelines

The Customer Master Data Management Implementation Guidelines provide organizations with baseline configurations and sample data to help get their STEP Systems off the ground and guide them in the creation of their Customer MDM solution.

This user guide offers a number of resources that can help achieve this goal:

- A generic Customer MDM use case that accounts for both B2C and B2B solutions, and provides context for system configuration and solution design. For more information, see the **Use Case Overview** section of this user guide.
- A series of component-specific use cases that tie into the generic Customer MDM use case mentioned above. For more information, see the **Use Case Appendix** section of this user guide.
- A comprehensive guide for how data flows in and out of the STEP system, including Matching. For more information, see the **CMDM Data Flow** section and the **Matching** section of this user guide.
- A comprehensive guide for modeling the data in STEP. For more information, see the **Data Modeling** section of this user guide.
- Configuration examples for Customer MDM Web UI components. For more information, see the **Web UI Configuration for CMDM** section of this user guide.
- For the configuration of a base CMDM solution and how to deploy it, see the **CMDM Configuration Guide** section of this user guide.

Important: This guide references supporting configuration documentation that exists in other sections of the STEP Online Help.

Use Case Overview

To illustrate how the CMDM solution can address common challenges an organization faces, the following use case presents a fictional company with typical business considerations and problems. Understanding this use case example will provide additional context to the configuration steps included in this guide, as well as the more granular use cases that concern specific CMDM components.

For more information on component-specific use cases, see the **Use Case Appendix** section of this guide.

ACME Company's Business to Customer Use Case

The ACME Company is a large clothing brand which markets and sells largely to North America. ACME sells directly to their consumers via various channels. As the company has grown, issues have developed due to poor data quality. Because of this, customer contact information is rarely in sync across business units. Furthermore, all customer shopping patterns are inaccurate because the data is not distributed and updated evenly. This asynchronization causes a missed opportunity to target customers for the best sales, promotions, and loyalty programs. Finally, due to these gaps in the data, the customer experience is quite poor as actions performed in one department by a customer are not known in another.

During an internal evaluation of the IT infrastructure, it was revealed that the customer data is owned and maintained by separate and distinct source systems. This setup causes data to not be updated across the enterprise, creating a large amount of conflicting duplicate customer records. Additionally, it was concluded that a lack of data governance has made it impossible to determine what degree of overlap and duplication is present within the customer database.

To improve customer data quality, ACME has targeted a few areas where data quality becomes less consistent:

- ACME's e-commerce system processing customers who self-register on the company website.
- Customers placing phone orders to a call center representative who then enters their data into ACME's ERP system.
- Customers signing up for the loyalty platform.
- Marketing uses customer data residing in ACME's CRM system for seasonal marketing campaigns.
- ACME uses loyalty points which can be redeemed for discounts and contain a monetary value. The financial data is owned by the CRM system which stores additional financial information.

Considerations

- North America is the primary market.
- ACME sells directly to customers via various channels.
- Multiple different source systems contain customer data.
- Currently, the customer data violates ACME's own data policies.

Difficulties and Problems

- Contact information not in sync across business units.
- Customer shopping patterns are of poor quality.

- Very poor user experience due to out of sync data across the business units.
- There is an excessive amount of duplicated customer records across the business.

ACME Company's Business to Business Use Case

ACME Company is a major clothing brand that not only sells directly to individual consumers, but also to both major and independent retailers and distributors across North America. To manage these various retailers, North America is divided up into several sales territories which are each managed by their own sales managers. Each sales territory contains several retailers, and information for each territory is managed by different instances of CRM. Since the same retailers may operate out of various different states or sales territories, ACME Company requires CMDM to not only deduplicate retail entities but also provide a trusted source of information for various source systems.

ACME also tracks and maintains contact persons' information for each retailer. These contact persons are designated sales representatives, account owners at each retailer, and purchasers in the procurement departments. The CMDM system must be able to account for contact persons that move from retailer to retailer, or those that get duplicated within the same organization, to ensure that accurate information is maintained.

Sales hierarchies are used to organize the customer portfolio in the sales staff and to aggregate bonus schemes in the sales staff, based on the revenue from each customer.

Considerations

- ACME sells to both major and independent retailers and distributors as well as to consumers.
- The company mostly services clients in North America.
- The North American region is divided into several sales territories with sales managers responsible for each territory.
- Contact persons are maintained in CRM as well as the ERP applications.

Difficulties and Problems

- There is an excessive amount of duplicate customer records.

Data Management Improvement Plan

The ACME Company has strategically planned the evolution of how their customer data will mature over the course of the next few years. To accomplish this long-term goal, an enterprise strategy plan is laid out in three phases, with each phase progressively building on top of the previous one.

Phase 1 – ACME would like to ensure their customer data is governed and can be trusted across the entire organization:

- Use MDM technologies to achieve a single, trusted customer view which is accurate, complete, and up to date.
- Establish baselines and improve analytical reporting accuracy for a better CRM and e-commerce experience.

- Identify customers who live under the same household for more efficient mail distribution and for better profiling of the customer.
- Reduce operational overhead cost by decreasing the amount of returned mail due to inaccurate mailing addresses.

Phase 2 – ACME would like to combine core customer master data with transactions, web site activity and social profiles:

- Enhance the customer experience by providing a personalized shopping experience.
- Utilize purchasing behavior analytics for increasing cross sell / up sell opportunities.
- Discover hidden associates and relationships to gather context for segmentation, targeted marketing, risk management, and regulatory reporting.

Phase 3 – ACME would like to further enhance the personalized shopping experience:

- Create an enterprise-wide single view of customers to serve all operational systems and use big data to generate insight and foresight based on advanced learning models.
- Contextualized customer data is then used to enhance customer experience through machine learning as well as digitally automating key business processes.

Use Case Appendix

The following use cases address a number of common challenges organizations face when handling customer data, and how they are solved by a CMDM solution.

Business to Customer Use Cases

- Merge Individual Customer Records via Automatic, Manual, or Advanced merging methods
- Rejecting Potential Duplicates (during Clerical Review)
- Customer Marriage and Relocation
- Trigger Clerical Review after Manually Editing Customer Record
- Address Validation

Business to Business Use Cases

- Identifying and Deduplicating Contacts Within the Same Organization and Handling Contacts across Different Organizations

Use Cases - Merging Individual Records

In these use cases the data steward is responsible for overseeing the deduplication of customer records via different merge processes.

Auto-Merge Individual Records

ACME has several external source systems which contain multiple records with separate identifiers (source record IDs). Despite these different identifiers, many of these records are in fact the same individual but with minor variations in demographic data.

These individual records are sent to STEP and automatically deduplicated via the configured matching logic, which identifies and merges identical records. These matched records are then merged and a single golden record is created.

Original source systems and respective source record IDs are maintained as part of the golden record for source traceability purposes.

Merge Individual Records

ACME has a number of records which are considered potential duplicates of one another, but the system cannot ultimately determine whether or not they are true duplicates without the input of a data steward. In these cases, a data steward must review potential duplicate records and corresponding demographics to intelligently determine whether they are a match or not.

Once STEP determines which records are potential duplicates of one another, a Clerical Review task list is presented to the data steward in the Web UI. Data stewards may review the demographics of each task to determine whether the records should be merged. If it is determined that the tasks are a match then the **Merge** action will merge the records into a single golden record. If it is determined that the records are not matches of each other then the **Reject** action will leave the records unmerged.

For matched records, a single resulting golden record will survive that contains the source system and source record ID of all records that were merged. For non-matched records, a Confirmed Non-Duplicate reference will be established to prevent future clerical review tasks from being created for the same records. The existing task will also be removed from the Clerical Review task list / workflow.

Advanced Merge Individual Records

As an alternative to the **Merge** function described in the above use case, data stewards can use the **Advanced Merge** option to manually dictate which demographics survive the merge between two or more potential duplicates. In addition they can choose to exclude individual records from the merge.

Once STEP determines which records are potential duplicates of one another, a Clerical Review task list is presented to the data steward in the Web UI. Data stewards may select a task, and by clicking **Advanced Merge**, they are presented with a table view of the associating records as well as corresponding demographics. As the data steward selects which demographics will survive the merge and be promoted to the golden record, a preview column displays the resulting merged golden record. Additionally, should the data steward determine that one or more records should not be considered as a potential duplicate, the option to **Exclude Record** is provided.

The resulting golden record reflects the demographics manually selected by the data steward. The task is removed from the Clerical Review task list once the merge is complete.

Use Cases - Marriage and Relocation

In these use cases, the data steward is responsible for overseeing customer record updates in the event that two customers get married or relocate.

Marriage

ACME must update their database in the event that two customers get married. A pre-existing customer named Catherine Yu got married to another pre-existing customer, Charles Johnson, and had her last name and address changed.

In this scenario, the change in status is initially updated in a source system by a customer service representative within a call center. The source system feeds this update to STEP, where it is then matched to an existing record via the source system record ID for Catherine Yu.

Once matched and updated, Catherine Yu is now Catherine Johnson and has a new address. Additionally, the household is updated to reflect that Catherine and Charles belong to the same household.

Relocation

ACME must update their database in the event that a customer moves to another address. In this scenario, Maxie Hadley was a pre-existing customer who moved from KY to UT. At some point a second record was created with the new address. STEP automatically identifies that the two records with different address are the same person.

Once flagged as potential duplicates, a data steward can handle the deduplication task via a clerical review. Concluding that they are in fact the same person, the data steward merges the two records into a single golden record.

Use Cases - Trigger Clerical Review after Manually Editing Customer Record

Sometimes ACME requires customer records to be manually edited in STEP to address errors in data quality. In this scenario, STEP identifies a new potential duplicate after the customer record for Kimberley Kaine had the spelling of her first name corrected. Once identified as a potential duplicate, a task is created in a clerical review workflow.

The data steward concludes that they are duplicate records and chooses to merge them into a single golden record.

Use Cases - Address Validation

ACME requires that address be validated via Loqate whenever a customer record is updated. In this scenario, Jason Kirby has moved to a new address. To account for this change, a data steward navigates to the customer record and manually updates the data. A Loqate request is sent, and new standardized data is returned based on the Jason Kirby's new address.

Typically, this kind of update would come through via a source system, in which case it would have already been validated upon import into STEP. In this scenario, there are no actions required by a data steward.

Use Cases - Rejecting Potential Duplicates

ACME may want to keep two matched records as separate records under certain circumstances. In such a scenario, a data steward can determine if two or more records in a clerical review workflow should merge or stay as separate entities. As an example, Jonathan Dorthy & Jenkin Doorthy are different people but were flagged as potential duplicates due to their similar names.

These potential duplicates are displayed via a clerical review task list in Web UI, where the data steward can reject the task to keep Jonathan Dorthy & Jenkin Doorthy from merging. Once rejected, the customer records remain as separate records.

Use Cases - Managing B2B Contacts

In these use cases, the data steward is responsible for managing organizational contacts (i.e. for vendors, distributors, etc.)

Identifying and Deduplicating Contacts Within the Same Organization

ACME must ensure that there are no duplicate contact records within the same customer organization. In this scenario, Sergio Bennett and Serge Bennet belong to the same organization and are identified as potential duplicates.

A source system sends a second record for Sergio Bennett with a different source record ID. STEP evaluates the incoming record and determines that it is in fact the same contact person for that specific vendor. Once STEP has determined that Sergio Bennett and Serge Bennet are one and the same, a data steward can merge them into one record via clerical review.

Handling Contacts Across Different Organizations

Emma Hoffman was previously a contact person for one of their suppliers, but now, works for another one ACME is associated with. Since the organizations are different, when the source system updates STEP with the new organization contact, Emma Hoffman's contact information will not be deduplicated across both organizations.

A new golden record contact for Emma Hoffman is created for that supplier and it is not deduplicated against her previously created golden record.

CMDM Data Flow

Having customer data authored and stored across separate source systems creates a major barrier to achieve a unified and trusted view of customer data. When data has multiple sources, it can become inconsistent, cause conflicts, inaccurate, and outdated.

