Informatica

28 May, 2024

Match and Merge Customer Use Cases in MDM SaaS

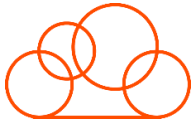
- Sourya Dass, Senior Solutions Architect
- Kamal Abrol, Senior Principal Customer Success Architect

Where data & AI come to LIFE

Housekeeping Tips

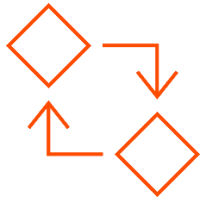
- Today's Webinar is scheduled for 1 hour
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the Q&A option and we will respond at the end of the presentation
- The webinar is being recorded and will be available on our Success Portal - where you can download the slide deck for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the post-webinar survey and provide your feedback and suggestions for upcoming topics.

Feature Rich Success Portal



Bootstrap trial and POC Customers

Enriched Customer Onboarding experience



Product Learning Paths and Weekly Expert Sessions

Informatica Concierge

Tailored training and content recommendations

More Information

Success Portal

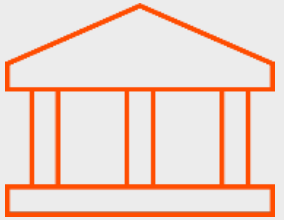
https://success.informatica.com

Communities & Support

https://network.informatica.com

Documentation

https://docs.informatica.com



University

https://www.informatica.com/in/services-and-training/informatica-university.html

Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

Industry Centric Data Model & Match Strategy

Thousands of organizations across more than 35 industries have implemented Informatica Intelligent MDM and 360 applications, services of Intelligent Data Management Cloud™ (IDMC). Based on this extensive experience, we came up with five steps to develop the right strategy to implement your MDM solution most effectively:

1. Define your business requirements.
2. Translate business requirements into technical requirements.
3. Select the rights tools and vendors.
4. Execute the project.
5. Evaluate your results.

Identify the business opportunities

Get input from your business stakeholders and leaders to determine the key business objectives they wish to support, the challenges they face and the potential opportunities MDM can deliver.

Write down the answers to these questions:

- What are the goals for the business leaders?
- What challenges are they experiencing that impede success?

Examples include:

- Deliver relevant experiences to attract, retain and grow customer base
- Reduce procurement costs
- Accelerate financial close
- Increase cross- and upselling
- Increase ROI from marketing campaigns
- Reduce supplier spend and management cost
- Accelerate new product introduction
- Reduce risk associated with large modernization efforts such as ERP

Define functional requirements

- Functional requirements describe what the software should do. For example:
- The MDM solution should automatically treat two customer records as duplicates if the names, dates of birth, addresses and phone numbers match with a 90% confidence level.
- Data stewards should be able to update multiple records at once to process them quickly and in a
- controlled manner.
- Business users should be able to capture and visualize person-to-person, person-to-company and company-to-company relationships along with historical views.

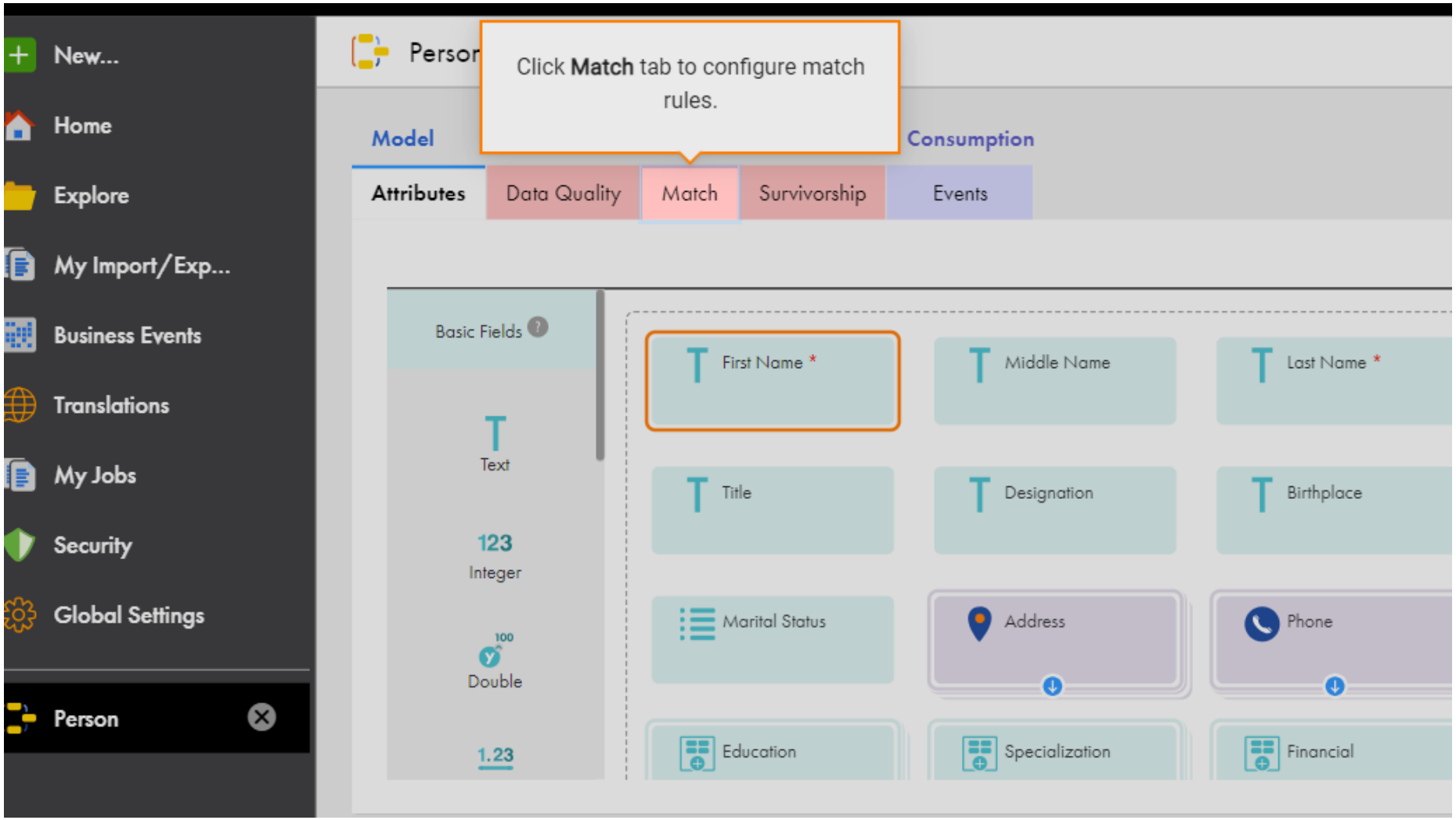
Before you configure match and merge, ensure that you meet the following prerequisites:

