



COmanage™

Managing duplicates with COmanage Match

Update: Fall 2023

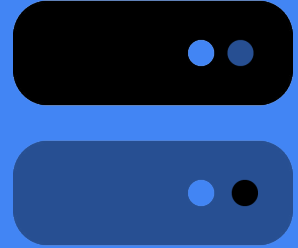
Introductions

What do you hope to get out of this workshop?

- Name
- Affiliation
- Why COmanage Match?

Assigning Virtual Machines

<https://node#.com/manage.incommon.training>



1 Shayna	2 Laura	21 Derrick J	22 Julio P	23 Debbie N
24 Jeri L	25 Rebekah A	26 Johnny L	27 David B	28
29	30			

About

Some background about Match

- The COmanage Project
- What is COmanage Match?

The C0manage Project

- Open Source Project with generous ongoing support from InCommon
- GOAL: Streamlining digital lifecycle management for your populations
- Two tools
 - **C0manage Registry**: Identity enrollment & Lifecycle Management. Used as central person registry, guest management system, or collaboration hub for scholarly collaborations.
 - **C0manage Match**: Inbound Record Cross-system Matching. Can be used with Registry or standalone to complement your other systems.

C0manage Match

v1.1.0, January 2023

- Matching across multiple authoritative systems of record (SORs)
- Heuristic-based system
- Flexible rule configuration
- Matching attributes defined by you
- Use w/Registry or standalone
- A full API

1 | Getting to Know Match

A Match Orientation

- Managing Duplicates
- CManage Match Capabilities
- The CManage Match Structure

Example Use Cases

- Information about the same person
- Tricky matching
 - Fuzzy matching
 - Same, but different (alternate forms of the same information)
 - Information stored in different fields
- Matching legacy information to current systems
- Non-person data matching

Terminology

- **System of Record (SoR)** - authoritative source of information for data set
- **Record** - a structured set Attributes values that represent a Data Subject
- **Data Subject** - The person, place, or thing described by a Record in a SoR
- **Attributes** - Data fields used to describe the Data Subject in a SoR
- **Matching Rules** - A set of instructions used to compare Attribute values in Records to determine if they describe the same Data Subject.

About Identity Matching

- Coordinated vs independent attributes
- Synchronous vs asynchronous match resolution
- Implementation at registry vs standalone
- Match before registry vs match at registry

About Identity Matching

Coordinated vs independent attributes

What attribute values are compared to during matching.

Coordinated Attributes

- “Golden” attributes
- All SoRs agree to use these attributes
- Matching is done against the golden set

Independent Attributes

- Each SoR is authoritative for its attributes
- Matching may be against a canonical representation of attributes
- Alternatively, matching may be against each SoR’s representation of attributes

About Identity Matching

Synchronous vs asynchronous resolution

When matches are resolved if there is a partial/potential match.

Synchronous Match Resolution

- Resolved by the person providing the data
- Resolution required before proceeding

Asynchronous Match Resolution

- Resolved by an “administrator”
- Notification sent for resolution later
- Must be used for batch resolution

About Identity Matching

Implementation at registry vs standalone

How the matching service is integrated into your workflow.

Implementation at Registry

- Identity matching is integrated into the person registry
- COmanage Registry and other tools have integrated Match

Standalone Implementation

- The Match engine independently provides information about matches
- The results of matches can be viewed or consumed by your systems

