Data Onboarding - Match Preview

10/01/2026 10:48 am EDT

Match Preview: Match Test Guide

This guide covers how to run a match test using ID5's Data Onboarding Service for an initial match test against the ID5 identity universe. The match test (**Match Preview**) is free and returns a coverage report showing how many of your hashed emails (HEMs) ID5 can resolve to ID5 IDs and optionally other identifiers.

Prerequisites

1. A file containing SHA-256 hashed emails, formatted as described in Input File Specification

2. One of the following for data transfer options:

Data transfer options

Option	Description
Your own S3 bucket (preferred)	Grant ID5 read access to a prefix where you upload the input file, and write access to a prefix where ID5 delivers the result
ID5-hosted bucket (`eu-central-1`)	See sub-options below

If ID5 hosts the bucket:

Your AWS status	What happens
You have an AWS account	ID5 creates a bucket and authorizes your AWS account on it directly. You read and write using your own IAM roles and users — no credentials are exchanged, and ID5 can revoke access at any time
You have no AWS account	ID5 creates the bucket and issues a dedicated access key and secret, restricted to your upload and delivery prefixes. The secret is shared once through a secure channel

---

1. Service Configuration

Before onboarding, confirm the following configuration preferences with your ID5 account manager:

Configuration	Description
Match methodology	**Current Matching** — Returns matches for the last observed identifier for the input signal **Individual Matching** — Returns matches for all identifiers associated with the individual linked to the input signal in the ID5 graph. **Household Matching** COMING SOON — Returns matches for all identifiers associated with the household linked to the input signal in the ID5 graph
Geography	Countries considered for the match test.
Lookback Days	The number of days to retrieve connections in the past (max 90 days).
Partners to match cookie IDs with	List of partner IDs to perform partner cookie ID matching with.

2. Input File Specification

HEM is the input signal type used throughout this guide; the field name `input_uid` is generic to support additional signal types in future.

Emails must be normalized and hashed using SHA-256 before sending to ID5:

- Remove leading and trailing spaces
- Convert all ASCII characters to lowercase

Your input file must contain one HEM per row, with a header row.

Format: Parquet (zstd compression supported)

Field	Required	Description

Field	Required	Description
`segment`	Optional	The segment associated with the input signal. Include a segment column if you want to to see match rate broken out by segment.
`input_uid`	Required	The signal to be matched.
`input_uid_type`	Required	The type of signal (e.g., `HEM`).

Upload Location & Structure

Location	S3 location agreed upon with ID5
Partitioning	Date-partitioned folders, format `/DATE=yyyy-mm-dd` (processing date)
Completion signal	An empty `_SUCCESS` file must be written to the S3 prefix once upload is complete, to signal the file set is ready for processing

3. Match Test Output

The match test output returns, for each identifier type, the count and percentage of HEMs with at least one match, along with the count of unique matched identifiers.

Delivery Details

Location	Configured prefix, defaulting to `/match-preview/DATE=yyyy-mm-dd`
Partitioning	By date (`DATE=yyyy-mm-dd`)
Format	CSV

Fields

Field	Description
`matched_uid_type`	The identifier type (or `(all)` for the aggregate row)
`num_input_uids_distinct`	Distinct count of input identifiers
`num_matched_input_uids_distinct`	Distinct count of input identifiers with at least one match per identifier type
`match_rate_pct`	`num_matched_input_uids_distinct` ÷ `num_input_uids_distinct`, as a percentage
`num_matched_uids_distinct`	Distinct count of matched identifiers of this type

Example output — no `segment` column in input

matched_uid_type	num_input_uids_distinct	num_matched_input_uids_distinct	match_rate_pct	num_matched_uids_d
(all)	10000	8000	80	30312
ID5_ID	10000	8000	80	24100