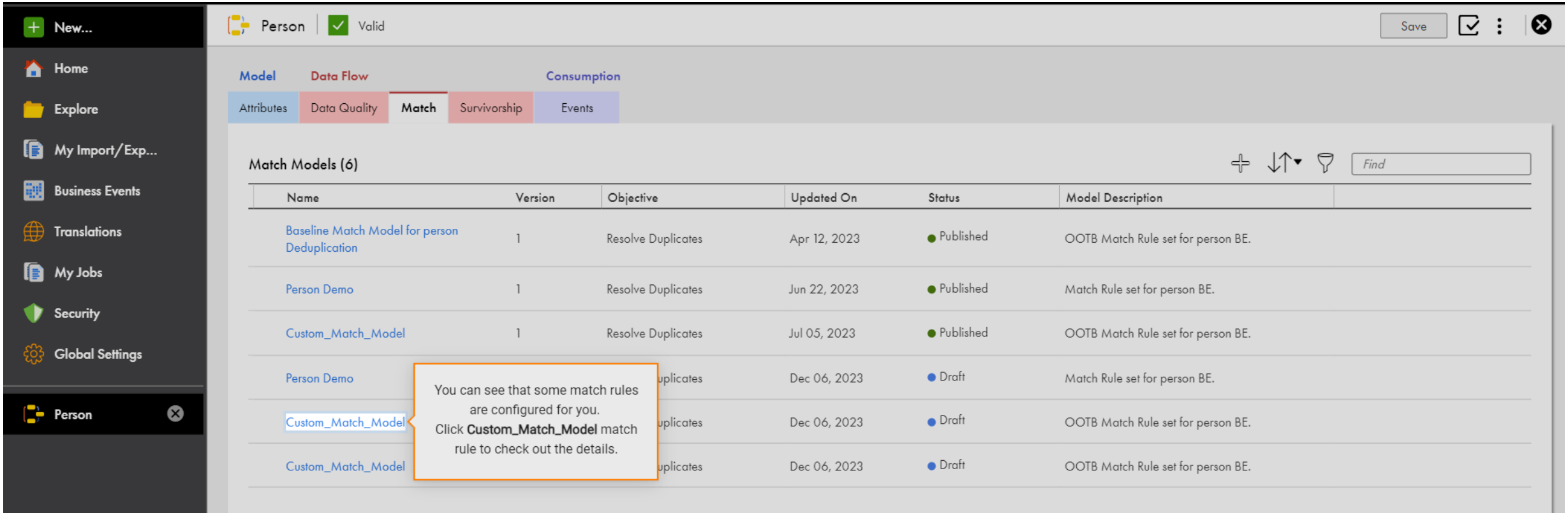
• Understand the business objectives and requirements for matching and merging duplicate data.

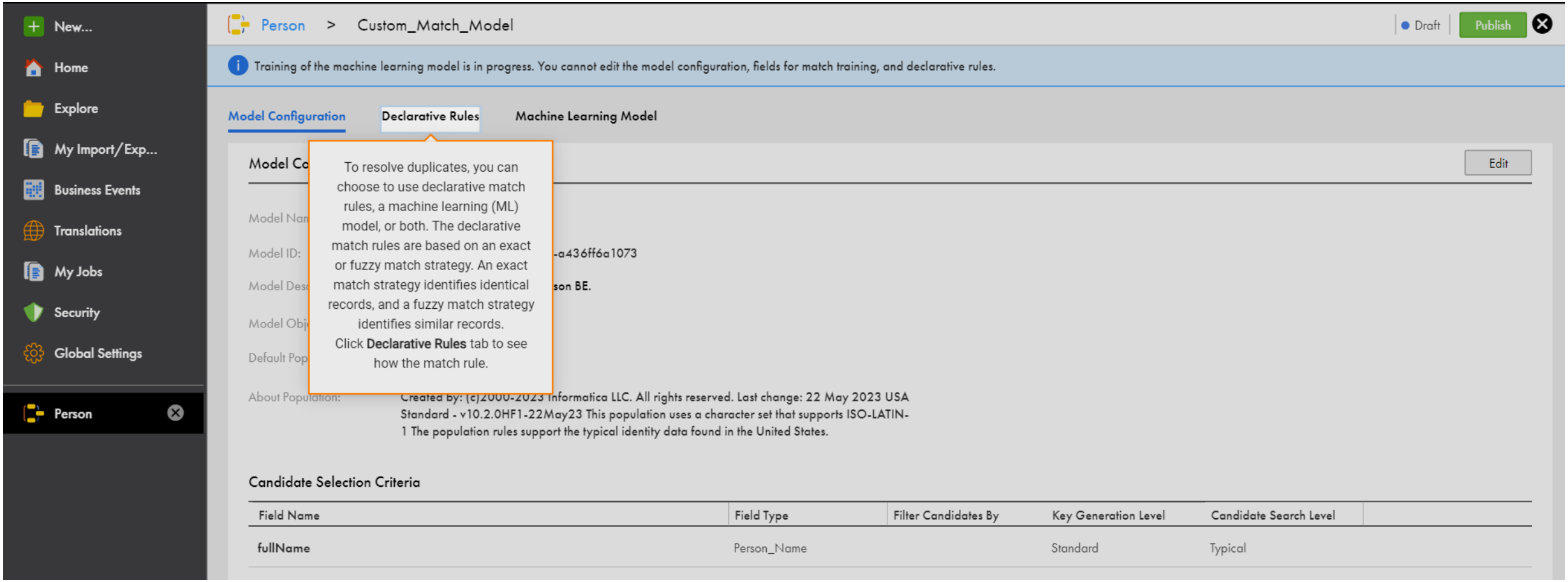
• Analyze the attributes and quality of the data that you want to consider for the match and merge process.