Per the Use Case Overview, ACME Company has multiple source systems from which customer data originates. The flow of customer data for ACME Company begins at these points of entry which are associated with the various channels, such as company website, phone direct, in-store, etc. (see Figure 1 below). ACME contains four source systems (Figure 2):

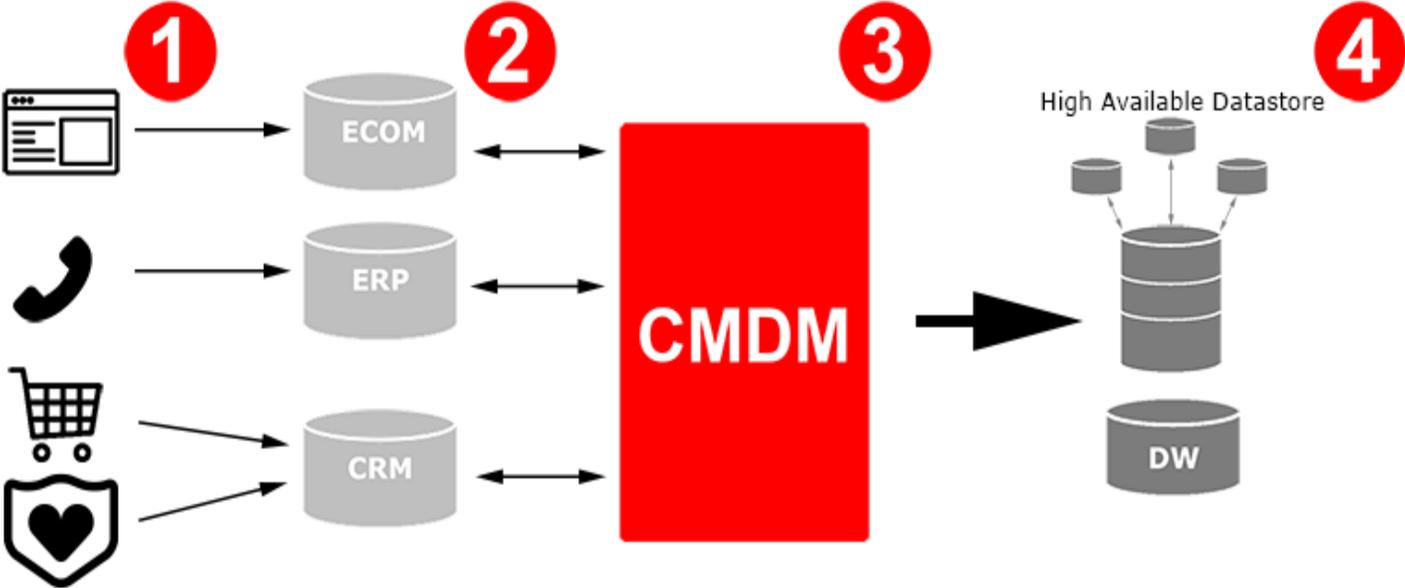
- E-commerce system for customers who place their orders and self-registers via the company's website.
- ERP system for customers who place orders by phone via ACME Company's call center.
- Customers who sign up for loyalty program on brand websites.
- CRM system for driving seasonal marketing campaigns as well as financial data which drives in-house financing.

Data Flow Methodology

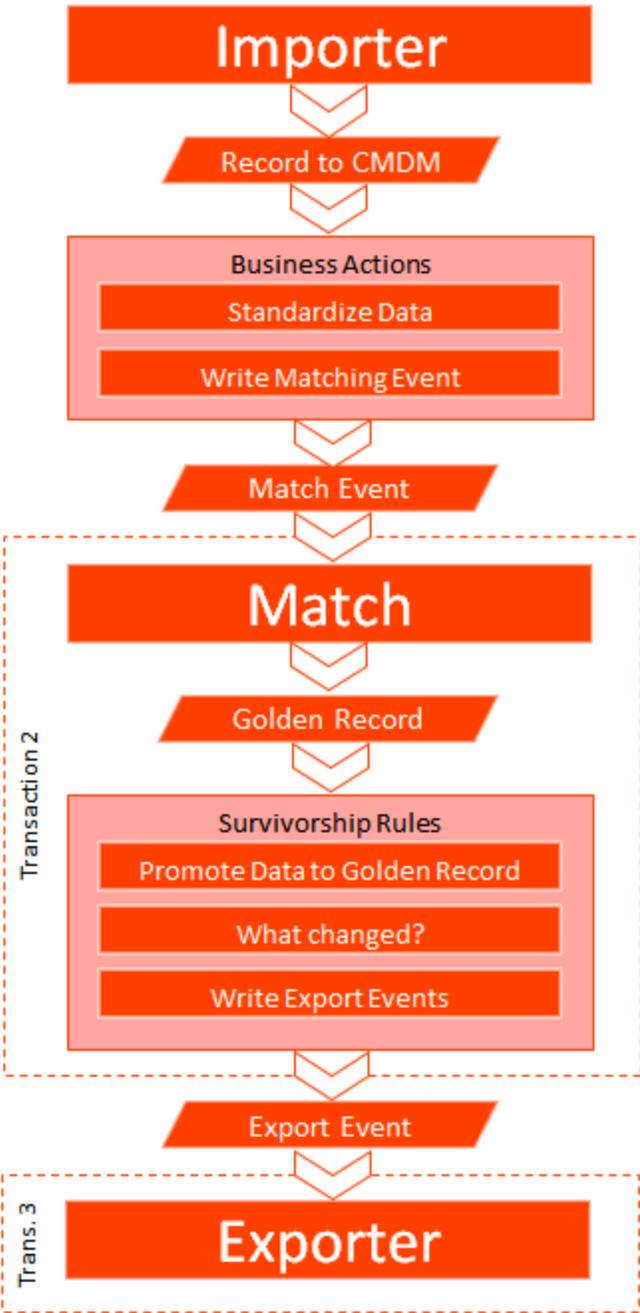
To understand the data flow into STEP and between applications, look at the two implementation methodologies: consolidation and coexistence. See the **Customer Master Data Management Overview** topic for information of how the different implementation methodologies function.

Data enters into STEP through external channels where customers will provide their information. In section 1 of the image below, these channels may be through a telephone order, a web order, when they sign up for a loyalty platform, or when they buy in store.

After their data is recorded, it is saved to various source systems. A **source system** is any system that stores data to be delivered to the CMDM solution. In this use case, the source systems are the CRM, ERP, and E-comm platforms (Figure 2). Once the data is inside of STEP's CMDM solution, the various applications in STEP will reform the data to make it standard, match it with possible other records, and create golden records. For more of what happens in Figure 3, see the Application path image following this section. Also, see the **Matching** topic and the **Loqate** topic in this CMDM documentation. Data will then be stored in a datastore / data warehouse. In a coexistence implementation, the data matched and merged in the CMDM solution (Figure 3) will be updated back at the source systems (Figure 2) as well as downstream solutions (Figure 4).



In Figure 3, as Data is imported into STEP, it will undergo a few operations before being made useable.



Through an Inbound Integration Endpoint, data enters the CMDM solution. Business actions will clean the data for matching. Once matching begins, the matching algorithm will check as to whether similar records exist. If a match is found above a set threshold, the records are merged into a golden records. If it is below a threshold, it is either made a golden record because it is unique or it enters a Clerical Review Work Flow. During import and during matching, addresses could be validated using Loqate. For more information on the match & merge importer, see the **IIEP - Configure Match and Merge Importer** topic in the **Data Exchange** documentation.

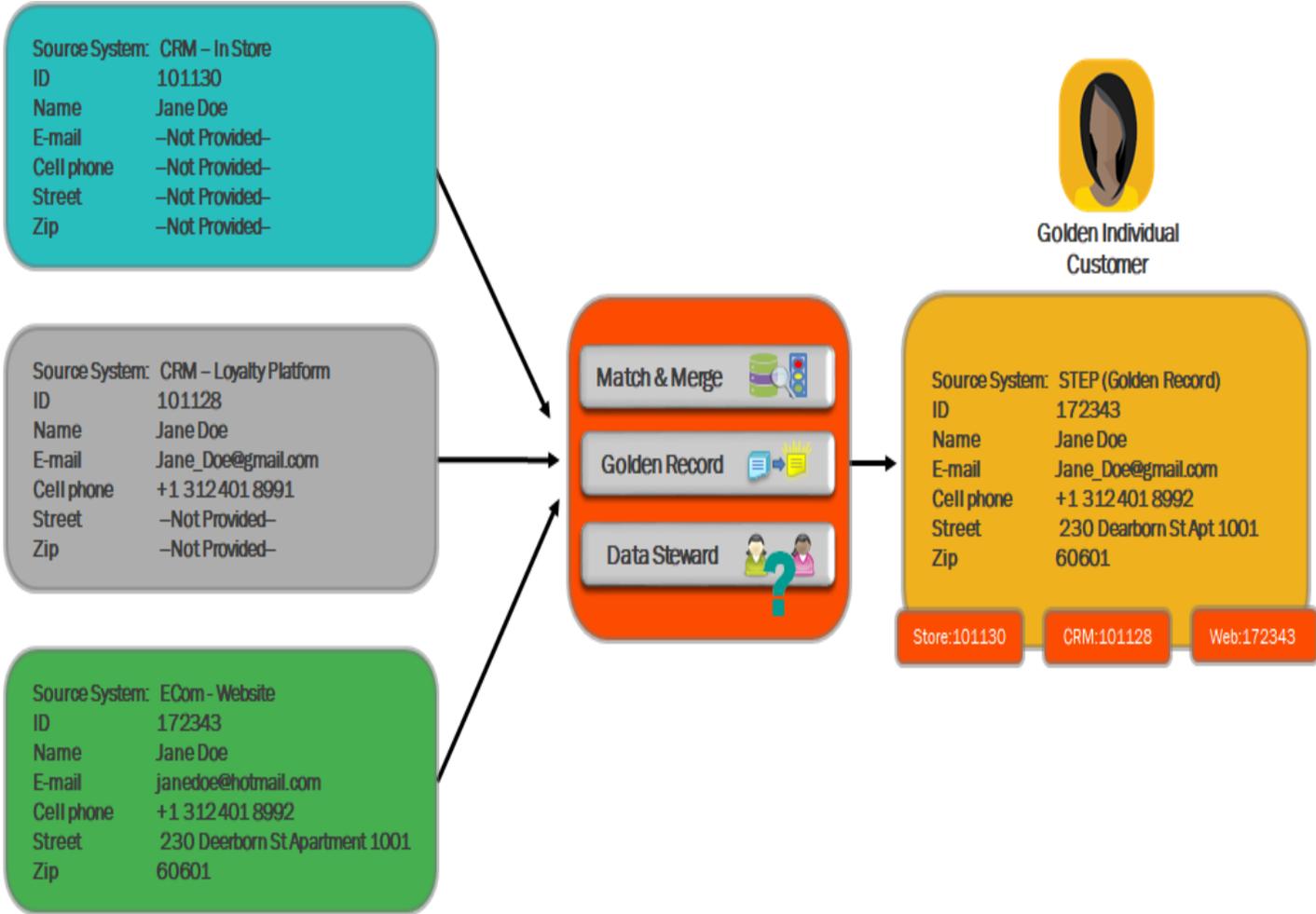
The following examples show two ways into how data moves to and from STEP using two different implementation methodologies.

Consolidation Example - Use Case 1: A new customer via the website

Jane Doe is a regular shopper at ACME's local store providing basic payment information and data like her full name and credit card information. Through her frequent shopping experience, she also signs up for ACME's loyalty program, requiring her phone and email address. Eventually, Jane Doe makes a purchase online. These online purchases require even more data such as her billing address, her home address, a contact email, and a phone number.

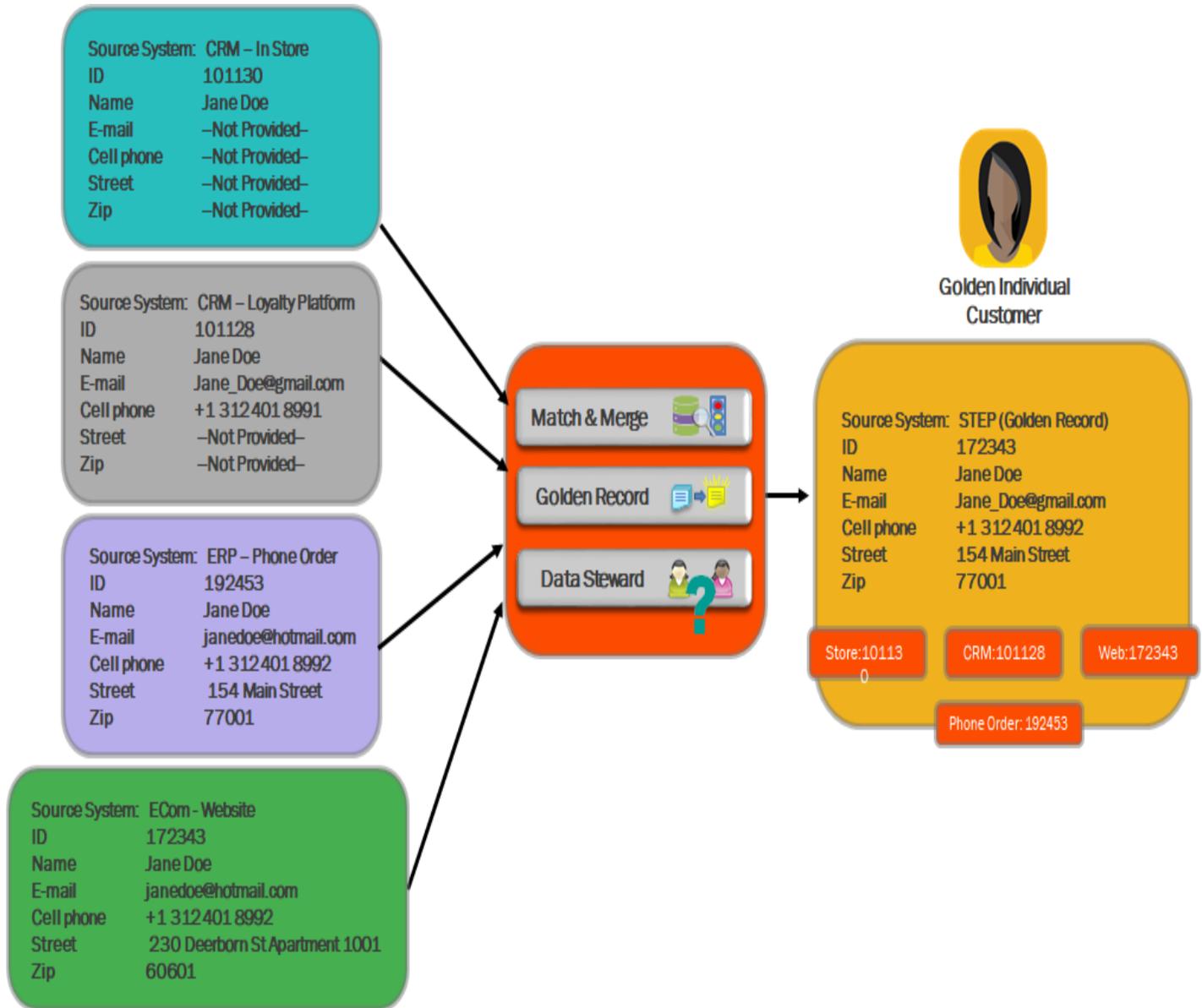
Using the CMDM solution, all of this data enters into STEP via three channels. At the local level, her data enters a CRM platform. When she entered her information on the ACME website, her data moves to an E-comm system. The E-comm and CRM systems are the source systems for STEP's CMDM solution.