About Identity Matching

Match before registry vs match at registry

When is matching done? Before or after the record is in the registry

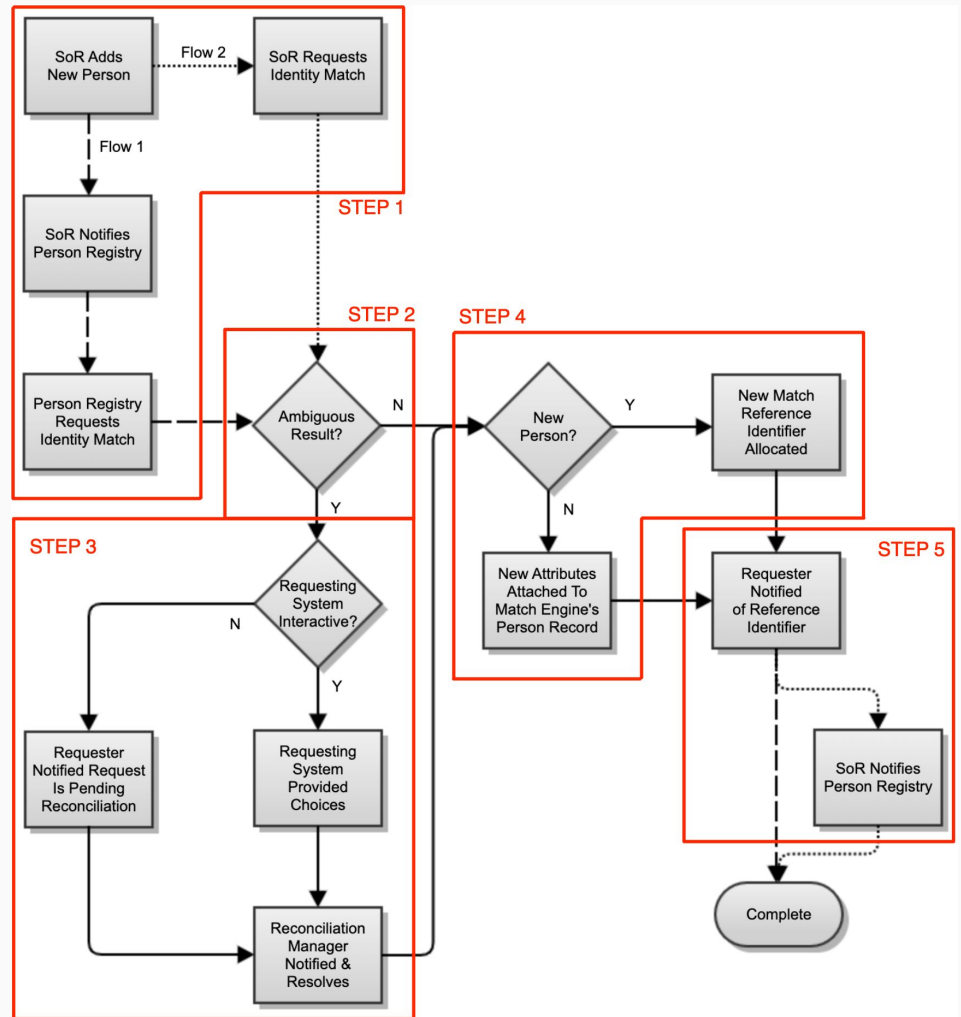
Match Before Registry

- Matching before the record is added
- May happen during an enrolment flow
- Useful if you don't want the SoR record added if there is a match

Match At Registry

- Matching at the time of adding the record
- Information from the source is added first and then matched to aggregate it

Identity Match Process Flow



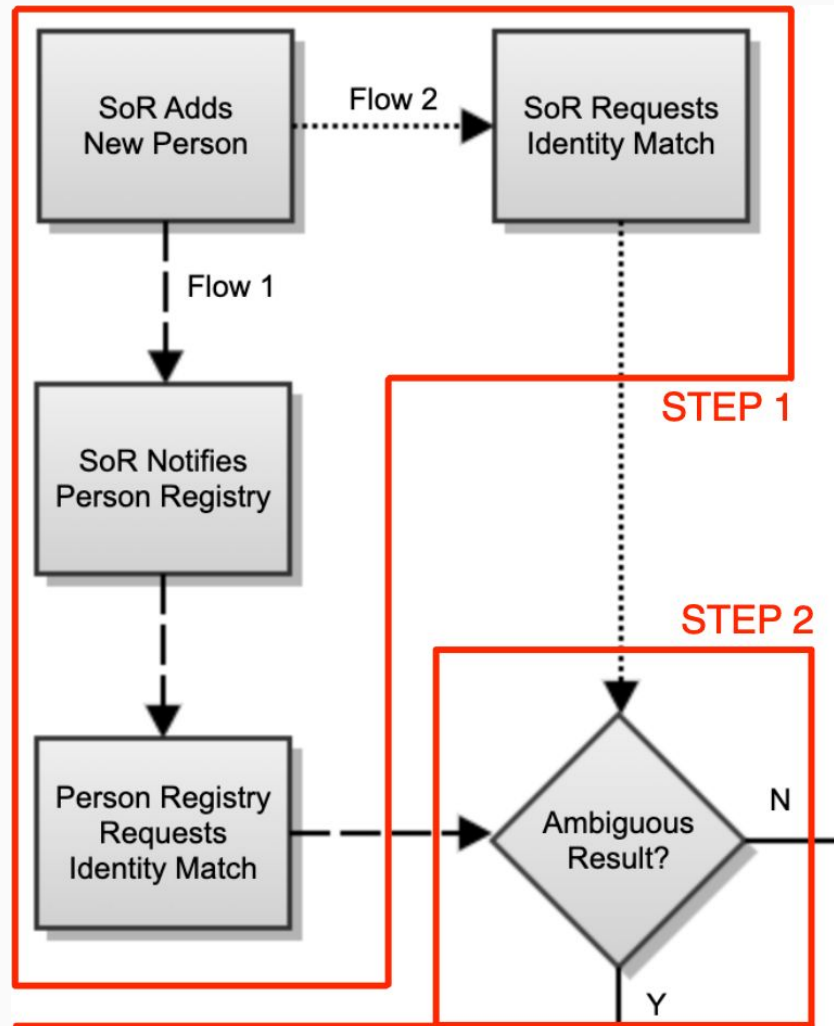
Identity Match Process Flow

STEP 1

New Record from a SoR is ready for Matching

FLOW 1 The System of Record provides a record for new data subject (person) to the (person) registry. The registry sends the information to Match for evaluation. This information is called an "Inbound Record".

FLOW 2 The System of Record (or other system) directly provides a record for new data subject (person) to Match for evaluation. This information is called an "Inbound Record".

