





# Identity Resolution

Training and Best Practices



# Contents

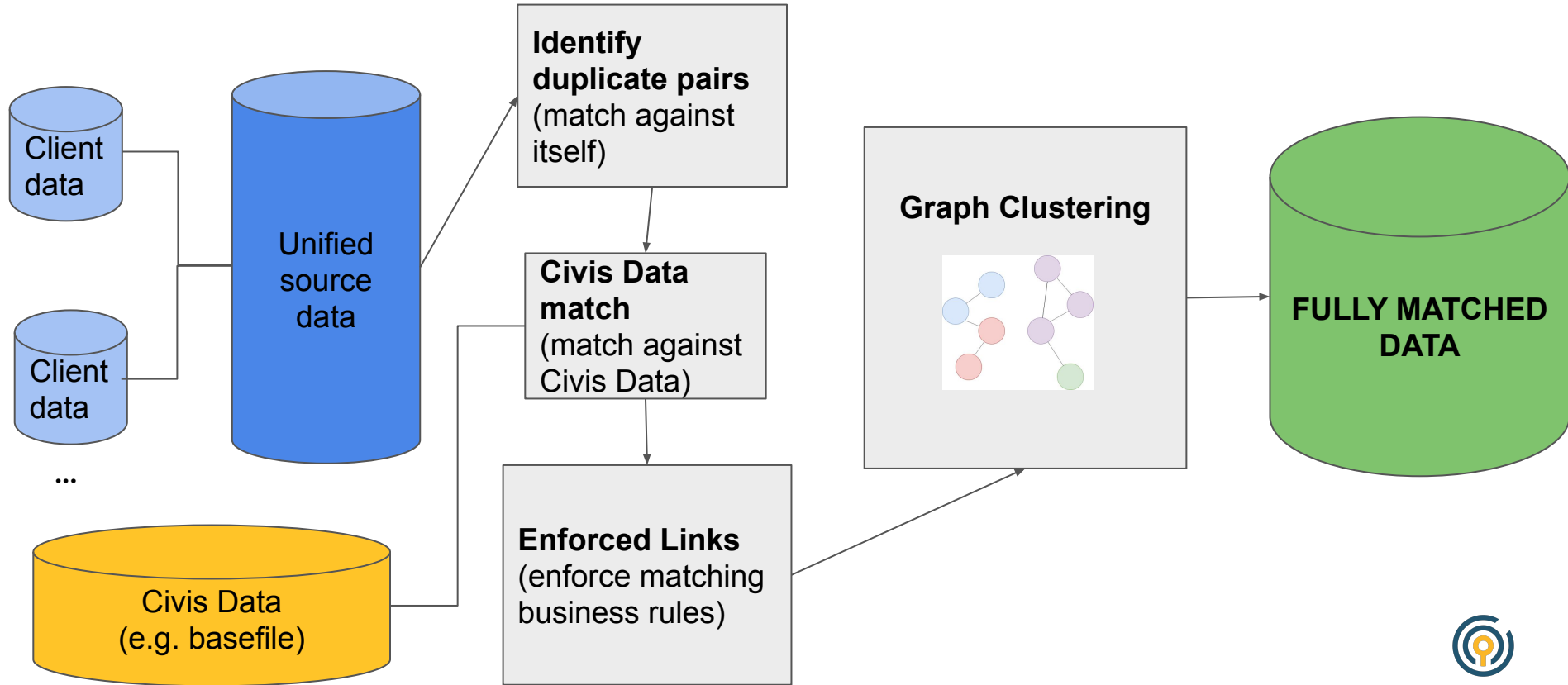
- What is IDR
- How to set up your data
- Running an IDR pipeline
- Interpreting results



# What is IDR?

1. Civis Identity Resolution (IDR) enables users to group together person records within and across multiple data sources to provide a comprehensive and deduplicated view of the individuals they care about.
2. IDR uses scalable distributed computing tools to deliver results efficiently and reliably for datasets of tens of millions of records or more.
3. To determine whether different person records correspond to the same individual, Civis Identity Resolution uses a precise statistical model that has been adjusted based on feedback from dozens of customers in multiple industries over a period of years.