Once her data enters the CMDM solution via an IIEP with Match & Merge importer, a matching algorithm is executed to detect that all three of her records are related to the same customer. These records are merged to create a golden record, which is then exported to downstream systems.



Consolidation Example - Use Case 2: Updating a Customer over phone in ERP

Some time later, Jane Doe calls ACME regarding a new order. During this call, the customer service rep learns that Jane has moved. A record is created, and this data is matched to the golden record, and it is forced to the clerical review, where a data steward confirms the changes. The address change is pushed to the golden record, and it is then exported downstream.



Coexistence Example - Use Case 1: Updating the Source Systems

Mary Sue has been a customer of ACME via their web site, exclusively. She signed up for a loyalty program with her email. When a new local location opens, Mary Sue begins shopping in store where she provides a new email. With the coexistence methodology, when this new record with a different email is imported into the CMDM solution, then a clerical review is initiated. A data steward confirms this email address and updates the golden record. The source systems are then updated with this new record as well as the downstream channels.

Matching

With a proper Matching strategy, organizations can consolidate their customer data from a variety of source systems and generate singular, definitive golden records for those individual customers.

Before creating a Matching configuration, it is important to get an understanding of the client organization's data and identify challenges the algorithm must account for. With this knowledge, a clear match strategy can be defined.

Once a strategy is defined, the Matching Algorithm is typically configured first, followed by the Match Codes. Both should be tailored to the client organization's data. The algorithm should be configured first in order to know what data points specific to the client organization will provide reliable matching. With the data points identified, you can define what attribute combinations will contribute to the match codes.

When the initial configurations are complete, they can be iterated upon via Matching Tuning sessions. The goal of these sessions is to gradually improve the accuracy of the matches being made, until you arrive at your final Matching configuration.

For more information on the Matching process and configuring the relevant components, see the **Matching, Linking, and Merging** documentation.

Matching Algorithm Configuration

The Matching Algorithm should be tailored to the data, and strive for precision. The CMDM Starter Package includes four pre-configured Matching Algorithms that can be used as a foundation to build a client organization's Matching solution:

- Individual Customer
- Organization Customer
- Contact Person
- Household

For more information, see the Algorithm & Match Codes topics.

When configuring the Matching Algorithm, it is important to consider the impact that Thresholds have on match results. If the Clerical Review threshold is set too high, a large amount of false negatives may be generated. Similarly, if the Auto Threshold is set too low, false positives could be generated. If the initial Matching configuration produces false negatives and/or false positives, the Thresholds should be reevaluated during the Match Tuning sessions.

For more information on configuring a Matching Algorithm, see the **Configuring Matching Algorithms** section of the **Matching, Linking, and Merging** documentation.

Match Codes Configuration

Match Codes should aim to ensure that anything the algorithm may score high is not left out, and that only those records that may score high get compared. The rules defined by the Match Criteria determine whether or not any given Match Code is relevant. The Match Codes should be considering the same data points that the Match

Criteria does. This ensures that the comparison pool that Match Codes generate is relevant to the data points the Match Criteria is matching on.

For example, if the Match Criteria is matching on a combination of Person Name and Address, it is not recommended to generate Match Codes based on unrelated attributes (e.g. Phone and Email).

Note: Configuring Match Codes is a development task, as it requires advanced scripting. However, the Starter Package comes with a library of scripts to get you started.

Example Match Codes

Typical attributes used to generate Match Codes for Individual (B2C) solutions include Name, Address, Email, and Phone. For Organization (B2B), these attributes potentially include Name, Address, Phone, DUNS (D&B Number), and Tax Identifier.

Email

Email is often used in combination with a person name for Match Criteria rules. In matching, no edit distance or phonetics is allowed. Email is potentially unique enough to warrant a high score on a match when included in a Match Criteria rule, removing the need for other Match Codes.

Example: A customer with the email 'HendrickBarnard@acme.com' becomes the Match Code 'EM~hendrickbarnard@acmecom'.

Phone

Phone shares many of the same benefits that email provides. However, it is common to have variations in whether country code is included or not. This can be solved by only including the last 7 digits of the phone number in the match code (for US phone numbers).

Example: A customer with the phone number '(337) 934-8392' becomes the Match Code 'PH~9348392'.

Person Name (First, Middle, and Last)

Person Name is often used in Match Criteria rules in combination with emails, phone numbers, or addresses. It has fairly low uniqueness compared to other Match Codes. When matching, techniques like edit distance, equivalent values and soundexes are used. Using person name as a Match Code requires a lot of simplification of the values in order to avoid false negatives and requires combination with other data elements (email, phone, etc.) to be sufficiently unique. For email and phone combinations, other Match Codes already capture the relevant candidates. It gets complicated when used in combination with addresses, as these suffer from some of the same complexities as person names.

Example: A customer with the name 'Hendrick Barnard' becomes the Match Code 'PN~h~PRNRT'.

Address

Address is often used in combination with person name or organization name. It is fairly unique if the full address is used, but has low uniqueness if only part of the address is used.

The accuracy of addresses varies (i.e. one entry may include suite no. while another may not). When matching, techniques like edit distance is used on city and street. Because of this, and the fact that accuracy varies, the full address does not make a good match code, as it will likely lead to false negatives.

Addresses are often abbreviated ('st' for street or station, etc.) in complex patterns that cannot be uniquely resolved easily. For high accuracy of address matching, it is therefore recommended to use STEP's Address standardization capabilities, which are integrated to Loqate.

Example: A customer with the zip code '92627' and country code 'US' becomes the Match Code 'ZICC~92627~us'.

Name and Address

Combining elements of a person name and elements of an address is often a good way to create Match Codes that are sufficiently unique, without causing false negatives. However, multiple permutations are often required to avoid false negatives.

Example: A customer with the name 'Hendrick Barnard' and zip code '92627' becomes the Match Code 'ZINM~92627~h~PRNRT'.

Other examples of Name and Address combination Match Codes:

- Zip + first letter of first first name + metaphone3 of last last name
- Zip + first letter of last last name + metaphone3 of first first name
- Metaphone3 of city + first letter of first first name + metaphone3 of last last name
- Metaphone3 of city + first letter of last last name + metaphone3 of first first name
- In case first name equivalents exists, it is likely necessary to add permutations for these as well

For more information on configuring Match Codes, see the **Configuring Match Codes** section of the **Matching, Linking, and Merging** documentation.

Starter Package Match Codes

The CMDM Starter Package includes four pre-configured sets of Match Codes that correspond to the four Matching Algorithms mentioned in the above section:

- Individual Customer
- Organization Customer
- Contact Person
- Household

For more information, see the Algorithm and Match Codes topics.

Matching Algorithm Tuning

Algorithm Tuning begins during the build phase of the implementation. Early on in this process, it is common to find a large number of invalid matches making it past the Auto-Merge threshold while valid matches fall short of the Clerical Review threshold. Thus, the ultimate goal of the algorithm tuning sessions is to perfect the matching logic's accuracy so that good and bad matches score within the appropriate thresholds.

This goal can be achieved by:

- Adjusting the scoring method and weighting of each scored attribute.
- Adjusting the relative weighting of scoring across all the scored attributes.
- Adjusting the Auto-Merge and Clerical Review Thresholds.

Note: It is important to have both the data steward and data owner present during algorithm review sessions. In this context, the data steward is the business user who has been tasked with the formation and execution of policies for the management of data and metadata. The data owner is the business user who typically has a direct line of responsibility for a functional area.

Algorithm Tuning Process

The Matching Algorithm Tuning process is as follows:

1. Configuration: Utilize a Match Tuning configuration to generate a Data Profile. Using this Data Profile, identify key data points to consider when configuring a baseline algorithm (Matching Algorithm and Match Codes).
2. Generate Sample Pair: Once the baseline algorithm is configured, generate the Random Sample Pair spreadsheet via a Match Tuning configuration. This baseline configuration is just a 'best-guess' configuration based on the analysis of the customer data so far.
3. Review Sample Pair: Review the Sample Pairs with the client. Each individual pair gets either a 'Yes', 'No', or 'Not Sure' indicating whether or not they should be considered the same entity by the algorithm and linked together.
4. Tuning the Algorithm: Tune the algorithm based on feedback from the Sample Pair review, and generate a new set of Sample Pairs based on the updated algorithm. Repeat steps 2 and 3 for two more cycles (or more, as needed).
5. Finalize: Make a decision on the final Auto Merge Threshold and Clerical Review Threshold.

Before the sample pair review can kick-off, the raw data from the output file should be formatted to make it human readable. The Sample Pair Formatter Excel sheet can optionally be used on the output file.

Note: To obtain the Sample Pair Formatter Excel sheet, please refer to the separately bundled Starter Package configuration files. Detailed instructions on how to use this Excel sheet is included within the sheet itself.

Once the random sample pair spreadsheet is generated and formatted, it is vital to review the sample pairs to see how the algorithm evaluates them. The primary purpose of the review is to assess the confidence of each merge, and modify the thresholds if the scores appear inaccurate. During the review process, it is also important to consider the following:

- The organization should mark each set with a decision as to whether (based on the data available) the records should be considered the same entity.
- It is best to approach this task from a 'human' standpoint as opposed to creating logic to help you achieve a certain score.
- This is not a data cleaning task.

The ultimate goal by the end of each Sample Pair Review session is to improve the quality of the matches found. It is much easier to identify false positives than false negatives in the pair export. Therefore, it is recommended to start with a wide net and narrow the Match Criteria during tuning.

Note: It is important to set expectations with the client organization that while those implementing the starter package can provide guidance, it is the organization's responsibility to determine which customer records should match and which should not.

The Sample Pair Review process can be a time-consuming task, but is critical in getting the algorithm tuned to meet requirements. Typically, 1,000+ Sample Pairs will need to be reviewed each cycle.

There are several personnel considerations to make during the sample pair review process:

- It is recommended that the consultant be present on-site for the Sample Pair Review sessions. Due to the highly interactive nature of these discussions, being on-site helps facilitate the process.
- Data Owners from different functional areas within the organization may have differing opinions on matching requirements. It may be difficult to get consensus among these various Data Owners.
- Some more sophisticated client organizations may introduce a Data Governance Board that can assist with reaching consensus among the Data Owners.

Other Considerations

It is recommended that the client organization re-engage with the applicable solution consultants post go-live for new Algorithm Tuning sessions when any of the following takes place:

- A new source is added that has different demographic or data quality.
- Substantial increase in False Positives or False Negatives.
- The Entity Size or Task Size metrics change significantly over time.

For more information on Matching Tuning and creating a Match Tuning configuration, see the **Match Tuning** section of the **Matching, Linking, and Merging** documentation.

Working with Customer Data

When developing a Matching strategy, it is important to consider the client organization's data and the potential challenges the algorithm will have to account for.

