



Unlock the full potential of your CRM
Improve data quality with Enlighten®

www.innovativesystems.com

Enlighten[®]

Introduction

Michael Ott

Senior Vice President

www.innovativesystems.com

CRM systems quickly become a landfill of data



Improved data quality drives revenue growth



Communications reach the right audience



Service & Support retains more customers



Outside sales territories are better planned



Marketing campaigns are personalized



Telesales connects with more prospects



Email sends see lower bounce rates



Inside sales sees more cross-sell opportunities



Decisions are based on better analytics

Enlighten Product Suite



**Creating
CRM Systems**

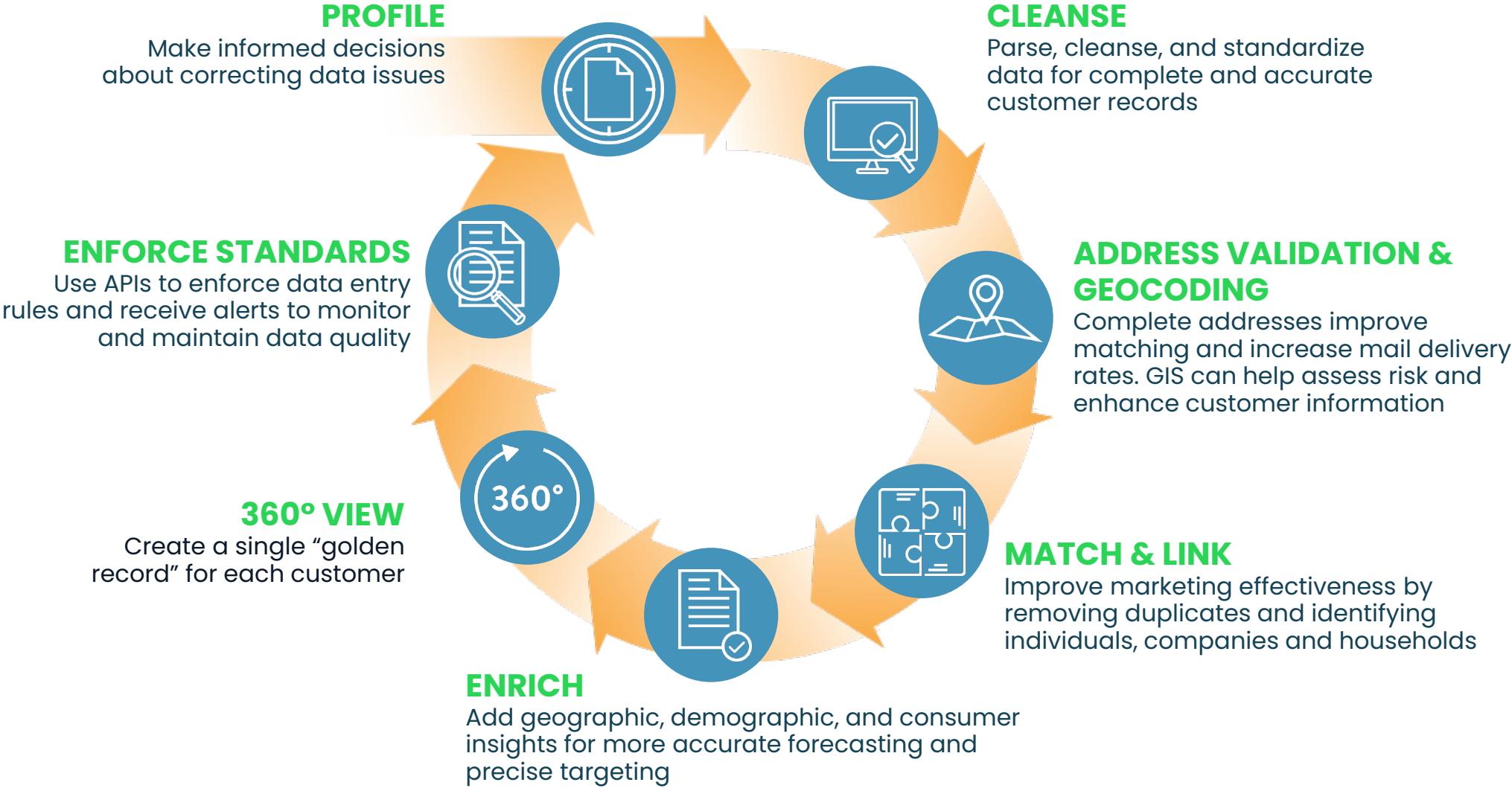


**Maintaining
CRM Systems**



**Auditing
CRM Systems**

Enlighten Product Suite: Creating CRM Systems



Enlighten Product Suite: Maintaining CRM Systems

- ✓ Monthly – Look for duplicates with matching software to combine matches
- ✓ Annually – Profile data to look for anomalies
- ✓ Real time – Onboarding APIs check for:
 - ✓ Already a customer
 - ✓ Address verification
 - ✓ Customer Identification Program (CIP)
 - ✓ Compliance warnings