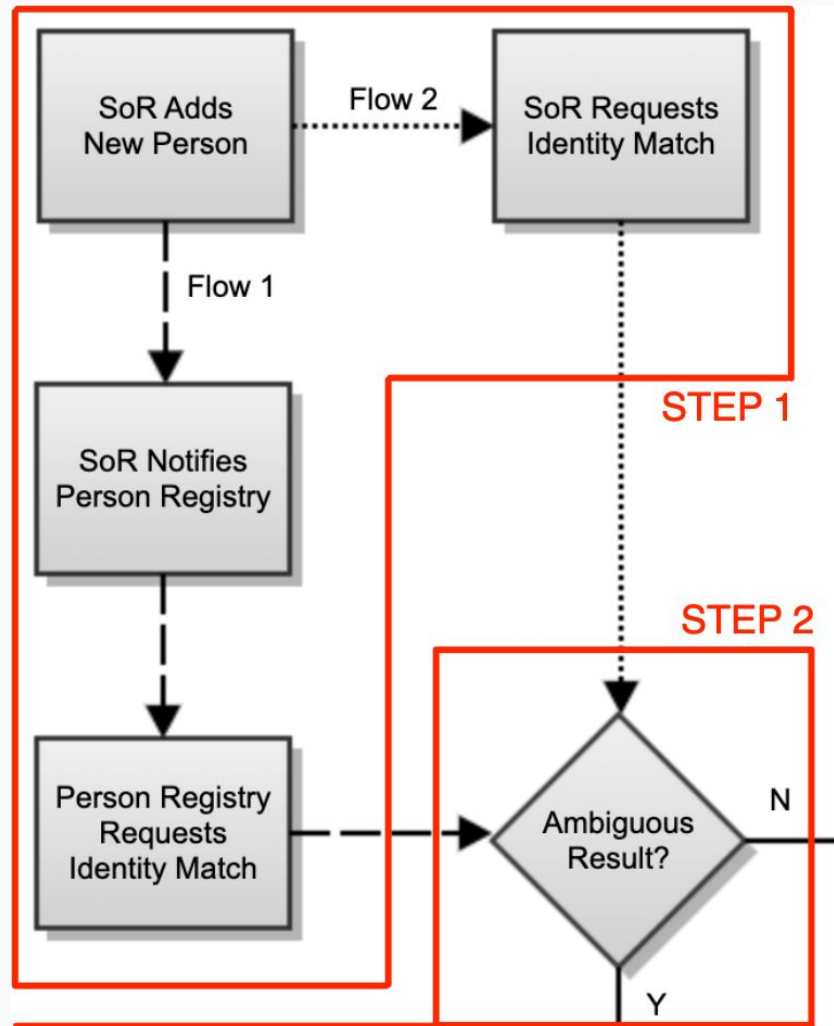


Identity Match Process Flow

STEP 2

Match Rules are evaluated

The Match engine searches for matching Records from the set it knows about. For each Inbound Record, the rules will produce either an ambiguous or unambiguous result. Records with ambiguous results will proceed to STEP 3; those with unambiguous results will proceed directly to STEP 4.



Identity Match Process Flow

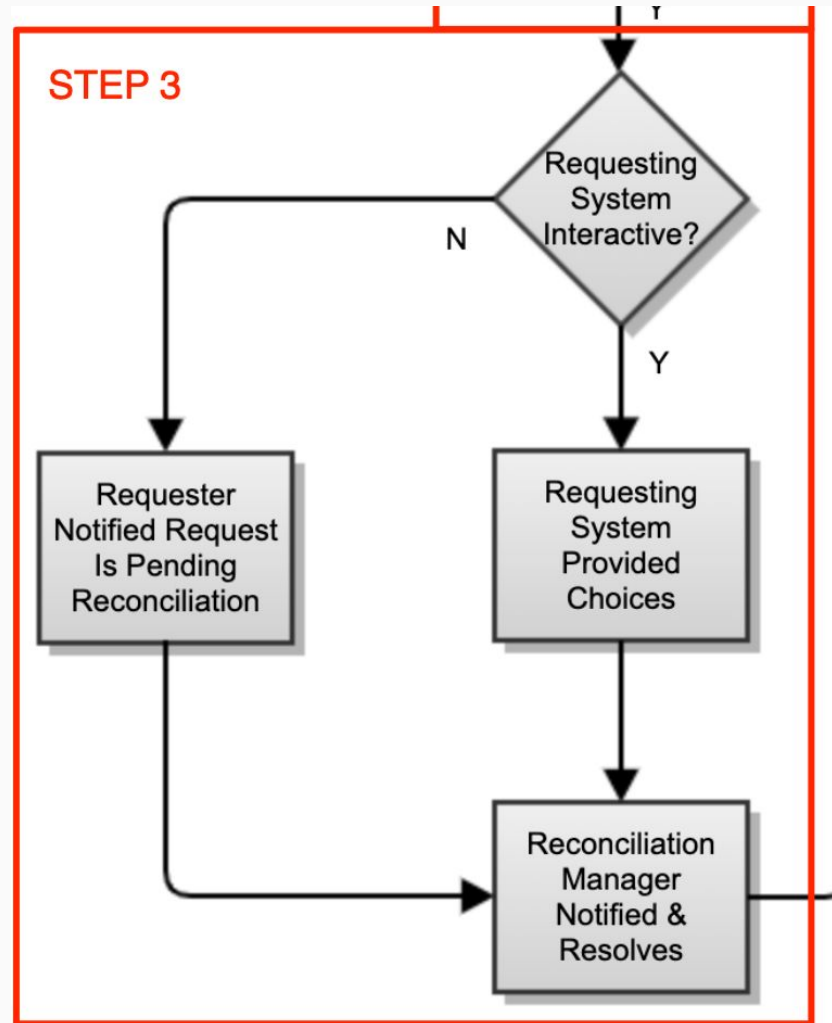
STEP 3

Potential Matches are Resolved

If the rules make an ambiguous match, an individual needs to resolve it. Match can resolve synchronously (interactive) or asynchronously.

ASYNCHRONOUS - A notification is sent to the configured email address that a potential match needs to be resolved. An individual will resolve the match before this Record will be assigned a Match Reference ID, but this resolution doesn't have to be done right away.

