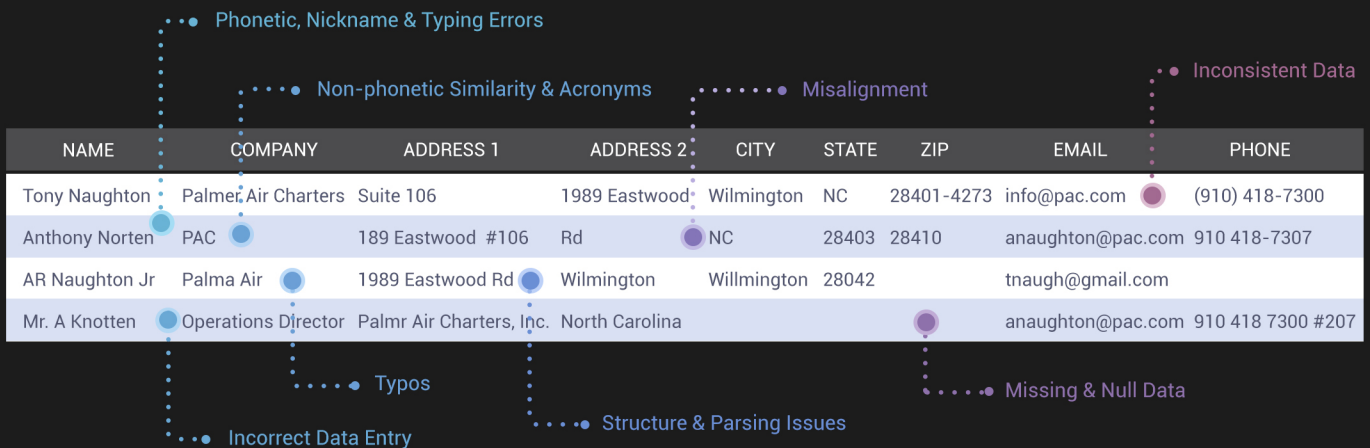# 360°Science

matchit®

THE MATCHING FRAMEWORK

It's **Science**.

Conventional wisdom would have you believe that matching and deduping are more art than science: accuracy purely a function of skill and experience and not the tools available to a user. We know the exact opposite to be true, that matching is a science, and with the right instruments, even a user who has never matched data before can generate incredibly accurate results, quickly and easily.

Conventional matching solutions require your data to be "analytics-ready:" standardized, schemas consistent, and certain datapoints validated. These exhaustive prerequisites not only slow you down, but they eat up valuable resources.

matchit®, on the other hand, is anything but conventional. The foundation of our matching software, matchit is designed specifically to deliver results that mirror human-like perception, at scale and without preprocessing.

Phonetic, Nickname & Typing Errors

Non-phonetic Similarity & Acronyms

Misalignment

Inconsistent Data

NAME	COMPANY	ADDRESS 1	ADDRESS 2	CITY	STATE	ZIP	EMAIL	PHONE
Tony Naughton	Palmer Air Charters	Suite 106	1989 Eastwood	Wilmington	NC	28401-4273	info@pac.com	(910) 418-7300
Anthony Norten	PAC	189 Eastwood #106	Rd	NC		28403 28410	anaughton@pac.com	910 418-7307
AR Naughton Jr	Palma Air	1989 Eastwood Rd	Wilmington	Willmington	28042		tnaugh@gmail.com	
Mr. A Knotten	Operations Director	Palmr Air Charters, Inc.	North Carolina				anaughton@pac.com	910 418 7300 #207

Typos

Missing & Null Data

Incorrect Data Entry

Structure & Parsing Issues

Human-like Perception, **Superhuman Scale.**

360Science has successfully replicated the natural intelligence humans use when forming comparisons - and we deliver that capability at scale.

matchit uses AI, a proprietary phonetic algorithm, customizable lexicons, and a contextual scoring engine to defeat the inconsistencies and challenges commonly found in contact and business data.

Natural Language Processing (NLP) refers to AI methods concerned with understanding human language as it is spoken or heard. matchit makes use of NLP techniques like lexical semantics to develop an understanding of your data based on what is and not where it resides in a table.

For example, matchit understands that Tony is a **variant form** of Anthony, Mr. is a **title or prefix**, MD is a **qualification**, and PAC is a possible **acronym** for Palmer Air Charters. matchit scours the data, looking for signals and indicators like these that inform how a single datapoint should be treated, extracted, and weighed.