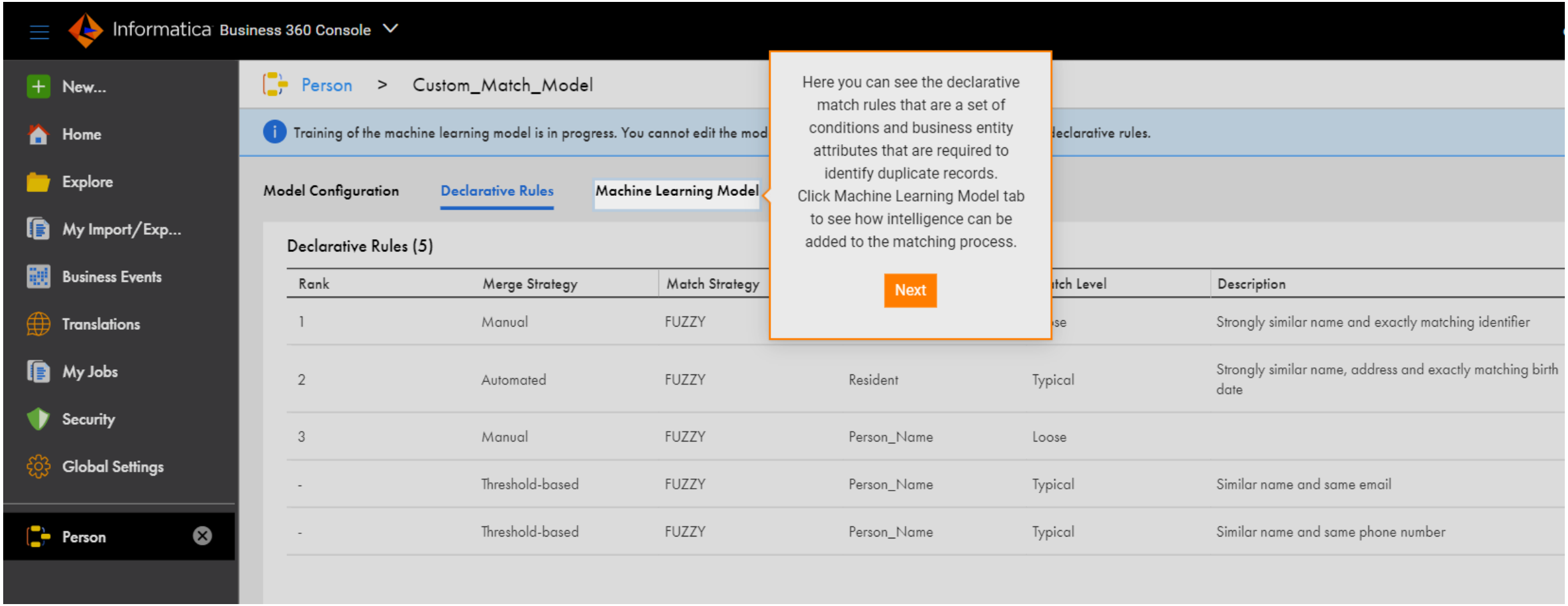
• Determine the size of the data set, which could impact the performance of the match and merge process.

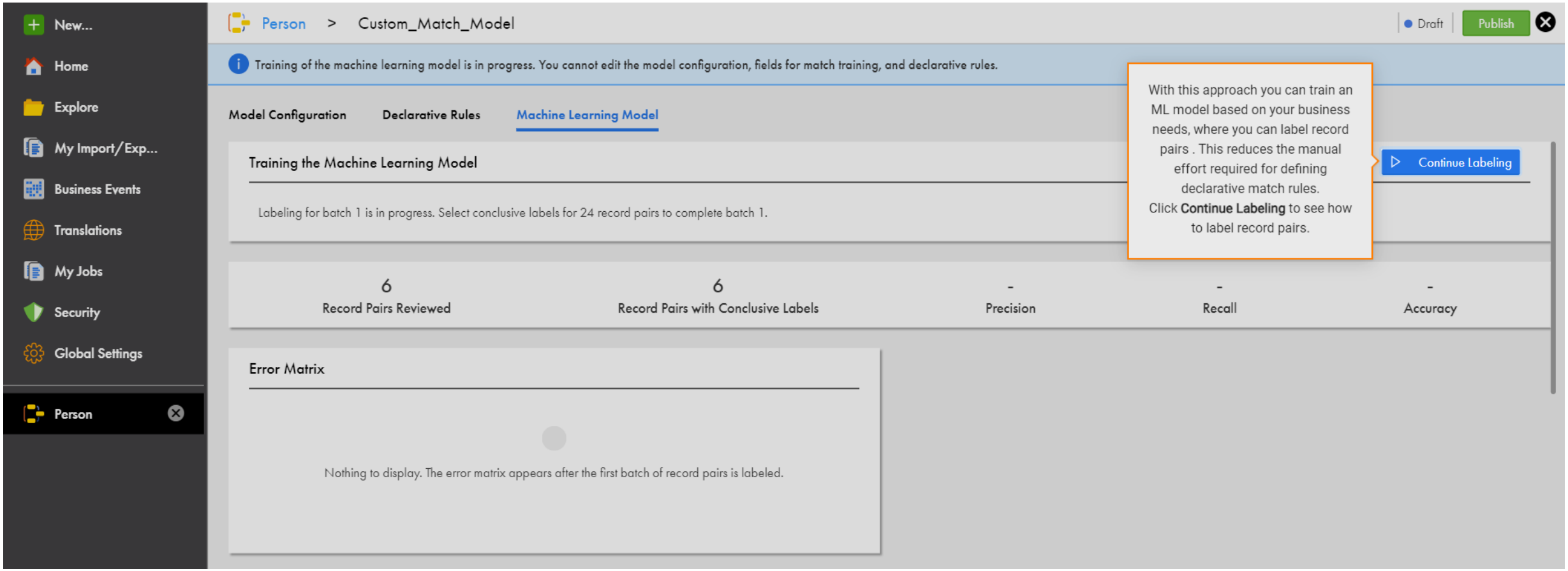
• Determine the match population, which improves match accuracy by accommodating variations and errors that are likely to appear in data for a particular population.