The following considerations and challenges are commonly encountered when implementing Matching:

- Algorithm Tuning is highly data-dependent, so real production data should be made available for analysis. As an example, start with 1000 records, eventually go to 100000, and in the end, at least 10% of the full data set should be taken in to consideration.
- Solution Consultants should expect delays in receiving customer data. They should work with the client organization early in the process to define the data to be delivered, and push to get it as early as possible. There may be both technical and process reasons for holding up data delivery. The ETL team may have issues in staging data from the source systems, and the legal or security team may introduce their own delays.
- Solution Consultants should establish a delivery date that the client organization agrees on, and emphasize that delays to that date will delay critical-path tasks. Having access to production data is a critical dependency to starting the Algorithm Tuning tasks.

- Solution Consultants should work with the client organization to determine the level of security needed around the data they provide. Note that the client may be held to a higher degree of security due to the regulation in their industry.
- Before Algorithm Tuning starts, the customer data should be analyzed to determine its quality and characteristics. This analysis will help in determining the baseline algorithm configuration. The Solution Consultant should analyze the data for attribute validity (percent populated), bad (or anonymous) data values populated, and any patterns in the data that may be instructive for Algorithm Tuning. The quality and characteristics of data can differ between sources, so data from each source system should be analyzed.
- Solution Consultants should not expect a client organization to be able to articulate their matching requirements. To arrive at a baseline algorithm configuration, a discussion with the client should occur focusing on what their priorities are for generating Golden Records. The Sample Pair review sessions will help facilitate the discussion around finalizing matching requirements.

Customer Data is required at key points during the implementation:

- 10-100 records for Data Modeling
- 20% of total data volume for Algorithm Tuning
- 100% of total data volume for go-live

Data from all sources must be included:

- Data quality and characteristics can vary from source system to source system, so getting sample from all sources is critical
- Samples of all object types in scope
- Other considerations for sampling data:
 - Data that crosses regions
 - The age of records (recently created records vs records created 20 years ago)
 - Records updated recently vs records not updated in years
 - Data captured through different means (call center, web, mobile, etc.)

Algorithm & Match Codes - Individual Customer

Matching Algorithm

The Individual Matching Algorithm delivered as part of this Starter Package is designed as a match algorithm for Individual Customer solutions. The most common data to have available for matching Individual Customers are First Name, Middle Name, Last Name, Address, E-mails, and Phone numbers

This algorithm consists of four normalizers and four matchers, with an Auto Threshold of 90.0 and a Clerical Review Threshold of 60.0.

For more information on configuring a Matching Algorithm, see the **Configuring Matching Algorithms** section of the **Matching, Linking, and Merging** documentation.

Normalizers

Normalizers are used to standardize values that are being compared. This ensures equal formatting is applied, increasing the accuracy of the comparisons being made. For more information, see the **Decision Table Normalizers** section of the **Matching, Linking, and Merging** documentation.

Person Name Normalizer

The Person Name Normalizer is configured to normalize the corresponding first, middle, and last name attributes (e.g. FirstName, MiddleName, LastName).

For customers with a large, non-English speaking consumer base, it is recommended to normalize accents and diacritic characters. Such characters may not 'play well' with the phonetic encoding of words (e.g. soundex or metaphone3) during the match process.

Address Normalizer

Because it is recommended to model addresses as data containers, the Address Normalizer will need to be configured to normalize data container attributes as defined within the Address component model.

Email Normalizer

Because it is recommended to model emails as data containers, the Email Normalizer will need to be configured to normalize data container attributes as defined within the Email component model.

Phone Normalizer

Because it is recommended to model phone numbers as data containers, the Phone Normalizer will need to be configured to normalize data container attributes.

Matchers

For general information on configuring Matchers, see the **Decision Table Matchers** section of the **Matching, Linking, and Merging** documentation.

Person Name Matcher

The Person Name Matcher is largely left with the default settings. However, an alias table is used to provide an equivalent names table.

Third party sources may be leveraged to build and enhance the equivalent names table within STEP. However, it is recommended that the client's expertise with their customer data be consulted to account for industry or business specific patterns. For example, if a company is based in southwestern United States then their equivalent names table may contain an emphasis on Hispanic name equivalents. Furthermore, equivalent name values may also be added as a result of a match tuning exercise with the client's customer data set.

Regarding middle names, an evaluation exercise with the customer is recommended to review the quality of data they have for customer middle names. It is possible that middle names are not collected from the consumers, or only middle initials are required but rarely provided. In such cases, it is recommended to reconsider the weight of MiddleName.

Note: The names in the equivalent names table provided by the Starter Package target the US market.

Address Matcher

The Address Matcher utilizes default configuration values.

Email Matcher

The Email Matcher utilizes default configuration values.

Phone Matcher

The Phone Matcher utilizes default configuration values.

Rules

When considering match rules, the recommended strategy is to dissect the customer's information into the smallest possible portions of data. These rules should not weigh the sum of *all* the customer's input data, and should instead be split so that it is possible to optimize each rule. Careful analysis of the customer dataset is required in order to determine what combinations of attributes present the best chance of uniqueness.

Three rules are provided to calculate the final score of the Individual Matching Algorithm. According to the configured conditions, if all respective Matchers resolve to true (i.e. scores above 70), then the algorithm will take the highest scoring rule as the final score of the algorithm. The rules are comprised of combinations of each Name score, Address score, Email score, and Phone score. The rules are then standardized to resolve to a value between 0 and 100.

These rules specifically include:

- Name & Address
- Name & Email
- Name & Phone

In this scenario, Name is used in all three rules because it is common for family members who use the same email and/or phone number to live at the same address. Thus email, phone number, or address are not reliably unique by themselves. By adding Name to these rules we are ensuring the uniqueness of the individual's name is taken into consideration, in addition to the other contact information elements.

It is possible to extend a rule by including various other combinations of matchers. This should be considered if there are specific requirements or use cases which requires specific combinations of matchers.

By adding another parameter, you have the ability to identify false positives. Unique identifiers such as Social Security, Passport, or Driver's License Numbers may be used as veto rules to further enhance the quality of the match process.

Other extensions include Date of Birth (DOB), which can be used in combination with other rules to be less strict on equality. For example, name and address runs the risk of a father and son having the same name which would be resolved by considering the date of birth. Additionally, you can accept lower scores of names and addresses, if DOB is equal.

Match Codes

For the Individual Customer entity type, three separate Match Codes are being generated. While based on the demographics of the customer record, these Match Codes are composed of: Email, Phone Number, and a combination of Individual Name and Address.

For information on how to configure Match Codes in STEP, see the **Configuring Match Codes** section of the **Matching, Linking, and Merging** documentation.

Email Match Code

The Email Match Code is simply the normalized value of the email address attribute. Both the username (or local part) of the email and the domain are normalized to ensure variations of the same email address (as a result of differing cases or special characters) are accounted for.

It is recommended to add a discernible prefix to each Match Code so the end-user may easily identify what attribute(s) the Match Code is based off of. The email Match Code contains the prefix 'EM~'.

Phone Number Match Code

The Phone Number Match Code is the normalized value of the individual's phone number attribute. The normalization removes any parenthesis and hyphenation in between numbers. Additionally, the area itself has been removed, leaving only the last 7 digits available as the Match Code. Phone number Match Code is prefixed by 'PH~'.

Address and Name Match Code

The Address and Name Match Code is a combination of elements of the individual's name and address. For example, the provided Match Code within this starter package is composed of: zip code + the first letter of the individual's first name + metaphone3 representation of the individual's last name.

The Address and Name Match Code contains the prefix 'ZINM~'.

Configuration Considerations

It is worth considering the use of the Equivalent Values Lookup Table and Anonymous Value Lookup Table. The Equivalent Value Lookup Table is used by both Match Codes and the Match Criteria in order to ensure that values that mean the same thing are evaluated as such. Equivalent values will score appropriately high, as if the values were actually the exact same.

Example:

- Name: Matt = Matthew

Note: Equivalent Values are only used for person & organization names.

The Anonymous Values Lookup Table is also used by both Match Codes and the Match Criteria in order to ensure that values that are anonymous, or not meaningful, do not contribute to identifying potential duplicates. Determining what these values should be is highly dependent on the organization's dataset.

Typically, these values will be default values that users of a Source System enter when they don't have the correct value, or don't want to enter a value. The actual anonymous values are not included in the baseline build of the Customer MDM configuration.

Examples:

- Phone: 999999999
- Address: DO NOT USE

Algorithm & Match Codes - Organizational Customer

Matching Algorithm

The Organization Matching Algorithm delivered as part of this Starter Package is designed as a match algorithm for Organization Customer solutions. This algorithm is most relevant in a Business to Business (B2B) style implementation and consists of two normalizers and two matchers, with an Auto Threshold of 90.0 and a Clerical Review Threshold of 60.0.

For more information on configuring a Matching Algorithm, see the **Configuring Matching Algorithms** section of the **Matching, Linking, and Merging** documentation.

Normalizers

Normalizers are used to standardize values that are being compared. This ensures equal formatting is applied, increasing the accuracy of the comparisons being made. For more information, see the **Decision Table Normalizers** section of the **Matching, Linking, and Merging** documentation.

Words Normalizer

The Words Normalizer for the Organization Algorithm is configured to normalize only the Legal Name (LegalName) attribute for organization entities.

Address Normalizer

Because it is recommended to model addresses as data containers, the Address Normalizer will need to be configured to normalize data container attributes as defined within the Address component model.

Matchers

For general information on configuring Matchers, see the **Decision Table Matchers** section of the **Matching, Linking, and Merging** documentation.

Words Matcher

The Words Matcher is used by the algorithm to match on organization names. This matcher is left with default settings, however it is recommended that the Words Matcher is tuned to the specifics of the organization.

Address Matcher

The Address Matcher utilizes default configuration values.

When matching organizations, keep in mind that multiple business entities may reside at a singular address. For example, Office Buildings where no suite number is provided.

Rules

When considering match rules, the recommended strategy is to dissect the customer's information into the smallest possible portions of data. These rules should not weigh the sum of *all* the customer's input data (which is likely inefficient), and requires a careful analysis of the customer dataset in order to determine what combinations of attributes present the best chance of uniqueness.

There is only one Rule associated with the Organization Algorithm. This Rule evaluates the scores of the Name and Address matchers. The resulting score is then weighed against the defined algorithm thresholds to determine whether the organization record should be created, auto-merged, or if it requires a clerical review.

Note: Common attributes to match for organizations on include Address, DUNS (D&B Number), and Tax Identifier.

Match Codes

For the Organization Customer entity type, three separate Match Codes are being generated. While based on the demographics of the organization record, these Match Codes are composed of the Email, Phone Number, and a combination of Legal Name and Address.

Since initial demographic data for organizations may be limited prior to enrichment (for example, Dun & Bradstreet detailed profile), generating these three Match Codes will help ensure proper identification of organizations during the match process.

For information on how to configure Match Codes in STEP, see the **Configuring Match Codes** section of the **Matching, Linking, and Merging** documentation.

Email Match Code

The Email Match Code is simply the normalized value of the organization's main contact email address. Both the username (or local part) of the email and the domain are normalized to ensure variations of the same email address (as a result of differing cases or special characters) are accounted for.

It is recommended to add a discernible prefix to each Match Code so the end-user may easily identify what attribute(s) the Match Code is based off of. The Email Match Code contains the prefix 'EM~'.

Phone Number Match Code

The Phone Number Match Code is the normalized value of the organization's main contact phone number. The normalization removes any parenthesis and hyphenation in between numbers. Additionally, the area itself has been removed, leaving only the last 7 digits available as the Match Code. Phone number Match Code is prefixed by 'PH~'.

Organization Address and Name Match Code

The Organization Address and Name Match Code is composed of: the zip code + metaphone3 representation of the organization's legal name.

The Organization Address and Name Match Code contains the prefix 'ZONM~'.

Algorithm & Match Codes - Contact Person

Matching Algorithm

The Contact Person Matching Algorithm delivered as part of this starter package is designed as a match algorithm for Contact Person records. This algorithm is relevant in a Business-to-Business (B2B) style implementation and consists of four normalizers and four matchers with an Auto Threshold of 90.0 and a Clerical Review Threshold of 60.0.

For further details regarding the Contact Persons use case, see the **Data Modeling** and **Use Case Overview** sections of this documentation.

For more information on configuring a Matching Algorithm, see the **Configuring Matching Algorithms** section of the **Matching, Linking, and Merging** documentation.

Normalizers

Normalizers are used to standardize values that are being compared. This ensures equal formatting is applied, increasing the accuracy of the comparisons being made. For more information, see the **Decision Table Normalizers** section of the **Matching, Linking, and Merging** documentation.

Person Name Normalizer

The Person Name Normalizer is configured to normalize the corresponding first, middle, and last name attributes (e.g. FirstName, MiddleName, LastName).

For customers with a large, non-English speaking consumer base, it is recommended to normalize accents and diacritic characters. Such characters may not 'play well' with the phonetic encoding of words (e.g. soundex or metaphone3) during the match process.

Address Normalizer

Because it is recommended to model addresses as data containers, the Address Normalizer will need to be configured to normalize data container attributes as defined within the Address component model.