# Core IDR Pipeline





# Preparing Your Data

## WORKFLOWS

### Workflow

Chain related jobs that depend on one another

## MODELS

## LABS

## ENHANCEMENTS

### Address Correction

Standardize and update mailing addresses

### Basefile Delivery Tool

Deliver Civis Data to your cluster

### Civis Data Match

Match records from your data to Civis data

### More Enhancements...

## IDENTITY RESOLUTION

## IDENTITY RESOLUTION (DEMO)

## CREATIVE FOCUS (POLITICAL DEMO)

# Navigating to IDR



DATA CODE **TOOLS** PUBLISH

**WORKFLOWS**  
Workflow  
Chain related jobs that depend on one another

**MODELS**  
**LABS**

**ENHANCEMENTS**  
Address Correction  
Standardize and update mailing addresses  
Basefile Delivery Tool  
Deliver Civis Data to your cluster  
Civis Data Match  
Match records from your data to Civis data  
More Enhancements...

**IDENTITY RESOLUTION**  
IDENTITY RESOLUTION (DEMO)  
CREATIVE FOCUS (POLITICAL DEMO)

# Starting a New Pipeline



## IDENTITY RESOLUTION

### All Pipelines

+ New Pipeline

NAME	ID	LAST UPDATED	SOURCES	TOTAL RECORDS	RESOLVED IDS	LAST RUN
Demo Pipeline	299	12/17/2021 9:49am	3	--	--	12/17/2021 9:49am
All Supporters	196	12/30/2021 9:33am	3	11,415	2,721	12/30/2021 9:33am
City Citizens	190	12/2/2019 3:44pm	2	--	--	





# Adding Sources

## Demo Pipeline

- 1 Data Sources
- 2 Links
- 3 Golden Table
- 4 Outputs
- 5 Review

Click 'Next' to unify the sources below or 'Add New Source' to add another source to this pipeline.

NAME	SOURCE	RECORDS	SELECTED COLUMNS	
customers	Civis Database.commercial_demo.customer_list	1,000	10	• • •
loyalty_data	Civis Database.airline.loyalty_data	13,000	10	• • •
bookings	Civis Database.airline.bookings	13,000	10	• • •

[+ Add New Source](#)

### Add Match Target

Civis Data (quarterly commercial)  

Match Target to match against your sources

Cancel

Next



# Adding Sources

## Demo Pipeline

- 1 Data Sources
- 2 Links
- 3 Golden Table
- 4 Outputs
- 5 Review

Select the column in your source that corresponds with each attribute below. Not all attributes are required, but more information will help improve match quality.

### REQUIRED

#### Unique ID

A unique identifier for each record, e.g. Customer ID

### SECONDARY

None of the PII fields below are required, but more information will help improve link quality. Mapping name fields is strongly recommended.

#### First Name

Recommended for best results.

#### Last Name

#### Full Address House Number

#### City

#### State Code State

2-letter code, e.g. IL

**\*\*\*Your data sources must include a Unique ID, First Name, and Last Name\*\*\***



# Adding Links

## Demo Pipeline

✓ Data Sources   2 Links   3 Golden Table   4 Outputs   5 Review

Linked records will be given a score between 0.5 and 1, representing estimated confidence that they are the same person. Choose the threshold, or minimum score that we should use to determine these links and tell us about any records that should automatically be linked.

### LINK SCORE THRESHOLD

Choose the minimum score you want to use to link records.

#### Score Threshold

Must be between 0.5 (lowest confidence) and 1 (highest confidence).

### ENFORCED LINKS (OPTIONAL)









You may already know fields that match exactly across your data sources. Select those columns below to ensure that records that match will automatically get a score of 1.

If you identify multiple links between the same two sources, only one pair has to match for the records to be linked.

Example Source 1 email_address	Example Source 2 Email	Example Source 3 phone_number	Example Source 4 Phone
-----------------------------------	---------------------------	----------------------------------	---------------------------

### ENFORCED LINK 1

Select two sources that have identical fields.

customers  	loyalty_data  
firstname  	fname  



# Golden Table Formatting

## Demo Pipeline

- ✓ Data Sources   ✓ Links   **3 Golden Table**   4 Outputs   5 Review

Create a single golden record for each resolved ID using the strongest personal identifiers from its linked records. Tell us how the values for each field should be chosen below. If you don't want to create a Golden Table, leave the Golden Table fields on the Outputs page blank.