New...
Home
Explore
My Import/Exp...
Business Events
Translations
My Jobs
Security
Global Settings
Person
Perso
Click Match tab to configure match rules.
Model
Consumption
Attributes
Data Quality
Match
Survivorship
Events
Basic Fields
Text
123
Integer
Double
1.23
First Name *
Middle Name
Last Name *
Title
Designation
Birthplace
Marital Status
Address
Phone
Education
Specialization
Financial


New...
Home
Explore
My Import/Exp...
Business Events
Translations
My Jobs
Security
Global Settings
Person
Person
Valid
Save
Model
Data Flow
Consumption
Attributes
Data Quality
Match
Survivorship
Events
Match Models (6)
Find
Name
Version
Objective
Updated On
Status
Model Description
Baseline Match Model for person Deduplication
1
Resolve Duplicates
Apr 12, 2023
Published
OOTB Match Rule set for person BE.
Person Demo
1
Resolve Duplicates
Jun 22, 2023
Published
Match Rule set for person BE.
Custom_Match_Model
1
Resolve Duplicates
Jul 05, 2023
Published
OOTB Match Rule set for person BE.
Person Demo
uplicates
Dec 06, 2023
Draft
Match Rule set for person BE.
Custom_Match_Model
uplicates
Dec 06, 2023
Draft
OOTB Match Rule set for person BE.
Custom_Match_Model
uplicates
Dec 06, 2023
Draft
OOTB Match Rule set for person BE.
You can see that some match rules are configured for you.
Click Custom_Match_Model match rule to check out the details.


New...
Home
Explore
My Import/Exp...
Business Events
Translations
My Jobs
Security
Global Settings
Person
Person > Custom_Match_Model
Draft
Publish
Training of the machine learning model is in progress. You cannot edit the model configuration, fields for match training, and declarative rules.
Model Configuration
Declarative Rules
Machine Learning Model
To resolve duplicates, you can choose to use declarative match rules, a machine learning (ML) model, or both. The declarative match rules are based on an exact or fuzzy match strategy. An exact match strategy identifies identical records, and a fuzzy match strategy identifies similar records. Click Declarative Rules tab to see how the match rule.
Edit
Model ID:
-a436ff6a1073
son BE.
About Population:
Created by: (c)2000-2023 Informatica LLC. All rights reserved. Last change: 22 May 2023 USA Standard - v10.2.0HF1-22May23 This population uses a character set that supports ISO-LATIN-1 The population rules support the typical identity data found in the United States.
Candidate Selection Criteria
Field Name
Field Type
Filter Candidates By
Key Generation Level
Candidate Search Level
fullName
Person_Name
Standard
Typical


Informatica Business 360 Console
New...
Home
Explore
My Import/Exp...
Business Events
Translations
My Jobs
Security
Global Settings
Person
Person > Custom_Match_Model
Training of the machine learning model is in progress. You cannot edit the mod
declarative rules.
Model Configuration
Declarative Rules
Machine Learning Model
Here you can see the declarative match rules that are a set of conditions and business entity attributes that are required to identify duplicate records. Click Machine Learning Model tab to see how intelligence can be added to the matching process.
Next
Declarative Rules (5)
Rank
Merge Strategy
Match Strategy
tch Level
Description
1
Manual
FUZZY
se
Strongly similar name and exactly matching identifier
2
Automated
FUZZY
Resident
Typical
Strongly similar name, address and exactly matching birth date
3
Manual
FUZZY
Person_Name
Loose
-
Threshold-based
FUZZY
Person_Name
Typical
Similar name and same email
-
Threshold-based
FUZZY
Person_Name
Typical
Similar name and same phone number


New...
Home
Explore
My Import/Exp...
Business Events
Translations
My Jobs
Security
Global Settings
Person
Person > Custom_Match_Model
Draft
Publish
Training of the machine learning model is in progress. You cannot edit the model configuration, fields for match training, and declarative rules.
Model Configuration
Declarative Rules
Machine Learning Model
With this approach you can train an ML model based on your business needs, where you can label record pairs . This reduces the manual effort required for defining declarative match rules.
Click Continue Labeling to see how to label record pairs.
Training the Machine Learning Model
Continue Labeling
Labeling for batch 1 is in progress. Select conclusive labels for 24 record pairs to complete batch 1.
6
Record Pairs Reviewed
6
Record Pairs with Conclusive Labels
-
Precision
-
Recall
-
Accuracy
Error Matrix
Nothing to display. The error matrix appears after the first batch of record pairs is labeled.