Email Normalizer

Because it is recommended to model emails as data containers, the Email Normalizer will need to be configured to normalize data container attributes as defined with the Email component model.

Phone Normalizer

Because it is recommended to model phone numbers as data containers, the Phone Normalizer will need to be configured to normalize data container attributes.

Matchers

For general information on configuring Matchers, see the **Decision Table Matchers** section of the **Matching, Linking, and Merging** documentation.

Person Name Matcher

The Person Name Matcher is largely left to the default settings. However, an alias table is used to provide an equivalent names table.

Third party sources may be leveraged to build and enhance the equivalent names table within STEP. However, it is recommended that the client's expertise with their customer data be consulted to account for industry or business specific patterns. For example, if a company is based in southwestern United States then their equivalent names table may contain an emphasis on Hispanic name equivalents. Furthermore, equivalent name values may also be added as a result of a match tuning exercise with the client's customer data set.

Regarding middle names, an evaluation exercise with the customer is recommended to review the quality of data they have for customer middle names. It is possible that middle names are not collected from the consumers, or only middle initials are required but rarely provided. In such cases, it is recommended to reconsider the weight of MiddleName.

Note: The names in the equivalent names table provided by the Starter Package target the US market.

Address Matcher

The Address Matcher utilizes default configuration values.

Email Matcher

The Email Matcher utilizes default configuration values.

Phone Matcher

The Phone Matcher utilizes default configuration values.

Rules

When considering match rules, the recommended strategy is to dissect the customer's information into the smallest possible portions of data. These rules should not weigh the sum of *all* the customer's input data, and should instead be split so that it is possible to optimize each rule. Careful analysis of the customer dataset is required in order to determine what combinations of attributes present the best chance of uniqueness.

Three rules are provided to calculate the final score of the Contact Person Matching Algorithm. According to the configured conditions, if all respective matchers resolve to true (i.e. scores above 70), then the algorithm will take the highest scoring rule as the final score of the algorithm. The rules are comprised of combinations of each Name score, Address Score, Email Score, and Phone score. The rules are then standardized to resolve in a value between 0 and 100.

- Name & Address
- Name & Email
- Name & Phone

In this scenario, Name is used in all three rules because it is common for family members who use the same email and/or phone number to live at the same address. Thus email, phone number, or address are not reliably unique by themselves. By adding Name to these rules we are ensuring the uniqueness of the individual's name is taken into consideration, in addition to the other contact information elements.

It is possible to extend a rule by including various other combinations of matchers. This should be considered if there are specific requirements or use cases which requires specific combinations of matchers.

By adding another parameter, you have the ability to identify false positives. Unique identifiers such as Social Security, Passport, or Driver's License Numbers may be used as veto rules to further enhance the quality of the match process.

Other extensions include Date of Birth (DOB), which can be used in combination with other rules to be less strict on equality. For example, name and address runs the risk of a father and son having the same name which would be resolved by considering the date of birth. Additionally, you can accept lower scores of names and addresses, if DOB is equal.

Match Codes

For the Contact Person entity type, three separate Match Codes are being generated. While based on the demographics of the Contact Person record, these Match Codes are composed of the Email, Phone Number, and a combination of Contact Person's Name and Address.

To ensure the algorithm will only match Contact Persons within the same organization, the Match Codes should contain the ID of the reference organization for a given Contact Person. This ensures that only Contact Persons within the same organization will be deduplicated.

For information on how to configure Match Codes in STEP, see the **Configuring Match Codes** section of the **Matching, Linking, and Merging** documentation.

Email Match Code

The Email Match Code is simply the referenced organization's ID and the normalized value of the email address attribute. Both the username (or local part) of the email and the domain are normalized to ensure variations of the same email address (as a result of differing cases or special characters) are accounted for.

It is recommended to add a discernible prefix to each Match Code so the end-user may easily identify what attribute(s) the Match Code is based off of.

The email Match Code contains the prefix 'EM~'.

Phone Number Match Code

The Phone Number Match Code is the referenced organization's ID and the normalized value of the Contact Person's phone number attribute. The normalization removes any parenthesis and hyphenation in between numbers. Additionally, the area itself has been removed, leaving only the last 7 digits available as the Match Code. Phone number Match Code is prefixed by 'PH~'.

Address and Name Match Code

The Address and Name Match Code is a combination of the referenced organization's ID and elements of the Contact Person's name and address. For example, the provided Match Code within this starter package is composed of: the zip code + first letter of the individual's first name + metaphone3 representation of the Contact Person's last name.

The Address and Name Match Code contains the prefix 'ZINM~'.

Configuration Considerations

It is worth considering the use of the Equivalent Values Lookup Table and Anonymous Value Lookup Table. The Equivalent Value Lookup Table is used by both Match Codes and the Match Criteria in order to ensure that values that mean the same thing are evaluated as such. Equivalent values will score appropriately high, as if the values were actually the exact same.

Example:

- Name: Matt = Matthew

Note: Equivalent Values are only used for person & organization names.

The Anonymous Values Lookup Table is also used by both Match Codes and the Match Criteria in order to ensure that values that are anonymous, or not meaningful, do not contribute to identifying potential duplicates. Determining what these values should be is highly dependent on the organization's dataset.

Typically, these values will be default values that users of a Source System enter when they don't have the correct value, or don't want to enter a value. The actual anonymous values are not included in the baseline build of the Customer MDM configuration.

Examples:

- Phone: 999999999
- Address: DO NOT USE

Algorithm & Match Codes - Household

Matching Algorithm

The Household Matching Algorithm delivered as part of this Starter Package is designed as a match algorithm for Household entities. This algorithm is most relevant in a Business to Consumer (B2C) style implementation and consists of two normalizers and two matchers, with an Auto Threshold of 90.0 and a Clerical Review Threshold of 80.0. The Clerical review threshold is both higher and tighter to ensure only customer records that are extremely likely to belong to the same household are considered.

It is worth noting that the Household Matching Algorithm utilizes the Link Golden Record solution engine rather than the Merge Gold Record approach. This is to ensure that individual customers that are determined to be of the same household are linked to a separate household entity object rather than merged.

For details regarding the Household entity and its purpose, see the **Data Modeling** section of this documentation.

For more information on configuring a Matching Algorithm, see the **Configuring Matching Algorithms** section of the **Matching, Linking, and Merging** documentation.

Normalizers

Normalizers are used to standardize values that are being compared. This ensures equal formatting is applied, increasing the accuracy of the comparisons being made. For more information, see the **Decision Table Normalizers** section of the **Matching, Linking, and Merging** documentation.

Last Name Normalizer

The Last Name Normalizer uses the Words Normalizer, and only evaluates the Last Name attribute of an individual record. This is because households are referred to only by the Last Name, or household name.

Address Normalizer

Because it is recommended to model addresses as data containers, the Address Normalizer will need to be configured to normalize data container attributes as defined within the Address component model.

Matchers

For general information on configuring Matchers, see the **Decision Table Matchers** section of the **Matching, Linking, and Merging** documentation.

Name Matcher

The Last Name Matcher is largely left with the default settings. An equivalent names look-up table is not utilized for Household entities since last names generally do not have shortened versions or a nickname alternative.

Address Matcher

The Address Matcher utilizes default configuration values.

Rules

When considering match rules, the recommended strategy is to dissect the customer's information into the smallest possible portions of data. These rules should not weigh the sum of *all* the customer's input data (which is likely inefficient), and requires a careful analysis of the customer dataset in order to determine what combinations of attributes present the best chance of uniqueness.

There is only one rule associated with the Household algorithm, which evaluates the scores of the Name and Address matchers. The resulting score is then weighed against the defined algorithm thresholds to determine whether the individual record should be created, auto-linked to an existing record, or it requires clerical review.

The Starter Package matches on Last Name and Address for households. Matching on Last Names may be advantageous in identifying family units within a large population such as a retirement home or a school. However, in some cases a married couple within a household may not have the same Last Name. It may be beneficial to make Address weigh more in scenarios like this.

Conversely, sometimes a large number of customer records with the same Address should not be considered the same household. This can include college campuses, retirement homes, and similar institutions. These scenarios should be handled on a case-by-case basis. However, one possible approach is to utilize an anonymous words table to anonymize aforementioned addresses, eliminating addresses from the matching logic for such records.

Match Codes

The Household Match Code is generated on the Individual Customer entity type along with the Individual Match Codes. Since a household is identified by its last name, the Household Match Code is composed of: the zip code + metaphone3 representation of the individual's last name.

The Household Match Code contain the prefix 'ZILNM~'.

Loqate

Loqate validations can occur at three different points during the customer record life cycle:

- During Import of Source Records
- After surviving attribute values have been determined for a newly created Golden Record
- When updating existing Golden Records

Note: It is recommended to validate address values upon Import and after Survivorship.

For more general information on Loqate, see the **Loqate Integration** section of the **Data Integration** documentation.

Validate on Import

By calling Loqate during Import, each Source Record will have its address values standardized before matching takes place. Raw values taken from Source Systems commonly include typos or inconsistent formatting. By standardizing these values beforehand, the matching algorithm is able to match records more accurately.

Validate on Survivorship

Validating addresses after survivorship ensures that the surviving address values are standardized. However, because address matching is less reliable when matching raw values taken from Source Systems, it is recommended to validate on Import as well.

Validate Updates to Existing Records

A Data Steward can trigger Loqate for single records or when running a bulk update for a collection. By using either of the maintenance methods listed below, address values can be standardized when records are updated.

Maintenance Validation Methods:

- A 'Run Business Rule' button on a Node Details screen
- Running a Bulk Update configuration (an Address Validation business rule can be configured for bulk updates)

By using Bulk Update, existing records can have their address values standardized whenever their data is updated.

Data Modeling

Customer data can be managed in two overall ways: either as an individual or as an organizational customer.

Note: There is no need to use different dimensions and contexts in CMDM solutions, nor should dimension specific attribute values and references be used. Instead, UI translations may use a language dimension with names of references and attributes being language specific.

When setting up records, ensure that the STEP IDs are set to autogeneration.

Customer Object Types

What is a Customer?

A customer is anyone who transacts with a business. In a CMDM sense, a customer is an object that represents these people. This customer object can represent an individual in B2C instances or an organization in B2B scenarios.

The typical customer attributes are:

Individual Customer:

- First Name
- Last Name
- Addresses
- Date of Birth
- Gender
- Contact Details
 - Phone Numbers
 - Fax
 - Emails
- Social Media IDs
- Loyalty Card Information
- Government IDs (social security, passport, driver's license)

Organization Customer:

- Legal Name of the Organization
- Date of Formation
- Address
- Contact Information
 - Main Phone
 - Fax

- Main Email
- Website URL
- GLN / ILN
- DUNS Number
- Tax Identification Number
- Risk Category

A **contact person** owns the relationship of the organization as governed by ERP and CRM systems. A contact person is authored in CRMs and deduplication is recommended. For contact persons, email and phone elements are separate data containers types. This setup allows for multi-entries around these contact details:

The Email Address data container usually has:

- Whether or not it is the primary address or not
- When was the last time the email was validated?

For the phone data container, the following attributes are included:

- Phone Number
- What type of phone number is this (primary and type)

What is a Household?

A household is a collection of individuals, usually in a family or other grouping, usually of individuals in the same home. Each of these individuals are called members of the household. In the CMDM sense, these household members are individual customers. The purpose of these household constructs are to help businesses target customers that share a household either as a family unit or as individual customers. This allows the company to limit extraneous mailing campaigns and aggregate buying patterns.

How are Addresses Structured?

In STEP, a location is a physical or logical position. The address is a description of this location wherein there is only one address per location. Additionally, locations may also have a contact person and opening hours in addition to an address. Address data containers will have two addresses, for example for main and delivery.

Note: The sample data provided in the Starter Package provides only main addresses. There are no delivery addresses.