NAME Mr. Tony Naughton MD	Title Detection
	Variant Form Name Lexicon
	Qualification Detection
COMPANY PAC, LLC	Acronym Detection
	Business Word Signal
ADDRESS 1 Operations Director	Job Title Signal
	Job Title Extracted
ADDRESS 2 1989 Eastwood Rd.	Premise Extracted
	Thoroughfare Extracted
EMAIL info@pac.com	Alias User Detection
	Corporate Email Address Detection
PHONE 910 418 7300 #207	Phone Normalized
	Extension Detection

The Problem with **Conventional Solutions**

Apart from needing impossibly clean and standardized data, conventional matching solutions require users to understand the various functions and off-the-shelf fuzzy algorithms used to produce match keys, the alphanumeric sequences that then become the basis for comparison and finding matches.

While this approach might be somewhat effective on limited data like product codes or order numbers, the nuances inherent to customer data result in a high number of missed matches and false positives. Even if you did master the intricacies of every fuzzy algorithm found in conventional matching solutions, the truth is, they were never designed for the phonetic complexities found in contact and business data. They are simply unable to adequately handle pronunciation and stress syllables.

Furthermore, these solutions don't truly grade matches, they grade the similarity of match keys. And for most, this is a binary operation. Either match keys match, or they don't.

What Makes
matchit Different.

Smart Settings

This collection of features enables matchit to configure itself optimally based on the nature of the incoming data with very little user input. This makes matchit incredibly easy to use and decreases the time it takes to ingest new data into an existing data quality flow.

Speed

Fortune 1000 executives agree that reducing time-to-insight is the primary driver for their investment in data. Doing so enables their firms to act faster when analyzing the business, extracting intelligence, making critical decisions, and bringing new capabilities to market. The ability to bring your data as it is without preprocessing greatly accelerates time-to-insight.

Accuracy

Accuracy lives at the very foundation of every record linkage, common view, and entity resolution initiative. For some organizations that rely on data to fuel business processes and strategic decision making, missed matches and false positives can be devastating. Conventional solutions are hamstrung because they don't actually compare your data, they compare keys.

Comparing Data, Not Keys

As you'll see in the following matching process overview, matchit doesn't rely on a single comparison between keys to find a match. Instead, matchit evaluates records contextually, running a variety of comparisons and scoring them individually to grade the similarity between all the relevant data points.

This contextual approach imitates human-like perception and enables matchit to consistently outperform conventional solutions. The best part?, matchit can deliver this at scale to process hundreds of millions and even billions of records.
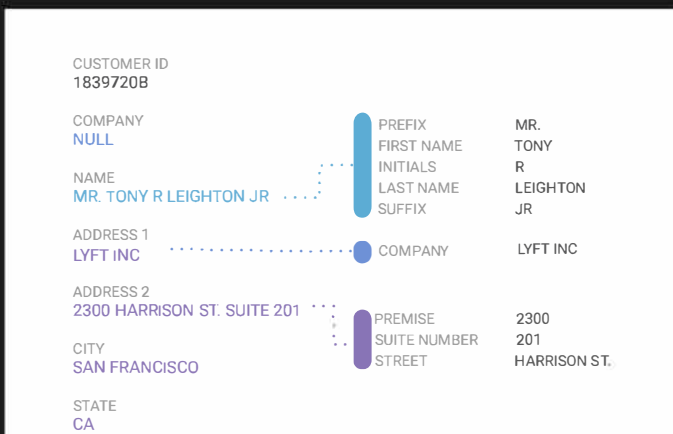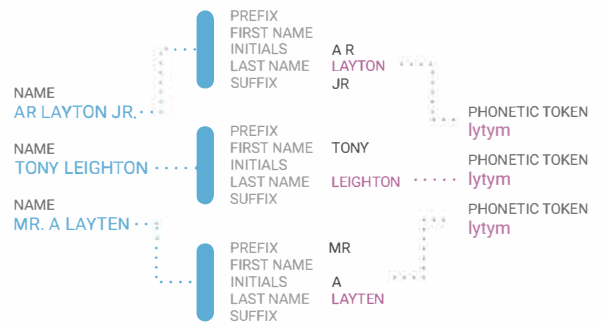
BINARY

Match Key	—	Match Key

CONTEXTUAL

Name		Name
Address		Address
Zip		Zip
Telephone		Telephone
Email		Email

A Smarter Way to **Match.**

CUSTOMER ID
1839720B

COMPANY
NULL

NAME
MR. TONY R LEIGHTON JR

ADDRESS 1
LYFT INC

ADDRESS 2
2300 HARRISON ST. SUITE 201

CITY
SAN FRANCISCO

STATE
CA

PREFIX	MR.
FIRST NAME	TONY
INITIALS	R
LAST NAME	LEIGHTON
SUFFIX	JR

| COMPANY | LYFT INC |

PREMISE	2300
SUITE NUMBER	201
STREET	HARRISON ST.