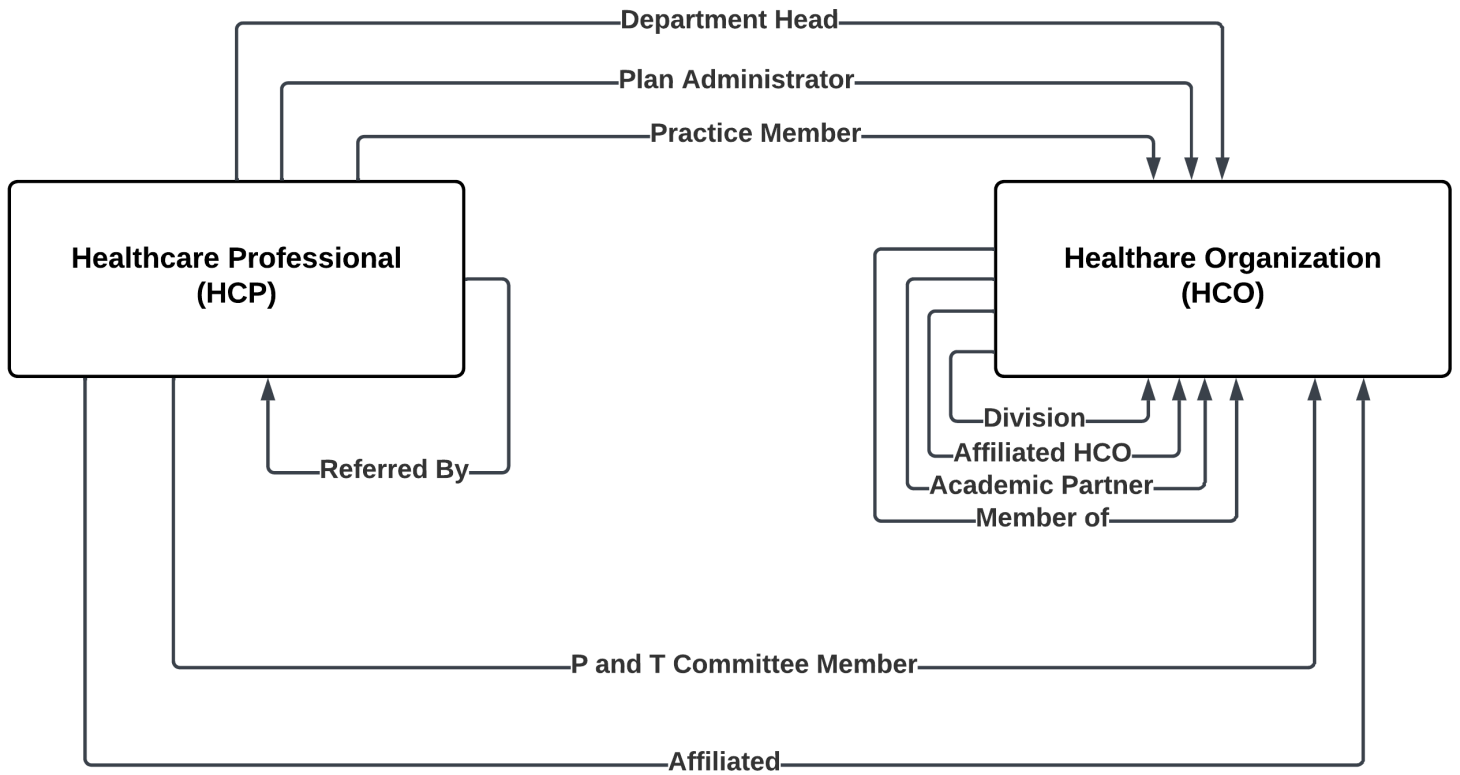
Rule Sequence	Rule Name
1	Ex-Name, Ex-Address, Ex-City, Ex-State, Ex-zip
2	Ex-Name, Ex-Address, Ex-zip
3	Ex-Name, Fuzzy Address, Ex-Zip (Conservative)
4	Ex-Name, Fuzzy Address, Ex-Zip (Typical)
5	Fuzzy-Name, Ex-Address, Ex-Zip (Conservative)
6	Fuzzy-Name, Ex-Address, Ex-Zip (Typical)
7	Fuzzy-Name, Fuzzy Address, Ex-Zip (Conservative)
8	Fuzzy-Name, Fuzzy Address, Ex-Zip (Typical)
9	Fuzzy-Name, Ex Address, Ex City, Ex State, Ex Zip
10	Match Quidel records with OneKey based on OneKey WKP ID

Rank	Merge Strategy	Match Strategy	Match Criterion	Match Level	Description
1	Automated	FUZZY	Person_Name	Typical	DrugEnforcementAgencyNumeric Exact, Full Name Fuzzy
2	Automated	FUZZY	Person_Name	Typical	DrugEnforcementAgencyNumeric Exact, StateLicenseNumeric Exact, FirstName Fuzzy
3	Automated	FUZZY	Person_Name	Typical	DrugEnforcementAgencyNumeric Exact, StateLicenseNumeric Exact, LastName Fuzzy
4	Automated	FUZZY	Person_Name	Typical	NationalProviderIdentifier Exact, Full Name Fuzzy
5	Automated	FUZZY	Person_Name	Typical	NationalProviderIdentifier Exact, StateLicenseNumeric Exact, FirstName Fuzzy
6	Automated	FUZZY	Person_Name	Typical	NationalProviderIdentifier Exact, StateLicenseNumeric Exact, LastName Fuzzy
7	Automated	FUZZY	Person_Name	Typical	StateLicenseNumber Exact, Full Name Fuzzy
8	Automated	FUZZY	Person_Name	Typical	DrugEnforcementAgencyNumeric Exact, SocialSecurityLastFour Exact, FirstName Fuzzy
9	Automated	FUZZY	Person_Name	Typical	DrugEnforcementAgencyNumeric Exact, SocialSecurityLastFour Exact, LastName Fuzzy
10	Automated	FUZZY	Person_Name	Typical	NationalProviderIdentifier Exact, SocialSecurityLastFour Exact, FirstName Fuzzy

11	Automated	FUZZY	Person_Name	Typical	NationalProviderIdentifier Exact,SocialSecurityLastFour Exact, LastName Fuzzy
12	Automated	FUZZY	Person_Name	Typical	StateLicenseNumeric Exact, SocialSecurityLastFour Exact, FirstName Fuzzy
13	Automated	FUZZY	Person_Name	Typical	StateLicenseNumeric Exact, SocialSecurityLastFour Exact, LastName Fuzzy
14	Automated	FUZZY	Person_Name	Conservative	Full Name Fuzzy, SSN Last Four Exact
15	Automated	FUZZY	Address	Typical	DrugEnforcementAgencyNumeric Exact, Geocoded Street Fuzzy, Geocoded City Fuzz, Geocoded ZIP Exact,
16	Automated	FUZZY	Address	Typical	NationalProviderIdentifier Exact, Geocoded Street Fuzzy, Geocoded City Fuzzy, Geocoded ZIP Exact
17	Automated	FUZZY	Address	Typical	StateLicenseNumeric Exact, Geocoded Street Fuzzy, Geocoded City Fuzzy, Geocoded ZIP Exact
18	Automated	FUZZY	Resident	Typical	Full Name Fuzzy, Geocoded Street Fuzzy, Geocoded City, Geocoded ZIP Exact,
19	Automated	FUZZY	Person_Name	Typical	Full Name Fuzzy , TaxID Exact