SYNCHRONOUS - The system requesting the match is sent data to be presented to the individual to resolve the match right away. Note that this flow, if chosen, requires integration with a system outside of COmanage Match.



Identity Match Process Flow

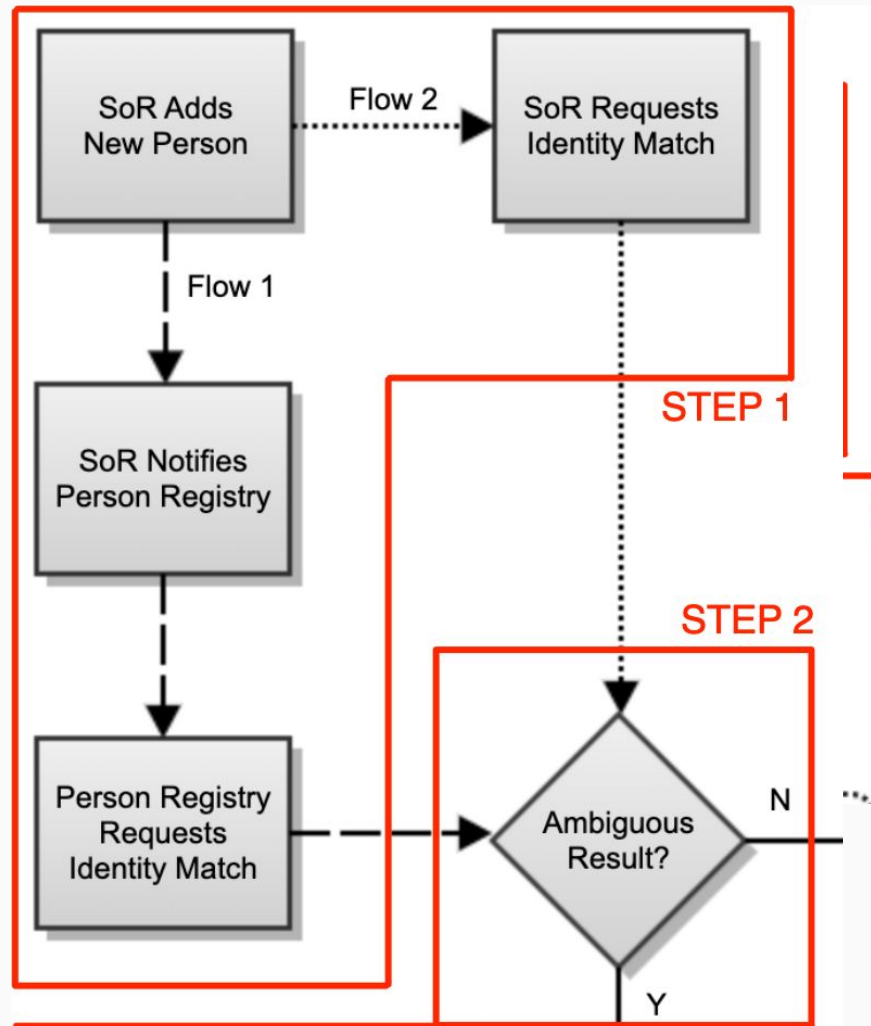
STEP 4

Match Reference Identifier is Assigned

Based on an unambiguous result from the match rules, or a resolve match by an individual, Match will assign a match reference identifier:

NEW RECORD - if the Inbound Record represents a unique data subject (person), Match will assign it a new reference identifier.

MATCHED RECORD - if the Inbound Record is determined to represent the same data subject (person) as one already identified, Match will assign the Inbound record the reference identifier of the matched existing Record. NOTE: the records remain distinct. The presence of the same reference identifier indicates that they represent the same data subject (person).