### FIELDS

First Name

Middle Name

Last Name

Address

Email

Phone Number


### GOLDEN TABLE RULES

The **Automatic** option will choose a value based on:

- Frequency (how often it shows across records)
- Formatting (how valid the particular data appears, depending on the field)

#### First Name

How should we choose which first name to use?

- Automatic 
- By Preferred Source

#### Middle Name

How should we choose which middle name to use?



# Configure Output Destinations

## Demo Pipeline

Data Sources  Links  Golden Table **4** Outputs  5 Review

Choose where your final output tables will be saved

### CUSTOMER GRAPH

This table will include all of your resolved IDs and their linked records.

Database

Civis Database



Schema

commercial\_demo



Table

idr\_demo\_graph

### GOLDEN TABLE

This table will include the golden records of your resolved IDs. If all fields below are left blank, a Golden Table will not be created.

Database

Civis Database



Schema

commercial\_demo



Table

idr\_demo\_golden

### LINK SCORES

This table will include link scores for the records between all sources. If all fields below are left blank, a Link Scores Table will not be created.

Database

Civis Database



Schema

commercial\_demo



Table

idr\_demo\_link

# Interpreting Results: Customer Graph

< Back to List of Pipelines

## Demo Pipeline: Run Summary

Edit Pipeline

Run Date

Run ID: 365824439 Threshold: 0.8 [More Run Details](#) ▾



Records Per Resolved ID [Sample Records](#) [Expand Cluster Query](#)

Showing 100 sample records for up to 100 sample resolved IDs. Some resolved ID groups may be truncated. [Query Code](#) ▾

[Load another 100 records](#)

RESOLVED ID	SOURCE	SOURCE PRIMARY KEY	FIRST NAME	MIDDLE NAME	LAST NAME	EMAIL	PHONE 0	FULL ADDRESS	HOUSE NUMBER	STREET	UNIT	STATE CODE	CITY	ZIP
lz7j7nqsumwt ahnr	Civis Data	IA- Y2945485515 7855												
lz7j7nqsumwt ahnr	bookings	30942939	ROBERT		NUSBAUM	BUZZBOYN@ CROSSPATHS. NET	6417522796	1211 W STATE ST APT 204				IA	MARSHALLT OWN	50158
lz7j7nqsumwt ahnr	bookings	30942939	ROBERT		NUSBAUM	BUZZBOYN@ CROSSPATHS. NET	6417522796	1211 W STATE ST APT 204				IA	MARSHALLT OWN	50158
lz7j7nqsumwt ahnr	loyalty_data	30942939	ROBERT		NUSBAUM	BUZZBOYN@ CROSSPATHS. NET	6417522796	1211 W STATE ST APT 204				IA	MARSHALLT OWN	50158

## Interpreting Results: Golden Table

Name	Type
resolved_id <span>distkey</span> <span>sortkey</span>	varchar(1024)
first_name	varchar(4096)
middle_name	varchar(4096)
last_name	varchar(4096)
house_number	varchar(4096)
street	varchar(4096)
unit	varchar(4096)
city	varchar(4096)
state_code	varchar(4096)
zip	varchar(4096)
email	varchar(4096)
phone	varchar(4096)

## Interpreting Results: Link Scores

Name	Type
source_name	varchar(1024)
source_primary_key distkey sortkey	varchar(1024)
match_name	varchar(1024)
match_primary_key	varchar(1024)
score	double



# Putting your Pipeline into a Workflow

IDR Demo Workflow



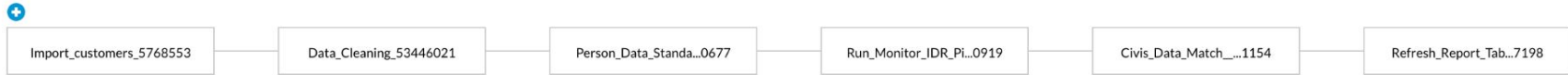
JS Jane Smith | [Settings](#)

[Manage Versions](#) [Execute](#) [View Executions](#)

CONFIGURATION

GRAPH

PARAMETERS



# Thank you!

If you have any questions, please contact Civis Client Success at [support@civisanalytics.com](mailto:support@civisanalytics.com).