Attribute Setup

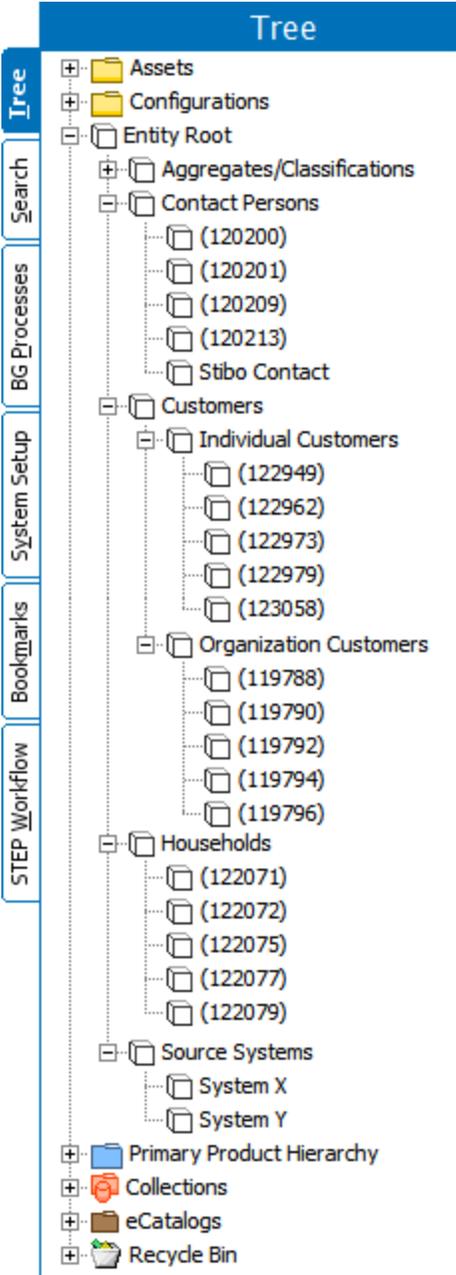
The following items should be set up:

- Domain Specific Attribute Groups
- Basic Domains
- Data Container Types

Domain Folders

In the Entity Root folder, create separate folders for each matching algorithm such as the individual, organization, contact persons, etc. Next, create a Household entities folder to ease traceability. Finally, add a source systems entity folder where all the source systems will reside without further configuring of the structure. There's no hierarchy to this structure so that the matching algorithm will point to the root node and sift through all records below.

Note: Source systems are often referenced and used in the CMDM solution. Source systems are the ERP, E-comm, and CRM type systems where data is authored based on interactions with the customers then sent to CMDM.



These domain folders should focus more on hierarchy constraints that separate the data domains per folder. These domain folders should be simple as possible since complexity isn't needed.

Entity Super Types

STEP uses four types of object super types to model data: products, classifications, assets, and entities.

Entities are the suggested object for modeling customer, contacts, and household data. Some benefits of entity object types include:

- Globally revisable, which requires no approvals, workspace revisions, and instantaneous updates.
- References can be configured as parent-child relationships.
- Data Containers can be used. Data Containers are designed with CMDM in mind to create a single object that would hold multiple address attribute values.
- Matching and Merging

Some deficits to using entities are that they do not support inheritance, translations, and specification attributes.

For more information on entities in STEP, see the **Entity Maintenance** topic in the **Getting Started** documentation.

Important Considerations for CMDM Data Modeling

Use Data Containers to maintain and collect all your customer data into easily grouped objects. It's a good idea to create multiple data containers when using a lot of unique contextual attributes, when using single / multiple value attributes, and when needing specific domain use cases.

Also, use different object types for individual customers and organization customers. This allows the attributes to be different per customer type, allows matching and merging and workflows to be different, and creates unique relationships in hierarchies.

Use caution when applying one-to-many relationships for information as multi-value attributes. Most values will use additional information to associate other values. Also, avoid creating data containers for outside entities. For example, contact persons are a separate object type and can be referenced by the customer separately.

This Implementation Guidelines assumes that customers, contact persons, and organizations are created and maintained by the merge golden records solution. Households are created and maintained by using the Link Golden Record solution.

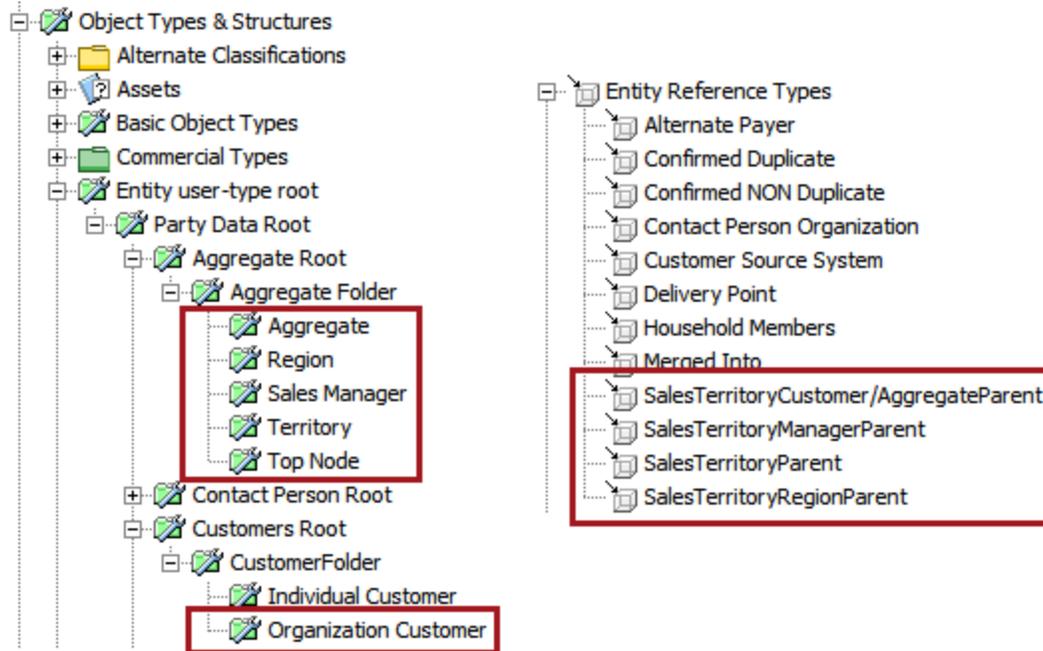
Customer Hierarchies

Hierarchies are a way to model data within a B2B solution. The key thing hierarchies require are references that link Organization Customers and Aggregate Objects. It's important to understand that Aggregate Objects are used for organizing objects within the hierarchies. These aggregates allow the user to customize their hierarchies as needed.

Note: While Aggregates exist in STEP, they do NOT represent real world customers.

An Organization may exist in multiple hierarchies, such as sales, billing, etc. It's important to remember that multivalued references **cannot** be used. Because of this limitation, separate reference types are required for multiple hierarchies.

When creating a hierarchy, the hierarchy references flow bottom up, meaning the child / target links to the parent / source. When complex hierarchies utilizing multiple aggregate types are required it may be necessary to restrict what object types can go above or below in a hierarchy. For this scenario, multiple reference types are required with appropriately configured valid source / target types.



In the above image, the highlighted object types and references are the ones used for ACME's hierarchy. By having multiple aggregate types, like the sales manager, region, etc., the data steward was able to create a hierarchy that visualizes what sales manager is responsible for which Organization Customers while also grouping geographically.

For another example in the CMDM system, in STEP search for record **ID X-12345**. This will show a company hierarchy with references.

Web UI Configurations for CMDM

The Web UI allows data stewards to view and maintain customer data, particularly for verifying the data model, as explained in the **Data Modeling** topic of this documentation, as well as to facilitate in conflict resolution. Since the primary user of the Web UIs are data stewards, the Web UI is configured to ensure the most thorough data quality. Even still, the Web UI is designed to be constrained to help less experienced users perform their needed activities.

Role of the Data Stewards

Experienced data stewards need access to all of their tasks that allow for the managing of data and ensuring data quality.

These tasks include:

1. Clerical Review - Data Stewards can review potential duplicates and non-duplicates.
2. Manually maintaining records - Data Stewards may use a series of components to update and edit records as needed.
3. Monitoring / Data Quality - Data Stewards should use data monitoring widgets to review customer data in the Web UI, and monitor Source system performance.

To do these tasks, the following Web UI components are used.

Clerical Review

Using the tasks, data stewards can ensure that clerical review of records are handled. For information on workflows or setting up a workflow task list, see the **Workflows** documentation or see the **Workflows in Web UI** topic in the Web UI documentation. These tasks are created in workflows that are defined in the Workbench.

Tasks	
Clerical Review	
Individual Review	16
Organization Review	1
Household Review	22
Contact Review	1

- **Icon-based Navigation Bar** - The icon-based navigation bar provides an enhanced User Experience to provide the most features while keeping the interface intuitive.
- **Task List** - The task list has been improved to provide grouping of tasks by listing all potential duplicate records. The task list shows the how many items are in each state of the work flow in a visual manner.

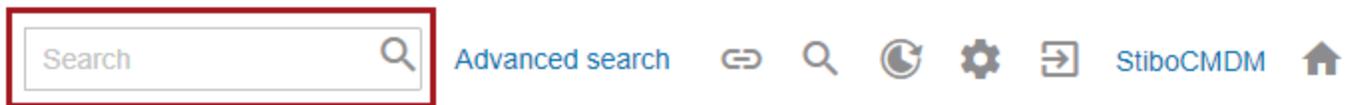
- **Advanced Merge** - As records fall within the Clerical Review threshold, Data Stewards are given the ability to manually dictate which attributes survive as the result of a merge. Within a data container, individual entries may be selected. For example, an address data container has one address from one source system and two from another. The Data Steward may multiselect from these entries for merging. Additionally, a preview panel displays what comprises the final merged record for Data Stewards to review prior to confirming the merge.
- **Exclude records from a potential duplicate task** - Data Stewards may select which records are excluded in the merge, such as a record with poor data quality or a record that the Data Stewards recognizes as not being a duplicate. Alternatively, Data Stewards may select which record survive merging allowing for source system IDs and other attributes.
- **Ability to re-assign tasks to different users** - Data Stewards can then select other users to take on the task to determine if merging of selected records should occur. This functionality delegates workload and assists in escalations if managerial approval is needed.
- **High Priority State** - A business rule that is written to govern what tasks are considered high priority based on criteria from the business user such as the Risk Flag attribute. This helps data stewards address critical tasks as these are presented in a different task list.

These tasks may be assigned to specific users, who are in charge of performing actions on them. For information on how to perform actions in a task list during a clerical review, see the **Golden Record Clerical Review Task List** topic of the Web UI documentation.

Manual Maintenance

The Data Steward can maintain any records as needed to ensure data quality or per requests from the business users.

- **Global Header Search Bar** is in the header of the Web UI. This is configured to allow users to search via Attribute Values as well as source record IDs. Data Stewards would be able to search using provided source record IDs from the business user rather than having to find the corresponding STEP ID, which may not be available / visible to business users and source systems. For more information on this search, see the **Global Header Search** topic of the Web UI documentation.



- **Customer Details Screen** - The summary screen provides an overview of each customer record. They can be customized to the needs of the Data Steward to show the relevant attributes in the STEP system.
- **Summary Cards** - Used on details screen to display key information about below the title.

Candace Fulgum INDIVIDUAL CUSTOMER • ID: 123249
 36 Garden St Apt 5A Sharon, MA 2067, US • (241) 502-8021 (251) 652-4606

- **Data Container Pop Up Editor** - When editing an address data container, an editor will pop up to allow multi-value editing.

Main Address Data Container
✕

<div style="display: flex; align-items: center;"> 36 Garden St Sharon, MA 0... 🗑️ </div>	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;">City</td> <td style="border: 1px solid #ccc; padding: 5px;">Sharon</td> </tr> <tr> <td style="padding: 5px;">Country</td> <td style="border: 1px solid #ccc; padding: 5px;">US</td> </tr> <tr> <td style="padding: 5px;">Country ISO Code</td> <td style="border: 1px solid #ccc; padding: 5px;"></td> </tr> <tr> <td style="padding: 5px;">State</td> <td style="border: 1px solid #ccc; padding: 5px;">MA</td> </tr> <tr> <td style="padding: 5px;">Street</td> <td style="border: 1px solid #ccc; padding: 5px;">36 Garden Street</td> </tr> <tr> <td style="padding: 5px;">Zip</td> <td style="border: 1px solid #ccc; padding: 5px;">2067</td> </tr> </table>	City	Sharon	Country	US	Country ISO Code		State	MA	Street	36 Garden Street	Zip	2067	<div style="border-bottom: 1px solid #ccc; padding-bottom: 5px;"> <p>Standardized Address</p> <p>36 Garden St Sharon, MA 02067-2654, United States</p> </div> <div style="padding-top: 5px;"> <p>Quality Index</p> <p>B</p> </div>
City	Sharon													
Country	US													
Country ISO Code														
State	MA													
Street	36 Garden Street													
Zip	2067													

✓ OK
✕ Cancel

Household Visualizations - In a B2C scenario, these visualizations give a quick view into which individual customers are in a set household.

Company Hierarchy Screen - In a B2B scenario, the company hierarchy screen allows users to see a visual representation of an organizations hierarchy with relationships for parents and children shown.

Golden Record Source Traceability Screen - This screen provides revision and historical information on the customer record based on the respective source systems. This screen shows from where and when the data comes into STEP on an attribute by attribute basis.

Confirmed Non-Duplicate Screen - This screen shows all the historic information on previously rejected potential duplicate records.

Data Monitoring

Data profiles can be used to monitor all changes that occur within STEP data. For example, if ACME imports data into STEP from their Ecom system, a collection for all recently changed items (such as for the last seven days) can

be created and monitored. This data can then be studied and watched to ensure that the data is improving or to make preventative decisions if the data quality is deteriorating. Through this collection, the full dataset can be tracked and data quality can be maintained.