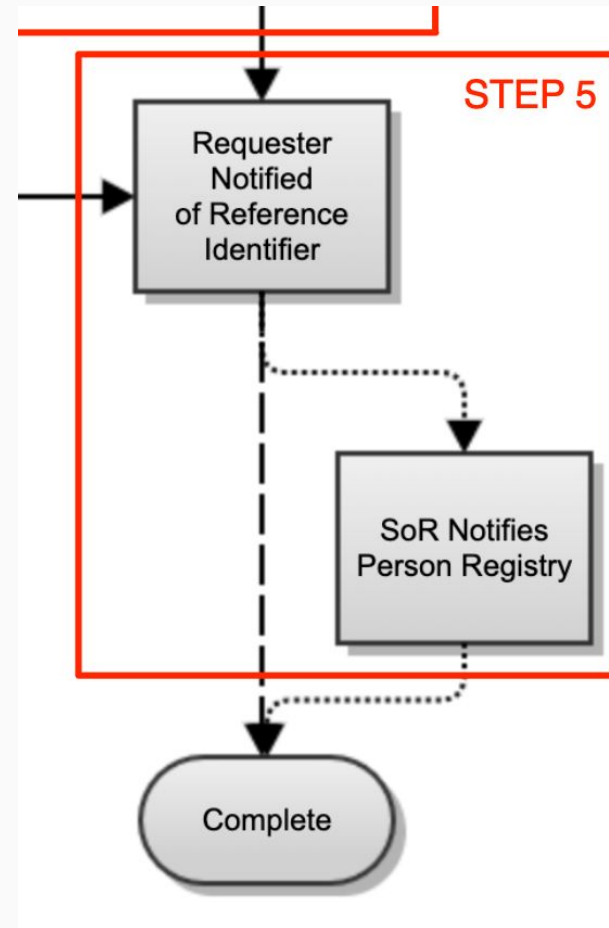


Identity Match Process Flow

STEP 5

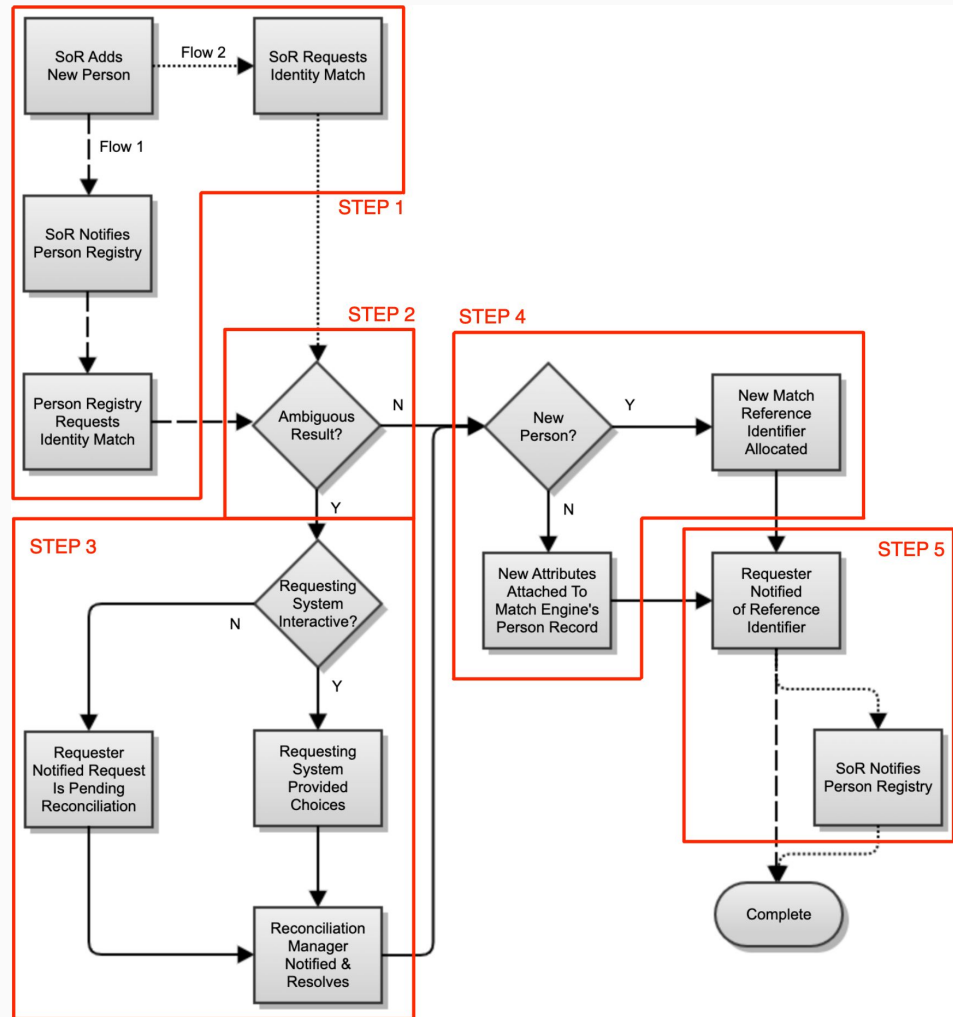
Notifications are Sent

Once a match reference identifier has been assigned to an Inbound Record, the system requesting the match analysis is sent a notification containing the reference identifier for the Record. For FLOW 1, the requesting system is the (person) registry. Match stores records that have been matched so they are available for matches to subsequent new Inbound Records.



Identity Match Process Flow

1. New record from a SoR is ready for matching
2. Match rules are evaluated
3. Potential matches are resolved
4. Match reference identifier is assigned
5. Notifications are sent



The Matchgrid

Where everything starts

Each Match tenet is a “Matchgrid”

The screenshot shows the COmanage Match web interface. At the top, a green header contains the text "COmanage Match". Below this, a breadcrumb trail reads "COmanage Match > Matchgrids". The main heading is "Matchgrids". On the right, a navigation menu is open, showing options: "Matchgrids" (selected), "Permissions", and "API Users". A red box highlights this menu. Below the heading, there is a button labeled "+ Add New Matchgrid". A table lists the matchgrids. The first entry is "demo" with status "Active". The table has columns for "Table Name", "Status", and "Action". The "Action" column contains links for "Edit", "Manage", and "Delete". Red annotations with arrows point to specific elements: "1" points to the "Matchgrids" menu item, "2" points to the "Matchgrids" breadcrumb, and "Create Matchgrids" points to the "+ Add New Matchgrid" button.

Platform admin

COmanage Match

COmanage Match > Matchgrids

Matchgrids

Matchgrids

Permissions

API Users

+ Add New Matchgrid

Manage Matchgrids 1

2

Create Matchgrids

Table Name	Status	Action
demo	Active	Edit Manage Delete

1 of 1

Add a Matchgrid

Add New Matchgrid

Table Name *

Unique name for matchgrid, must be a valid SQL identifier (will be prefixed mg_ for actual table name)

Description

Status *

Active



** Denotes Required Field*

SAVE

Match Users / Permissions

The screenshot shows the 'CManage Match' interface. At the top right, a dropdown menu is open, showing 'Matchgrids', 'Permissions' (highlighted with a red box and arrow), and 'API Users'. Below the header, the breadcrumb 'CManage Match > Permissions' is visible. The main heading 'Permissions' is also highlighted with a red arrow. To the right of the heading is a button '+ Add New Permission' with a red arrow pointing to it. Below this is a table with columns: Username, Matchgrid, Permission, and Action. The table contains two rows: 'admin' with 'Platform Administrator' and 'demouser' with 'Matchgrid Administrator'. Each row has 'Edit' and 'Delete' links in the Action column. A green footer bar at the bottom right shows '1 of 1'.