NORMALIZATION
Making Sense of Your Data

The journey begins as your data enters matchit where it is first normalized. Using AI to understand individual data elements, matchit breaks down complex and concatenated fields into their constituent parts. The quality of fields like name, address, and email are calculated and the resulting qualitative scores inform matchit on the trustworthiness of specific attributes.


NAME
AR LAYTON JR.

NAME
TONY LEIGHTON

NAME
MR. A LAYTEN

PREFIX	
FIRST NAME	
INITIALS	A R
LAST NAME	LAYTON
SUFFIX	JR

PHONETIC TOKEN
lytym

PREFIX	
FIRST NAME	TONY
INITIALS	
LAST NAME	LEIGHTON
SUFFIX	

PHONETIC TOKEN
lytym

PREFIX	MR
FIRST NAME	
INITIALS	A
LAST NAME	LAYTEN
SUFFIX	

PHONETIC TOKEN
lytym

PROPRIETARY PHONETIC ALGORITHM
Truly Understanding Pronunication

This process of normalization plays directly into phonetic tokenization. Now that we've isolated values like first name, company name, street name, and town name, we can generate phonetic tokens on these fields to help circumvent errors in your data.

Instead of using off-the-shelf algorithms like Soundex or Metaphone that were never designed for the nuances of contact and business data, matchit uses a proprietary phonetic algorithm that has an enhanced understanding of stress syllables and pronunciation.
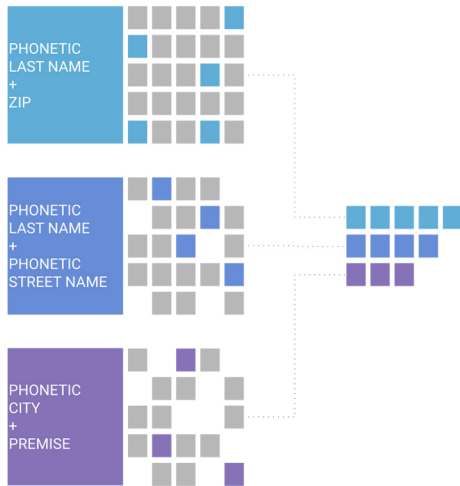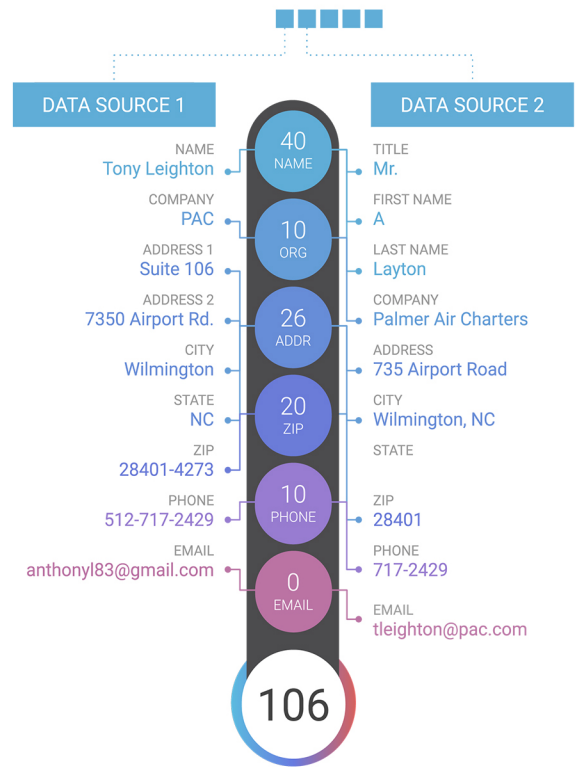
CANDIDATE GROUPING
Detecting Similarity

Once matchit has made sense of the data, the normalized and tokenized values are used to seek out similar records. It's important to note that we aren't finding matches yet, we're simply identifying groups of records that are good candidates for further comparison.

Unlike conventional solutions, this clustering process doesn't depend on a single datapoint being accurate, consistent, and present.

Leveraging the generated values from the previous steps, matchit is able to locate two matching records that have nothing exactly in common.

Why candidate groups? The short answer is scale. matchit can run efficiently on over a billion records and perform real-time lookups on massive datasets. Without candidate grouping, this wouldn't be possible.

CONTEXTUAL SCORING
Human-Like Perception

As candidate groups are created, they're funneled into the contextual scoring engine where records are compared two at a time. All the available data that make up your records are graded for similarity and assigned a composite score.

Where conventional approaches rely on the comparison and scoring of match keys, matchit is comparing and scoring the contact name, company name, address, zip, telephone, email, website, and even custom fields individually.

The composite scores are brought together into a single match score which establishes the overall similarity between two records.