Industry Centric Data Model & Match Practices

Life Sciences Data Model –HCO & HCP



Informatica

Comprehensive Life Science Data Model -Key Entities

Health Care Organization

- Subtype
- Location
- Classification
- Segmentation
- Taxonomy
- Taxonomy
- Compliance
- Affiliations

Integrated Delivery Networks/ Accountable Care Organization

- Subtype
- Key people
  - Medical/pharmacy directors, physicians, VPs/Directors of Patient Safety/Innovation, CEOs of affiliated hospitals
- Segmentation
  - Patient demography, geography, organizational structures, payment model

Health Care Professional

- Subtype
- Designation
- Address
- Phone
- Specialty
- Taxonomy
- License
- Degree/Certificate
- Therapeutic areas
- Segmentation
- Indicators (KOL)
- Preferences (Channel, Content)
- Affiliations

Group Purchasing Organization

- Subtype
- Relationships

Contact

- Designation
- Address
- Phone
- Communication

Match Model for HCP Deduplication

HCP > Match Model for HCP Deduplication

Published

Model Configuration | Declarative Rules | Machine Learning Model

Declarative Rules (7)

Edit

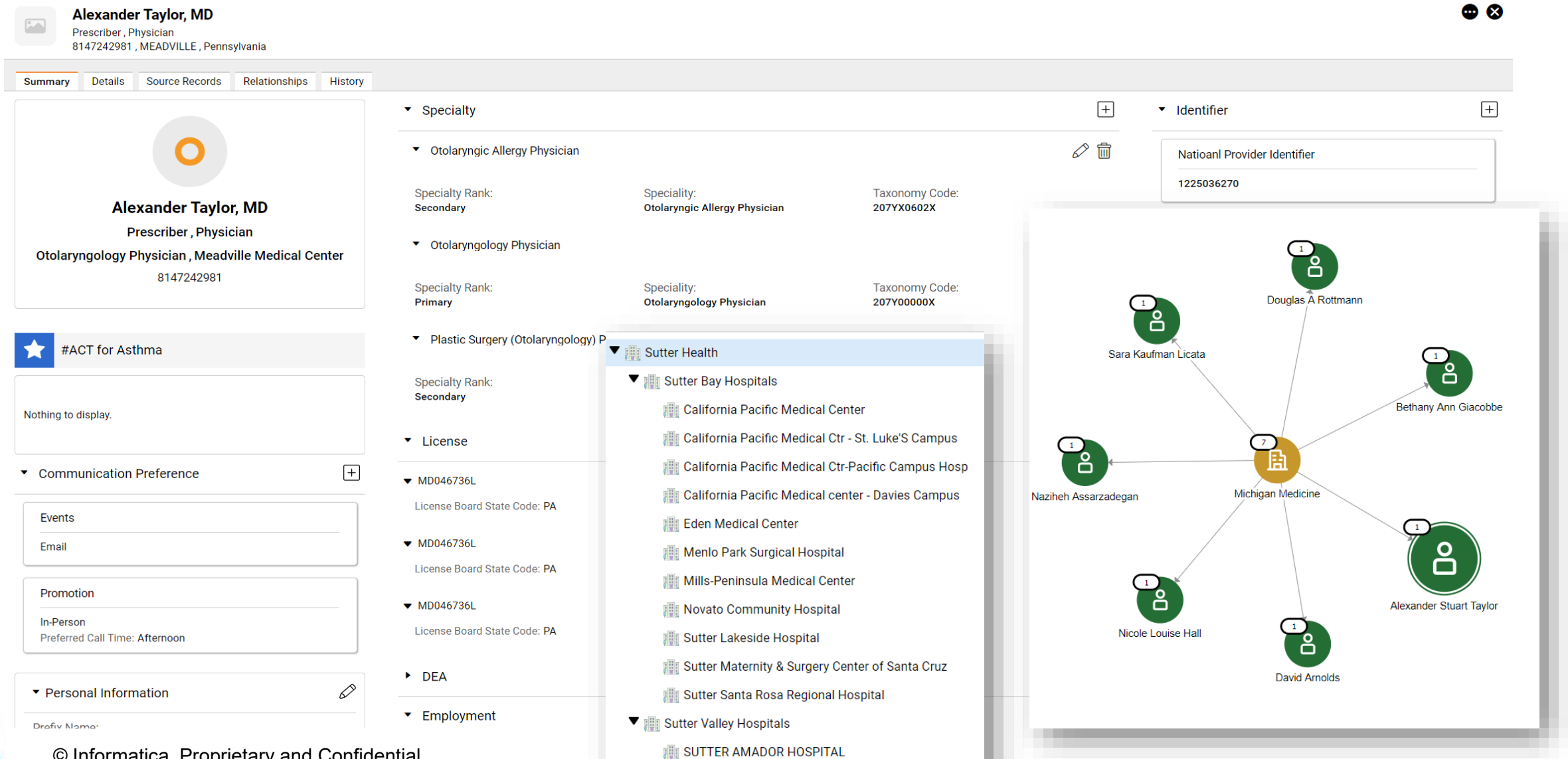
Rank	Merge Strategy	Match Strategy	Match Criterion	Match Level	Description
1	Automated	EXACT	-	-	Exactly similar name and address
2	Automated	FUZZY	Person_Name	Typical	Strongly similar name and license no.
3	Automated	FUZZY	Person_Name	Typical	Strongly similar HCP name and alternate id
4	Automated	FUZZY	Person_Name	Typical	Strictly similar HCP name, address and specialty
5	Manual	FUZZY	Person_Name	Typical	Similar HCP name and address
6	Manual	FUZZY	Address	Typical	Similar HCP last name and address
7	Manual	FUZZY	Person_Name	Typical	Similar HCP alternate name, address and specialty