Platform admin

CManage Match

CManage Match > Permissions

Permissions

Matchgrids
 Permissions
 API Users

Add New Permission

Username	Matchgrid	Permission	Action
admin		Platform Administrator	Edit Delete
demouser	demo	Matchgrid Administrator	Edit Delete

1 of 1

Match Users / Permissions

CManage Match > Permissions > Edit demouser

Edit demouser

Username *	<input type="text" value="demouser"/>
Permission *	<input type="text" value="Matchgrid Administrator"/>
Matchgrid *	<input type="text" value="demo"/>

** Denotes Required Field*

SAVE

Enter the Matchgrid

COmanage Match



Matchgrid Selection

Select the Matchgrid you would like to work with.

Matchgrid	Description
demo	Demo Matchgrid for use with Demo Registry

How CManage Match indicates a match


Match Reference ID

- Two possible formats
 - Sequence
 - UUID (type 4)
- Records that match will be assigned the same Reference ID

Matchgrid Settings

Matchgrid: demo



 Manage

 Display

 Reconcile

 Configure

COmanage Match > demo > Configure demo

Configure demo

 [Matchgrid Settings](#)

 [API Users](#)

 [Attributes](#)

 [Attribute Groups](#)

 [Attribute Maps](#)

 [Rules](#)

 [Systems of Record](#)

 [Build](#)

Matchgrid Settings

Matchgrid: demo



 Manage

 Display

 Reconcile

 Configure

[COmanage Match](#) > [demo](#) > [Configure](#) > [Matchgrid Settings](#) > Edit Matchgrid Settings

Edit Matchgrid Settings

**Reference ID Assignment
Method**

*

Notification Email

*On potential match, notify this
address of the request requiring
resolution (used if no SOR specific
value is set)*

* Denotes Required Field

SAVE

2 | Systems of Record

Data comes from somewhere


- Defining Systems of Record (SORs) in Match
- Configuring Match Resolutions for SORs
- Overview: Systems of Record and the Match API

Configure Systems of Record

Matchgrid: demo



 Manage

 Display

 Reconcile

 Configure

COmanage Match > demo > Configure demo

Configure demo

 Matchgrid Settings

 API Users

 Attributes

 Attribute Groups

 Attribute Maps

 Rules

 Systems of Record

 Build

Add System of Record

Matchgrid: demo



 Manage

 Display

 Reconcile

 Configure

[COmanage Match](#) > [demo](#) > [Configure](#) > Systems of Record


Systems of Record


[+ Add New System of Record](#)


Label	Resolution Mode	Action
hrms	External	Edit Delete
sis	External	Edit Delete


1 of 1


Defining Systems of Record

Matchgrid: demo

 **Manage**

 **Display**

 **Reconcile**

 **Configure**

[COmanage Match](#) > [demo](#) > [Configure](#) > [Systems of Record](#) > [Edit hrms](#)

Edit hrms

Label *	<input type="text" value="hrms"/>
Resolution Mode *	<input type="text" value="External"/> ▼
Notification Email <i>On potential match, notify this address of the request requiring resolution</i>	<input type="text"/>

** Denotes Required Field*

SAVE

System of Record metadata

Systems of Record in Match are defined by three attributes:

- **Label** - The name that will be displayed within Match to refer to the SOR.
- **Resolution Mode** - An indication of how potential matches are handled.
- **Notification Email** - The email address where notifications should go when there is an inbound record from this source that results in a potential match that needs resolution.

Resolution Modes

Each Systems of Record will expect a specific resolution mode:

- **External:** You can think of this resolution mode as being asynchronous. Resolution is handled external to the request.
- **Interactive:** You can think of this resolution mode as being synchronous. The record's System of Record handles the resolution of potential matches at the time of submission.

3 | Match Attributes

Where comparisons are made

- What attributes to use
- The anatomy of an attribute
- Attribute groups
- Attribute maps
- Other attribute topics

Attribute schema | API name

Attributes may be simple or complex

- **Simple** - No sub attributes
- **Complex** - Contain sub attributes

Attributes may have single or multiple values

Match is an expression of an API. The attributes define the structure of the API data message.

Schema Details in the [TAP Attribute Dictionary](#)

Add and configure attributes

Matchgrid: demo



- Manage
- Display
- Reconcile
- Configure**

COmanage Match > demo > Configure > Attributes

Attributes

[+ Add New Attribute](#)

Name	Attribute Group	Action
dob		Edit Duplicate Delete
firstname	official	Edit Duplicate Delete
lastname	official	Edit Duplicate Delete
ssn		Edit Duplicate Delete

1 of 1

Attribute Configurations

Matchgrid: demo



 Manage

 Display

 Reconcile

 Configure

COmanage Match > demo > Configure demo

Configure demo

 Matchgrid Settings

 API Users

 Attributes

 Attribute Groups

 Attribute Maps

 Rules

 Systems of Record

 Build

Configure attributes

CManage Match > demo > Configure > Attributes > Edit ssn

Edit ssn

Name *

ssn



Value must be a valid SQL identifier, as it will be used to construct the matchgrid column name

Description

API Name *

identifiers:identifier/national

Display Field in Matchgrid Index

☐

Alphanumeric

☒

Case Sensitive

☐

Null Equivalents

☒

Search Distance

2



Search Exact

☒

Search Substring From



Search Substring For



Attribute Map



Attribute Group



** Denotes Required Field*

SAVE

Anatomy of an attribute

Descriptive Information

- Attribute Name
- Description
- API Name
- Attribute Group
- Display field in the Matchgrid?