Recently Updated Records Collection

105 Items • Last run : 2018-01-19 15:12:16

Overview **Content**

Object Type Count

A pie chart showing the distribution of object types. The chart is almost entirely green, representing 'Organization Cus...' at 98.10%. A very small orange slice represents 'Individual Custo...' at 1.90%. A legend to the right identifies the colors: a green circle for 'Organization Cus...' and an orange circle for 'Individual Custo...'. An information icon (i) is located to the right of the chart.

Object Type	Percentage
Organization Cus...	98.10%
Individual Custo...	1.90%

Average Object Completeness 51%

A horizontal bar chart showing the distribution of average object completeness. The y-axis lists completeness ranges: 0-10%, 20-30%, 40-50%, 60-70%, and 80-90%. The x-axis represents the number of objects, ranging from 0 to 200. The bars are green. The 40-50% range has the highest count, followed by 60-70%, 20-30%, 80-90%, and 0-10%. An information icon (i) is located to the right of the chart.

Completeness Range	Count (Approximate)
0-10%	25
20-30%	130
40-50%	190
60-70%	175
80-90%	140

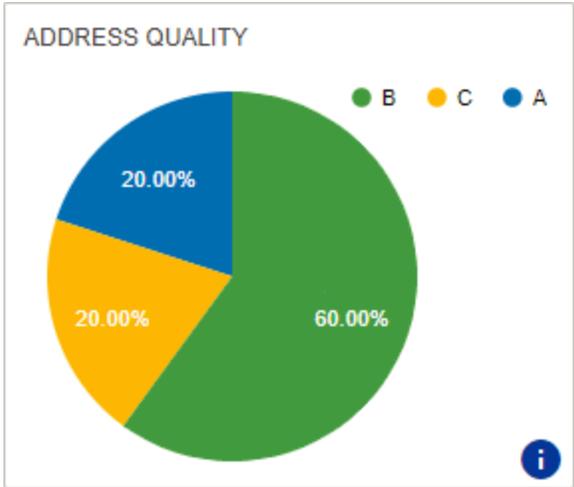
[Update profile](#)

KPI Widgets

KPI (Key Performance Indicator) Widgets are used to measure the quality of the data set as well as how complete the object's data set is. These widgets can be drilled in further to search records and action on any data needing updates. For information on how to set up a KPI Widget, see the **Data Profile Widgets in the Web UI** topic in the Web UI documentation.

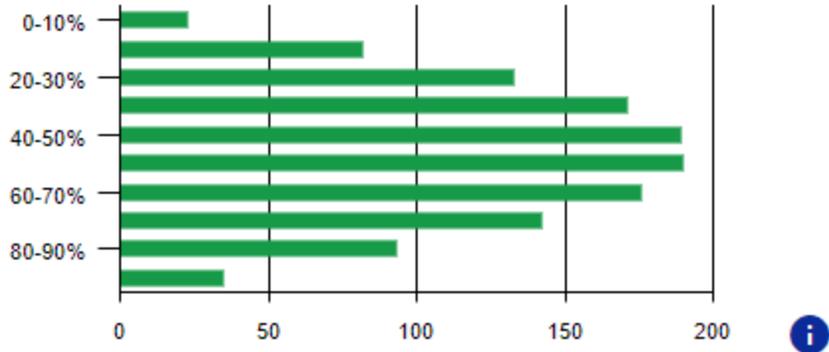
Some of the suggested widgets are:

- the Attribute Value Profile, which is based on the input address. The address quality of source systems can be monitored by the data steward using this widget. Attribute Value Profiles can also be used to measure email quality that is received from source systems as they are validated from a third party integration when applicable.



- the completeness bar chart, which is based off the individual completion score on each record. With a snapshot of the status of the customer records in the system, the data steward can have an informed view of the data they control in STEP as well as in their respective source system.

Average Object Completeness 51%



Each record in the Web UI will have a completeness meter.



Key Identifiers

Golden Record ID 123015 - Active

Data stewards will see the percentage of completed data every time they open a collection. This visually reminds them to track down where the data quality gaps may be. Data stewards also may evaluate characteristics based on the dimensions associated with those attributes. For example, it's possible to set up a time dimension that would show KPIs for only items that have recently changed.

Collections

One of the most powerful tools for a data steward is the ability create an ad-hoc collections based off of saved Advanced Search results. This allows Web UI users to save a batch of search criteria to recall as needed. Using these collections and the collection screen, users may also bulk update all the objects in the collection. For more information on collections and bulk updates, see the **Bulk Updates** documentation as well as the **Collections** topic of the Getting Started documentation.

Consider the following use case from ACME, Inc. Their marketing team wants to target customers that have an annual income of more than \$100,000. This would allow them to tailor advertising. To do this, perform an advanced search for greater than 100,000 on the Income attribute.

Once this is saved, a data steward can monitor these customers, then pass along the information to marketing for their campaign.

High Income Customers Collection

81 Items • Last run : 2018-01-19 10:05:24

Overview Content

Select all Clear filter Refresh collection

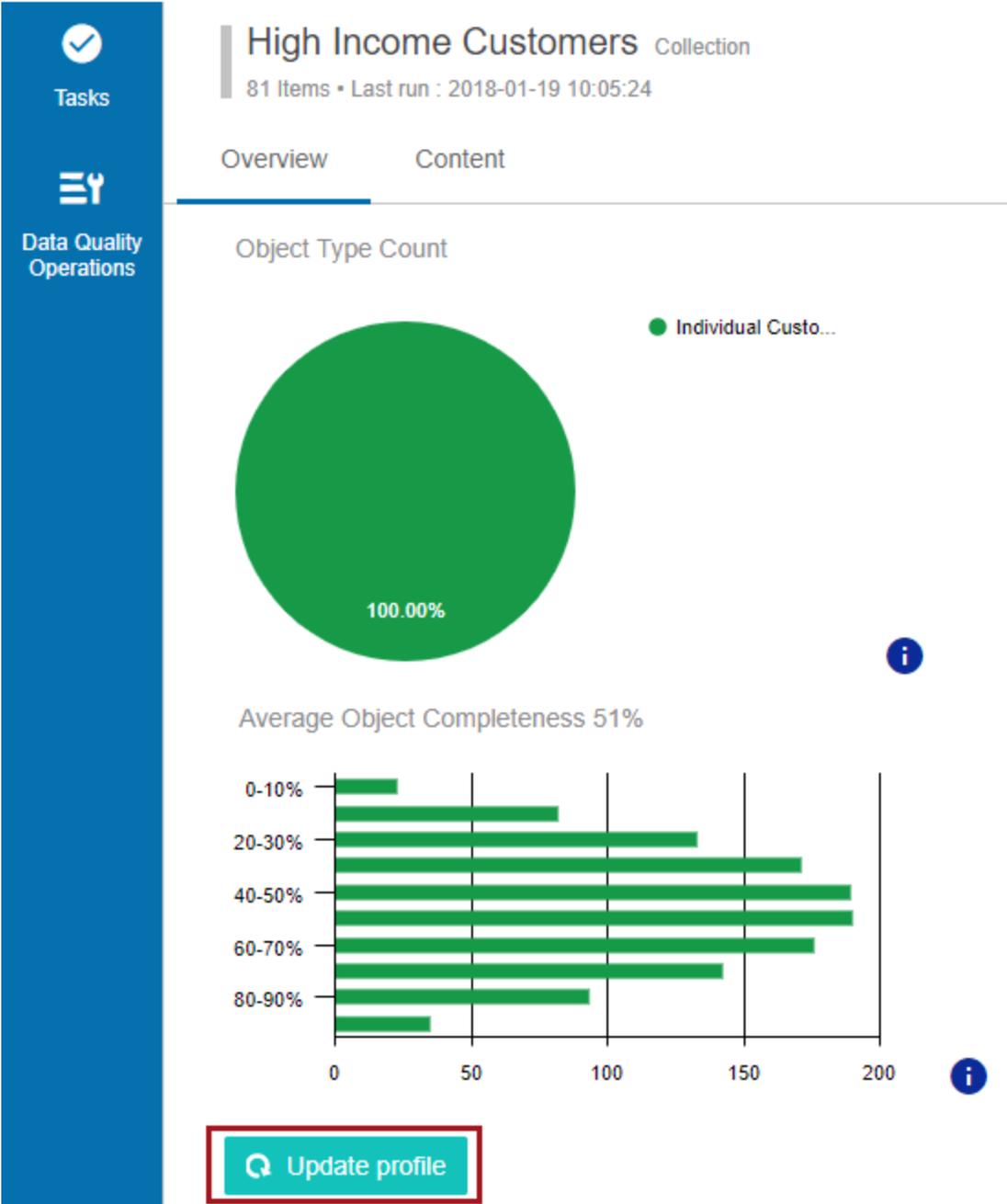
ID	Name	Address
<input type="checkbox"/> 122949	Abra Jonathan	101 Main St Apt 5 Bowling Green, KY 42101, US
<input type="checkbox"/> 122950	Deana Jonathan	101 Main St NW Apt 5 Bowling Green, KY 42101, US
<input type="checkbox"/> 122961	Naida Kristen	9821 US Highway 60 West Owensboro, KY 42301, US
<input type="checkbox"/> 122979	Philip Kristen	9821 US Highway 60 Owensboro, KY 42301, US
<input type="checkbox"/> 122985	Colbie Allistair	50 Ashley Av Woodland, CA 95695, US
		50 Ashley Ave

Update profile

Selecting any collections from the Data Quality Operations pane will show a Data Collection screen.

The screenshot shows the 'Data Quality Operations' pane on the left with a 'Tasks' section containing 'Create collection'. Below this, a list of collections is shown under 'Adhoc Collections', with 'High Income Customers' highlighted in blue and circled with a red '2'. A red '1' is also present near the 'Data Quality Operations' header. On the right, the 'High Income Customers Collection' screen is visible, showing '81 Items • Last run : 2018-01-19 10:05:24' and 'Overview' and 'Content' tabs. A red '3' is circled next to the last run timestamp.

The collection screen requires an update to show any changes done to the collected items.



In the above screenshot, the Overview tab shows all the data on the selected Ad-Hoc Collection, which contains all the customer records who have a reported income of more than \$100,000. It shows how much data is complete across the records, in this example only 51%, and also if configured, it can show the Most Complete Attributes as well as the Least Complete Attributes.

The content tab of the collection will show all the items saved as part of this collection. Once again, any changes that were made, require pressing the Update Profile button to see the updates.



Tasks



Data Quality Operations

High Income Customers Collection

81 Items • Last run : 2018-01-19 10:05:24

Overview Content

 Select all  Clear filter  Refresh collection

ID	Name	Address
<input type="checkbox"/> 122949	Abra Jonathan	101 Main St Apt 5 Bowling Green, KY 42101, US
<input type="checkbox"/> 122950	Deana Jonathan	101 Main St NW Apt 5 Bowling Green, KY 42101, US
<input type="checkbox"/> 122961	Naida Kristen	9821 US Highway 60 West Owensboro, KY 42301, US
<input type="checkbox"/> 122979	Philip Kristen	9821 US Highway 60 Owensboro, KY 42301, US
<input type="checkbox"/> 122985	Colbie Alistair	50 Ashley Ave Woodland, CA 95695, US
<input type="checkbox"/> 122991	Colby Alistair	50 Ashley Ave Wodland, CA 95695, US
<input type="checkbox"/> 122992	Jim Kristen	9821 US Hwy 60 W Owensboro, KY 42301, US
<input type="checkbox"/> 123009	Lewis Bradley	348 S Lexington Ave White Plains, NY 10606, US
<input type="checkbox"/> 123015	Irene Bradley	348 S Lexington Ave White Plains, NY 10606, US
<input type="checkbox"/> 123021	Darryl Lilah	1595 Old Norcross Road Apt 23 Norcross, GA

1-25 of 81

 Update profile

Another use case from ACME Inc. is when they decided to create a marketing campaign with the strategy of targeting teenagers from the ages of 10 to 20 with their new clothing and apparel line for the upcoming Spring line. To correctly target these households with the marketing campaign, the data steward should create a collection with all households that have members whose birth dates fall between 1/1/1998 and 1/1/2018.

STEP Work Flow

In the Web UI, a data steward will create a collection based off of set criteria in the Advanced Search. Once the collection is saved and selected, a Completeness Metric will be calculated based on the attributes of each record in this collection. Periodically, the data steward will update the profile to reflect any of these changes. Then, using the data on these records, various widgets will be configured to provide easy to see information.