Candidate Selection Criteria

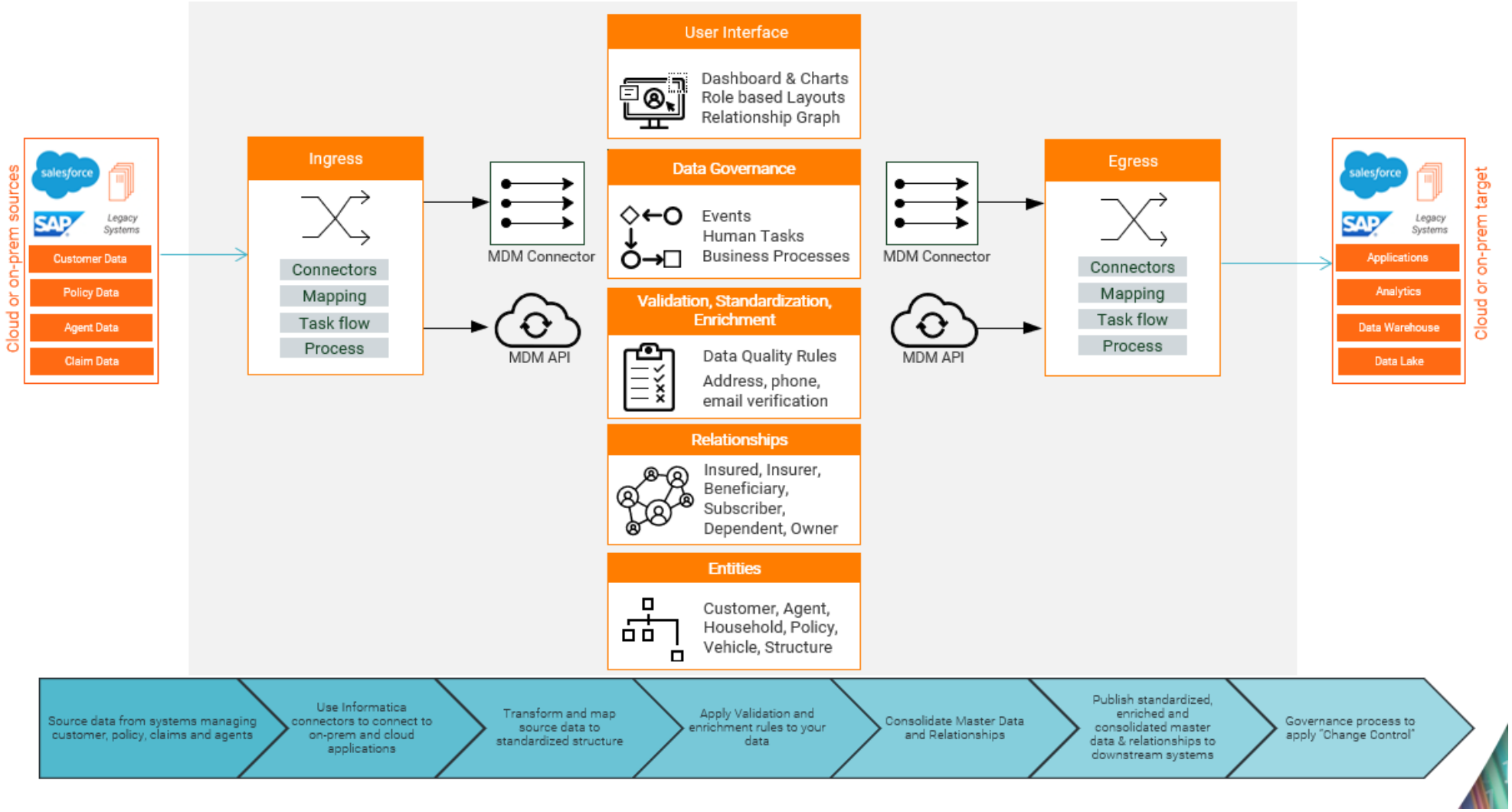
Field Name	Field Type	Filter Candidates By	Key Generation Level	Candidate Search Level
fullName	Person_Name		Standard	Typical