Matching Traits

- Alphanumeric
- Case Sensitive
- Null Equivalents
- Search Distance
- Search Exact
- Search Substring from/for
- Attribute Map



We'll focus on these first!

Attribute Groups

Matchgrid: demo



 Manage

 Display

 Reconcile

 Configure

[COmanage Match](#) > [demo](#) > [Configure](#) > Attribute Groups

Attribute Groups

[+ Add New Attribute Group](#)


Name	Action
official	Edit Delete
1 of 1	

Attribute Mappings

Matchgrid: demo



 Manage

 Display

 Reconcile

 Configure

[COmanage Match](#) > [demo](#) > [Attribute Maps](#) > [Nicknames](#) > Attribute Mappings

Attribute Mappings

[+ Add New Attribute Mapping](#)

[↓ Install English Nicknames](#)

Query	Value	Action
aaron	erin	Edit Delete
aaron	ron	Edit Delete
aaron	ronnie	Edit Delete
ab	abbie	Edit Delete
ab	abe	Edit Delete

BREAK

4 | Developing Match Rules

What constitutes a match?

- Rules background and setup
- Configuring rules

About Match rules

- Each Matchgrid will have one or more rules
- Each rule will have a configuration for one or more attributes
- Rules are configured with a confidence mode
 - **Canonical** - definitively defines if records are matched or not
 - **Potential** - may identify matching records.
- Rules behave like searches. An attribute configuration are the settings to find the incoming attribute value in the set of records in the Matchgrid.

Processing Match rules

I. Process Canonical Rules

1. Each rule is evaluated
2. Processing stops when a rule finds one or more matches (*i.e., a search produces a results set*)
3. If exactly one match is found, the record is given the same Reference ID as the match
4. If more than one is found, it is treated as a potential match

II. Process Potential Rules

1. ALL potential rules are evaluated
2. Each record that matches (items found during a search for the attributes) is a potential match
3. If processing is synchronous, the match must be resolved: select a match, or indicate no match
4. If processing is asynchronous, a notification is sent for later resolution

Managing Rules

Matchgrid: demo



Manage

Display

Reconcile

Configure

















CManage Match > demo > Configure > Rules

Rules

Add a rule

Configure a rule


+ Add New Rule

Name	Confidence Mode	Order	Action			
Canonical Exact	Canonical	1	 Edit	 Duplicate	 Rule Attributes	 Delete
Canonical Nickname	Canonical	2	 Edit	 Duplicate	 Rule Attributes	 Delete
Potential Distance	Potential	1	 Edit	 Duplicate	 Rule Attributes	 Delete
Potential Nickname	Potential	2	 Edit	 Duplicate	 Rule Attributes	 Delete

Set Rule Metadata

Matchgrid: demo



 Manage

 Display

 Reconcile

 Configure

[COmanage Match](#) > [demo](#) > [Configure](#) > [Rules](#) > Edit Canonical Exact

Edit Canonical Exact

Name *	<input type="text" value="Canonical Exact"/>
Description	<input type="text"/>
Confidence Mode *	<input type="text" value="Canonical"/> ▼
Order	<input type="text" value="1"/> ▼

** Denotes Required Field*

SAVE

Define the search criteria for each attribute

Matchgrid: demo



Manage

Display

Reconcile

Configure

[CManage Match](#) > [demo](#) > [Rules](#) > [Canonical Exact](#) > [Configure](#) > Rule Attributes

Rule Attributes

[+ Add New Rule Attribute](#)

Attribute ▲	Crosscheck Attribute	Search Type	Action
dob		Exact	Edit Delete
firstname		Exact	Edit Delete
lastname		Exact	Edit Delete
ssn		Exact	Edit Delete

1 of 1

Searching for matches

Four search types are supported in v1.0.0 of COmanage Match:

- Exact
- Distance
- Mapping
- Substring

The search is partially configured in the attribute, and partially in the rule.

Anatomy of an attribute

Descriptive Information

- Attribute Name
- Description
- API Name
- Attribute Group
- Display field in the Matchgrid?

Matching Traits

- General
 - Search Exact
 - Alphanumeric
 - Case Sensitive
 - Null Equivalents
- Search Distance
- Attribute Map
- Search Substring from/for

Now these are important



Searching for matches

EXACT

Attribute Settings

Alphanumeric	<input type="checkbox"/>
Case Sensitive	<input type="checkbox"/>
Null Equivalents	<input type="checkbox"/>

For this attribute...

- Should non-alphanumeric characters be stripped out?
- Does case matter?
- Is a blank "" the same as null 0000-00-00?

Rule Setting

Attribute *	firstname
Crosscheck Attribute	
Search Type *	Exact
Required	<input checked="" type="checkbox"/>

Searching for matches

DISTANCE

Attribute Settings

Search Distance

2

For this attribute...

- What Levenshtein Distance should be used to find transpositions (e.g., *Smith* vs *Simth*)

Rule Setting

Attribute *

firstname

Crosscheck Attribute

Search Type *

Distance

Required



Searching for matches

MAPPING

Attribute Settings

Attribute Map

Nicknames

For this attribute...