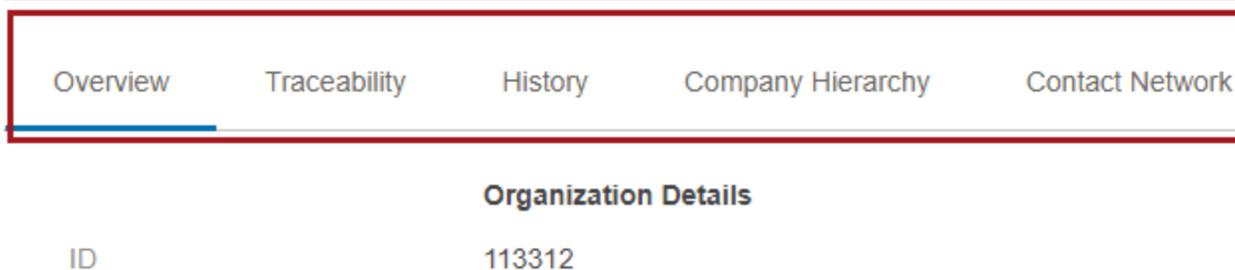
Extra Considerations

It is recommended to schedule a data profile update so that any of the changes to the customer records will be reflected in the collection. For scheduling updates, see the **Scheduling Bulk Updates** topic of the Bulk Update documentation.

Best Practices and Considerations when Configuring Web UI

When configuring any Web UI, consult the **Web UI** documentation. This documentation contains details on how to add various components of the Web UI.

Sub-screens, when configuring data screens, are very useful configuration methods that create tabs on a data entry's page. In the following screenshot, each of the highlighted items was configured as a sub-screen.



When using a references rather than data containers, the details screen is configured for that referenced object type. For information on the data containers, see the **Data Containers** topic in the System Setup documentation. Data containers allow an object to associate with multiple attribute values, for example a data container of all the addresses. Also, when configuring the root nodes and folders, ensure that the Web UI screen for each child object is configured as a node list. All the same, each dialog list screenshot be configured to 'Find Similar' records.

When creating the structure, use a Display Attribute Group rather than individual party data. This allows users to use the root node rather than each child node when configuring objects. To avoid tree navigation, which is only used to view the Collections root node, ensure that at least two searches are configured, one for simple searches based on ID and one for advanced searches with more configurations. For more information on searching in the Web UI, see the **Search Widget** topic in the Web UI documentation.

CMDM Configuration Guide

The purpose of this document is to provide an implementation party a detailed guide to the installation of the starter package configurations in a freshly installed STEP environment. This document should be interpreted as a complimentary peer to the collective enablement collateral provide by the CMDM Implementation Guidelines.

STEP System Components

This document assumes the target system is on the latest release of STEP 8.3.

The following components are required to be activated in your STEP system prior to beginning configuration within the application:

1. Party Data Matching (partydatamatching)
2. Company Data Visualization (companydata-visualization)

Note: Other CMDM-related features require dedicated components which are not in the scope of the CMDM Implementation Guide. Dedicated documentation for these components is available separately:

- Local Loqate (local-loqate)
 - Experian Email Validation (experian)
 - Dun & Bradstreet Company Match and Profiling (dnb-integration)
-

Deployment of STEP components require adequate working knowledge of patching STEP environments. For deploying STEP components it is recommended to solicit the assistance of a Stibo Systems developer or Technical Services representative.

STEPXML Configuration Files

The following is the complete list of the STEPXML import files required to provide a clean STEP system with the necessary configurations for a functional, base CMDM solution.

File Name	File Content
sharedconfig.properties	System Configuration Properties
Link, Reference, and Object Types.xml	Link, Reference, and Object Types
List of Values.xml	List of Values
Attributes and Attribute Groups.xml	Attributes and Attribute Groups
Setup Entities.xml	Setup Entities
Setup Groups.xml	Setup Groups

File Name	File Content
Data Containers.xml	Data Containers
Entity Root Nodes.xml	Entity Root Nodes
Classifications.xml	Classifications
Business Rules and Libraries.xml	Business Rules and Libraries
Component Models.xml	Component Models
Import Configurations.xml	Import Configurations
Transformation Lookup Tables.xml	Transformation Lookup Tables
Match Codes.xml	Match Codes
Match Algorithms.xml	Match Algorithms
Integration Endpoints.xml	Integration Endpoints
Users and User Groups.xml	Users and User Groups
Event Processors.xml	Event Processors
Collection Definitions.xml	Collection Definitions
Workflows.xml	Workflows
WebUI.xml	WebUI
SampleContacts.xlsx	Sample Contacts Data
SampleIndividuals.xlsx	Sample Individuals Data
SampleOrganizations.xlsx	Sample Organization Data
Sample_Pair_Formatter_ver3.1	Sample Pair Formatter Excel Sheet

STEPXML Import Sequence

Due to the nature of STEP configuration dependencies, to properly import the STEPXML configuration files a strict sequence of import must be adhered to. Please note the sequence of tasks presented in this section contains

both STEPXML files to be imported as well as manual configuration tasks that are required but not captured by STEPXML.

Furthermore, it is strongly recommended to only import the latest released version of each STEPXML file. To determine whether a newer version is available, please contact Stibo Systems.

1 - Shared Config Properties File

Each STEP environment contains unique properties specific to the environment, hence, no sharedconfig.properties file will be provided as part of this Starter Package. The following CMDM properties need to be manually added to the existing sharedconfig.properties file.

The following are standard Loqate Cloud properties that must be added in order for Loqate Cloud integration to function properly. These properties assume the deployment of the Loqate Cloud solution. For details on configurations for the Loqate Local solution, please refer to the Loqate set-up guide.

Address.AddressQualityPlugin=LoqateAddressQuality

Address.Service.Loqate.License=[Please obtain a License key for Loqate Cloud]

Address.Service.Loqate.GeocodeAll=[true or false]

Note: The Loqate.Geocode property is Optional. The default value is true. Specifies if all addresses should be geocoded. A geocode is a pair of coordinates for the latitude and longitude of a location.

Additionally, for the STEP environment to access Loqate servers for validation through the cloud, the following URL must be added to the firewall to ensure connectivity:

<http://saas.loqate.com>

Once these properties are in place, a restart of STEP is required. Finally, if a pre-existing instance of STEP is already running on your local computer then it is recommended to relaunch STEP.

2 - Link, Reference, and Object Types

The first step after modifying the sharedconfig properties file is to ensure basic object types are available in the system.

This file includes not only references but also metadata attributes associated with references. Since attributes are not yet available in the system, it is expected that the import process will encounter some errors pertaining to missing attribute IDs which may be ignored. The reason for these errors is that Object Types must be present in order for attribute validity within the *Attributes and Attribute Groups* file to properly import. Once Attributes have been imported, a re-import of Object Types will take place to establish the metadata attribute associated with references.

3 - List of Values

This file contains all CMDM-specific Lists of Values.

4 - Attributes and Attribute Groups

Attributes, including all metadata attributes, will be imported via this file.

5 - Link, Reference, and Object Types

This is a re-import of the same file from import sequence #2 above. The reason is to ensure all metadata attributes for references are accounted for now that attributes are present in the system.

6 - Setup Groups

Setup groups are the root nodes within the System Setup tab. All root nodes must be present before dependent Setup Entities are imported.

7 - Setup Entities

This file contains CMDM-specific Setup Entities.

8 - Data Containers

Data containers utilize a number of attributes that should now be present in the system.

9 - Entity Root Nodes

This file provides a base entity structure within the Tree Navigation tab of the STEP application. Also included are basic entities such as default source systems and hierarchy visualization aggregate objects.

10 - Classifications

A specific classification folder is required to support transformation look-up tables.

11 - Business Rules and Libraries

All business rules and business rule libraries required for the starter package.

12 - Component Models

Included in the Component Models STEPXML file are components for Address, Party Data Matching, Linked Golden Record, and Merged Golden Record components. Importing the Component Models STEPXML file means the user is no longer required to deploy each individual component model as all necessary attributions are included. However, if it is determined that separate attributes are more desirable to a client / on-going project, then manual deployment of each component model is also available.

13 - Import Configurations

Three import configurations are provided and are associated with each of the three customer entity types: Individuals, Organizations, and Contact Persons. These import configurations provide attribute mappings for the corresponding sample files and are utilized by the match and merge importer.

Since Household entities are generated within the CMDM solution, there is no import configuration for Households.

14 - Transformation Lookup Tables

An equivalent names table is provided with basic and common name equivalents. It is recommended that clients and implementation parties maintain and expand this table with new entries as needed.

15 - Match Codes

Match codes for the four customer entity algorithms.

16 - Match Algorithms

Four algorithms for Individuals, Organizations, Household, and Contact Persons.

17 - Integration Endpoints

Upon importing the Integration Endpoints STEPXML file, the endpoints will need to be manually invoked in order for the hotfolders to be created on the filesystem.

18 - Users and User Groups

A dedicated 'Data Stewards' user group is provided. However, no specific privileges are given to this user group. It is recommended the clients and/or implementation parties determine the specific privileges the data stewards require.

19 - Event Processors

A matching and address standardization event processor is provided. Event processors and queue status' must be manually set to 'Enable' and 'Read,' respectively.

20 - Collection Definitions

A number of collection definitions are provided to support data profiling within the WebUI

21 - Profile

Data Profile configurations are not supported by STEPXML, hence a manual task to configure the completeness metrics with each collection is required.

The following configurations are done from the data profiles tab of a collection after entering the settings. Dashboard and profile configurations both require manual configuration.

All Individual Collection

Profile Configuration

1. Add individual Completeness under 'Select Completeness Metrics to profile'
2. Add Individual Completeness attribute group under 'Include Data in Attribute Groups'

Dashboard Configuration

3. Object type count
4. Least complete attribute
5. Most complete attribute
6. Average Object completeness

- a. Right click > Configure > Set completeness metric to Individual Completeness and Intervals to 10 (Or desired interval)

All Objects Collection

Profile Configuration

1. Add General Metric Completeness under 'Select Completeness Metrics to profile'
2. Add Individual Completeness, Organization Completeness, Household Completeness, Contact Completeness attribute group under 'Include Data in Attribute Groups'

Dashboard Configuration

3. Object type count
4. Least complete attribute
5. Most complete attribute
6. Average Object completeness
 - a. Right click > Configure > Set completeness metric to General Metric Completeness and Intervals to 10 (Or desired interval)
7. Attribute Values Profile
 - a. Right click > Configure > Set Attribute ID to 'CalcAQI' and chart type to Pie Chart

Recently Updated Contact Persons Collection

Profile Configuration

1. Add Contact Completeness under 'Select Completeness Metrics to profile'
2. Add Contact Completeness attribute group under 'Include Data in Attribute Groups'

Dashboard Configuration

3. Object type count
4. Least complete attribute
5. Most complete attribute
6. Average Object completeness
 - a. Right click > Configure > Set completeness metric to Contact Completeness and Intervals to 10 (Or desired interval)

Recently Updated Households Collection

Profile Configuration

1. Add Household Completeness Metric under 'Select Completeness Metrics to profile'
2. Add Household Completeness attribute group under 'Include Data in Attribute Groups'

Dashboard Configuration

3. Object type count
4. Least complete attribute
5. Most complete attribute
6. Average Object completeness
 - a. Right click > Configure > Set completeness metric to Household Metric Completeness and Intervals to 10 (Or desired interval)

Recently Updated Organizations Collection

Profile Configuration

1. Add Organization Completeness under 'Select Completeness Metrics to profile'
2. Add Organization Completeness attribute group under 'Include Data in Attribute Groups'

Dashboard Configuration

3. Object type count
4. Least complete attribute
5. Most complete attribute
6. Average Object completeness
 - a. Right click > Configure > Set completeness metric to Organization Completeness and Intervals to 10 (Or desired interval)

Recently Updated Individuals Collection

Profile Configuration

1. Add Individual Completeness under 'Select Completeness Metrics to profile'
2. Add Individual Completeness attribute group under 'Include Data in Attribute Groups'

Dashboard Configuration

3. Object type count
4. Least complete attribute
5. Most complete attribute
6. Average Object completeness
 - a. Right click > Configure > Set completeness metric to Individual Completeness and Intervals to 10 (Or desired interval)

22 - Workflows

A total of four clerical review workflow are provided; one for each of the four customer entity types.

23 - Web UI

A basic Web UI is provided to assist data stewards with their day to day responsibilities of clerical review, manual maintenance of customer data, and data monitoring.

24 - Sample Data Load Files

Three sample data Excel files are available for each of the three customer entity types to be imported via the match and merge import configuration (i.e. Individuals, Organizations, and Contact Persons).

The names of these sample data files are:

- SampleContacts.xlsx
- SampleIndividuals.xlsx
- SampleOrganizations.xlsx

Note: Users should import SampleOrganizations.xlsx before importing SampleContacts.xlsx.

Sample data files may be imported with the Match and Merge Importer via the Asset Importer Widget available within the Web UI.

CMDM Starter Package Configuration Delivery Mechanism

All STEPXML configuration and sample data files will be provided by Stibo Systems CMDM team in the form of a zip file. Subsequent updates to the configuration files will be made available upon request.

Important: Any CMDM Starter Package configuration files received from sources other than the Stibo Systems CMDM team should not be trusted as new revisions may be available. Please contact the Stibo Systems CMDM team for the latest versions.