Informatica

360 View

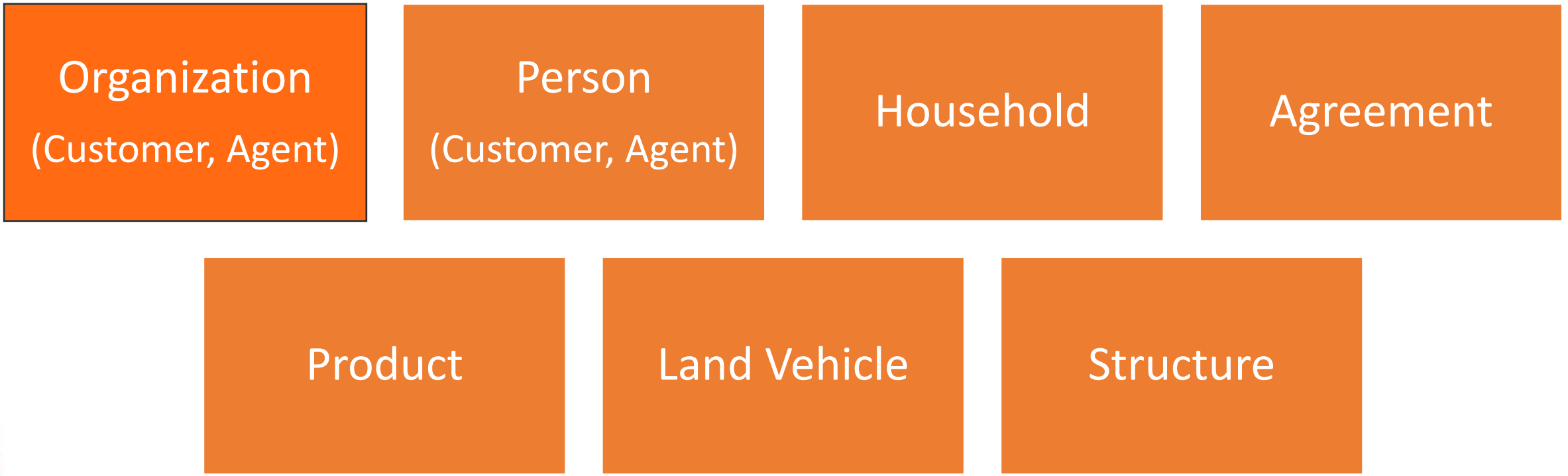


Insurance Data Model- OOB Extension for IDMC

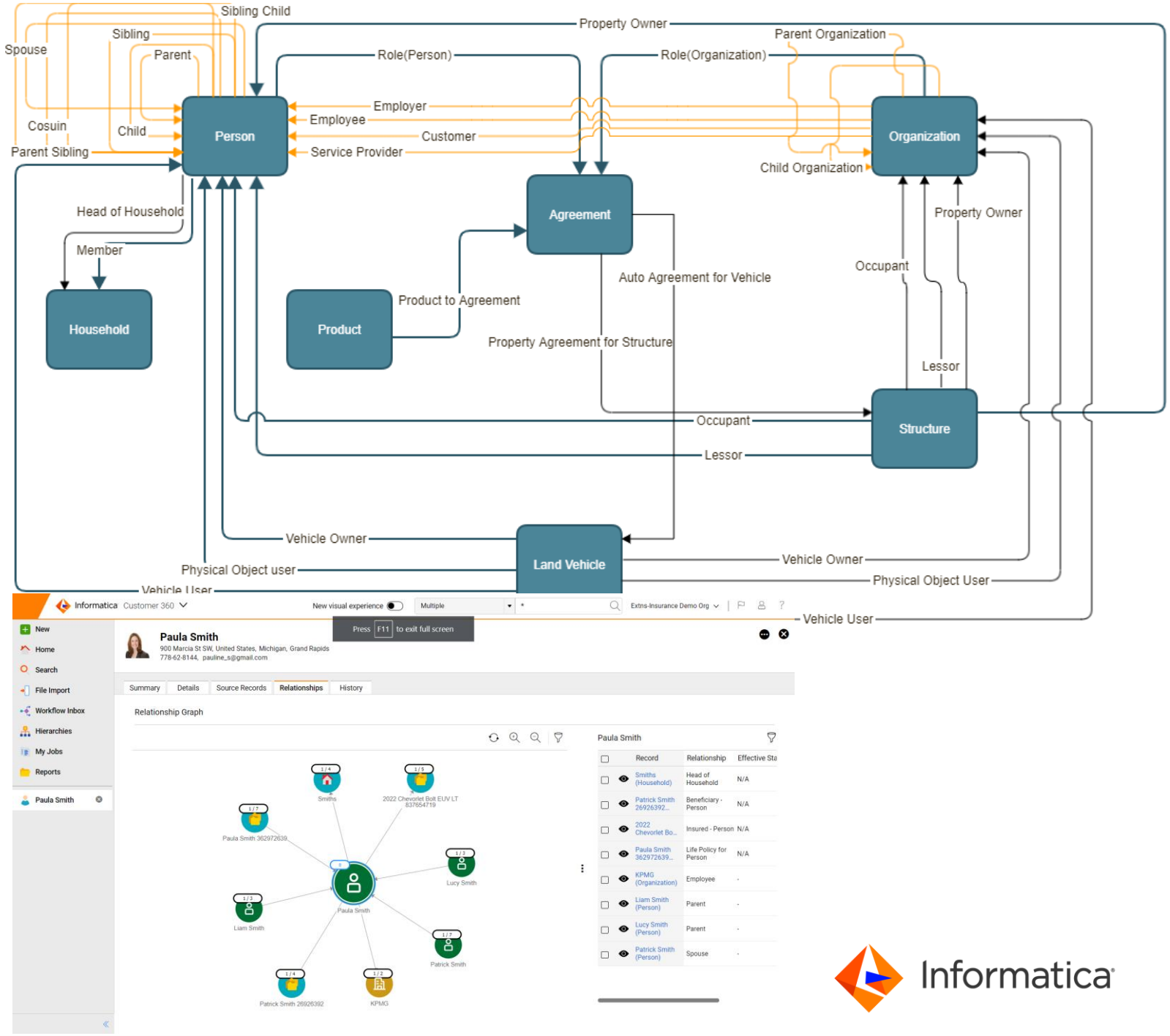


Informatica

Key Business Entities



Insurance Model-Key Relationships



Informatica

Match Model –customer/Agent

Model Configuration | Declarative Rules | Machine Learning Model

Declarative Rules (5) Edit

Rank	Merge Strategy	Match Strategy	Match Criterion	Match Level	Description
1	Automated	EXACT	-	-	Exactly similar name and address.
2	Automated	FUZZY	Person_Name	Typical	Strongly similar person name and exactly similar government ID.
3	Automated	FUZZY	Resident	Typical	Strongly similar name, address, and exactly similar alternate ID
4	Manual	FUZZY	Person_Name	Typical	Similar person name and same address
5	Manual	FUZZY	Address	Typical	Similar address and same last name

Match/Merge- Key Pointers

Avoid Hotspots

- Must investigate the high frequency values during the profiling & fix it before matching i.e. Must not use garbage/Test data without enforcing DQ execution for matching ,to avoid job hitting 10k match candidates blocking limit and max 1k Xref for each master record
- Avoid loading transactional Information
- Must not randomly use exact fields like ”DOB”,”CITY”,”STATE”,”ZIP” may results in hotspots
- Follow Iterative Process- Fine tune match model with smaller data sets

Match Configuration

- Keep exact match rules even when you configure Fuzzy match Strategy aim to send reduce records for fuzzy matching
- Per profiling results, keep highest matching percentage rule with top Rank.
- Leverage Segment Matching
- Keep only Entity relevant Field Groups , higher number of field groups will slow the match and merge process,

Benefit - Enhance match accuracy + optimal jobs performance

Questions?

Speaker: Kamal Abrol, Sourya Das

Thank You!

- Customer Success Architect

References

https://informatica.csod.com/LMS/UserTranscript/OnlineClassView.aspx?qs=%5e%5e%5eYjseWjceYYIHv1Aa%2bnK1ilhZcf2c8x7o7uAmmwFLfpsr2qxeOtvi%2bC6pdEoY4VTFnWdXNbFNB0SVEZ6H4Pkcpog68%2bhi4FIx2zS7SJ%2bjOM5GlIVpvYxl1yU%2fJyoRp%2byc

https://success.informatica.com/explore/tt-webinars/mdm-saas-match-and-merge-best-practices.html

https://docs.informatica.com/master-data-management-cloud/business-360-console/current-version/configure-match-and-merge/preface.html

https://infawiki.informatica.com/display/DF/Extension+Releases