- What data dictionary (attribute map) should be used to find matches

Rule Setting

Attribute *

firstname

Crosscheck Attribute

Search Type *

Mapping

Required



Searching for matches

SUBSTRING

Attribute Settings

Search Substring From	6
Search Substring For	4

For this attribute...

- Where in the field should the search start, and for how many characters - for example, search from 6 for 4 to compare values of the last 4 characters.

Rule Setting

Attribute *	ssn
Crosscheck Attribute	
Search Type *	Substring
Required	<input checked="" type="checkbox"/>

5 | Building Test Sets

Bedrock of effective
configuration


- Approaching tests
- Testing one record at a time
- Testing in bulk


Designing test sets


- **A baseline record:** A record against the other test records will be compared.
- **Positive test cases:** Records that you expect to result in a match.
 - the exact same record;
 - the same Attribute data but from a different System of Record;
 - data that shouldn't result in a match, but may match anyway based on the way that the rule is configured
- **Negative test cases:** Record that you would not want to result in a match.
 - an alternate spellings of values in one or more fields;
 - the same Attribute data but from a different System of Record;
 - data that may be present in a different Attribute field


Matchgrid Records


Matchgrid: demo



 Manage


 **Display**

 Reconcile










 Configure

CManage Match > demo > Matchgrid Records

Matchgrid Records



Q Filter

ID	System of Record (SOR)	System of Record ID	Reference ID	firstname	Action
3	hrms	hrms1001	B0B42CAB...		 Edit  Delete
28	hrms	hrms10011	070247FA...		 Edit  Delete
29	hrms	hrms10012	60EAD5C9		 Edit  Delete

6 | Resolving Matches

When the humans get involved

- The Match interface
- Resolutions in other tools
- How the API helps out

Match Resolution Interface

Matchgrid: training



Manage

Display

Reconcile

Configure

COManage Match > training > Pending Requests > Reconcile Unresolved Request sis/potential1001

Reconcile Unresolved Requests

Highlight: ☒ differences ☐ matches ☐ both

	New Record	Suggestion 1	Suggestion 2
Reference ID	New	8e343435-602f-482b-a7b6-52daab91bfc9	9737b31a-e75a-48a1-a7f6-929aee9f135f
Action	<button>GENERATE NEW REFERENCE ID</button>	<button>ASSIGN THIS REFERENCE ID</button>	<button>ASSIGN THIS REFERENCE ID</button>
id	152	22	153
sor	sis	hrms	sis
sorid	potential1001	N-00007	potetential1002
firstname	Bella	Isabella	Isabel
lastname	Stokes	Stokes	Stokes
dob			
nationalid		280-78-7615	
email	isa@gmail.com	Isabella.Stokes@redcurrant.edu	lstokes@redcurrent.edu
firstnamepref		Bella	
lastnamepref		Stokes	
emailpers			
request_time	2023-03-30 04:32:26	2023-03-23 01:11:35	2023-03-30 04:33:35
resolution_time		2023-03-29 19:49:50	2023-03-30 04:33:35

Resolution in other settings

- **In CManage Registry:** Match is well-integrated with Registry; resolutions can happen in the Registry interface
-
- **API:** The Match API enables matches to happen within any tool or with a custom interface.

BREAK

Your Turn

Where the REAL learning begins!

Design a (set of) rules that will ensure that the record on the right creates an **exact match** with the following record.

Hint: Distance Match

hrms N-00009

Eliana Carter

Eliana.Carter@redcurrant.edu

Ntl ID: 269-49-0033

Try developing other test cases too!

System of Record	sis
SOR ID	test5002
dob	
email	
emailpers	
firstname	Elaina
firstnamepref	
lastname	Carter
lastnamepref	
Identifier	269-49-0033

You have a system that masks the first 5 digits of the national identifier. Design a (set of) rules that will identify matches with the records in your Matchgrid that have the full number. What should the confidence level of this match be?

Hint: Substring Match

Try developing other test cases too!

System of Record	sis
SOR ID	test5002
dob	
email	
emailpers	
firstname	Abby
firstnamepref	
lastname	Jimenez-Carter
lastnamepref	
Identifier	xxx-xx-5196

Design a (set of) rules that will create a **potential match** with the following record.

Hint: Cross-attribute matching

hrms N-00184

Henry Martinez
henry.martinez@redcurrant.edu
Ntl ID: 422-01-5726

Try developing other test cases too!

System of Record	sis
SOR ID	test5003
dob	1953-07-04
email	Enrique.martinez@mailinator.com
emailpers	henry@gmail.com
firstname	Enrique
firstnamepref	Henry
lastname	Martinez
lastnamepref	Martinez
Identifier	

Use Case Brainstorm

What use cases do you have? Let's try to build rules to address them.

System of Record	
SOR ID	
dob	
email	
emailpers	
firstname	
firstnamepref	
lastname	
lastnamepref	
Identifier	

...and now...

You know some.
What about the rest?

This doesn't have to be the end of your CManage Match journey!



[Technical Manual](#)



[User Guide](#)

...and other resources

- [CManage Mailing lists & slack](#)
- [Monthly Newsletter archive](#)
- CManage Open Office Hours (monthly)
- **Open Office Hours - Monthly: 1st Friday**
Friday, Nov 3 at 12:00 ET | 9:00 PT

Feedback Please!!

<https://www.surveymonkey.com/r/Match0323>