Sanctions



Block List



Death List

Enlighten Product Suite: Creating/Maintaining CRM Systems

Innovative Systems performs assessments to audit data quality



- Data to be migrated into an existing system or used to create a new CRM system
- Comes from various sources with different field configurations and quality standards
- CRM systems start with clean data



- Data in a CRM system that has degraded over time from user interactions
- Monitor and maintain data quality

Our dictionaries have the definitions

- ✓ We have been creating customer-centric databases for over 50 years
- ✓ Enlighten dictionaries contain millions of definitions for words, phrases, and patterns, built from billions of records reviewed
- ✓ Our technology continues to evolve, now in its 14th generation
- ✓ Users don't have to test or tune their system to improve results
- ✓ Enlighten processes data faster with more accurate results

We don't use weighted field scoring

Threshold
80%

- ✓ Other products compare field pairs and assign a percentage
- ✓ Anything above a set threshold percentage is considered a duplicate
- ✓ There is no transparency for the reason for the match
- ✓ Matching criteria cannot be fine tuned to improve results
- ✓ The result is either over-matching or under-matching

Match 1: 85% and a True Match

Charles	Taylor	Male	19480128	Liberia	321-45-9876	SCORE 85
Charles	Taylor		19480128	USA	321-45-9876	

Match 2: 85% and a False Match

Charles	Taylor	Male	19480128	Liberia	321-45-9876	SCORE 85
Charles	T	Male	19480128	Liberia	321-79-8990	

Percentage based matching just doesn't work!

We use pattern-based matching and AI

First Name

Last Name

Title

Date of Birth

Equal

Close

No Match

Blank



Automatic

Match

Review

No Link

Critical



Skip

Each field pairing is compared and assigned an attribute

Each pair is assigned a rank string showing the likelihood of being a duplicate

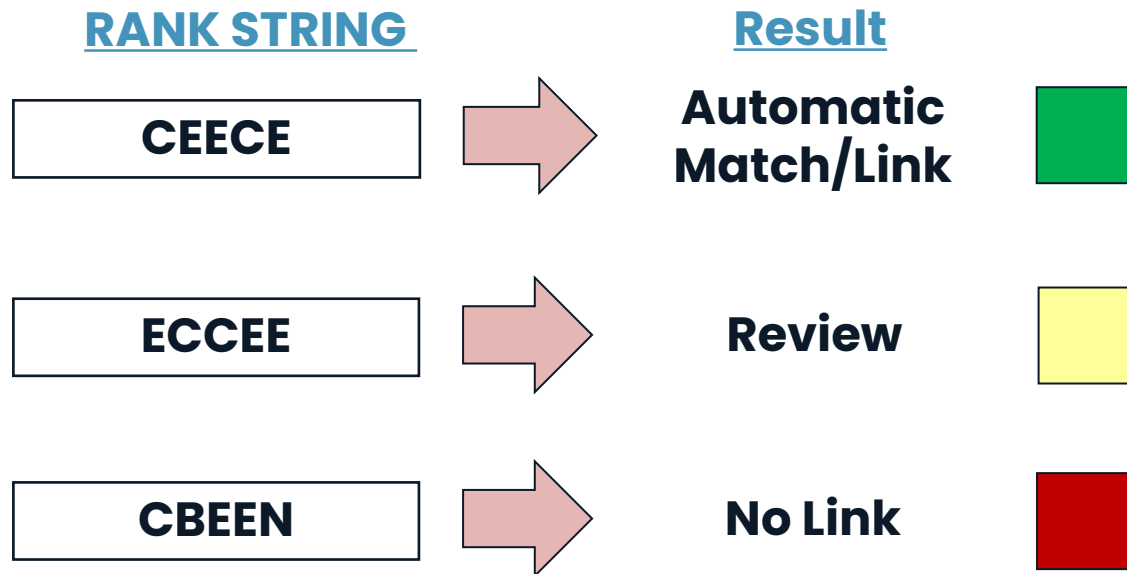
Our AI engine mimics how humans process data to determine a match

Rank strings determine if pairs are automatically matched or reviewed

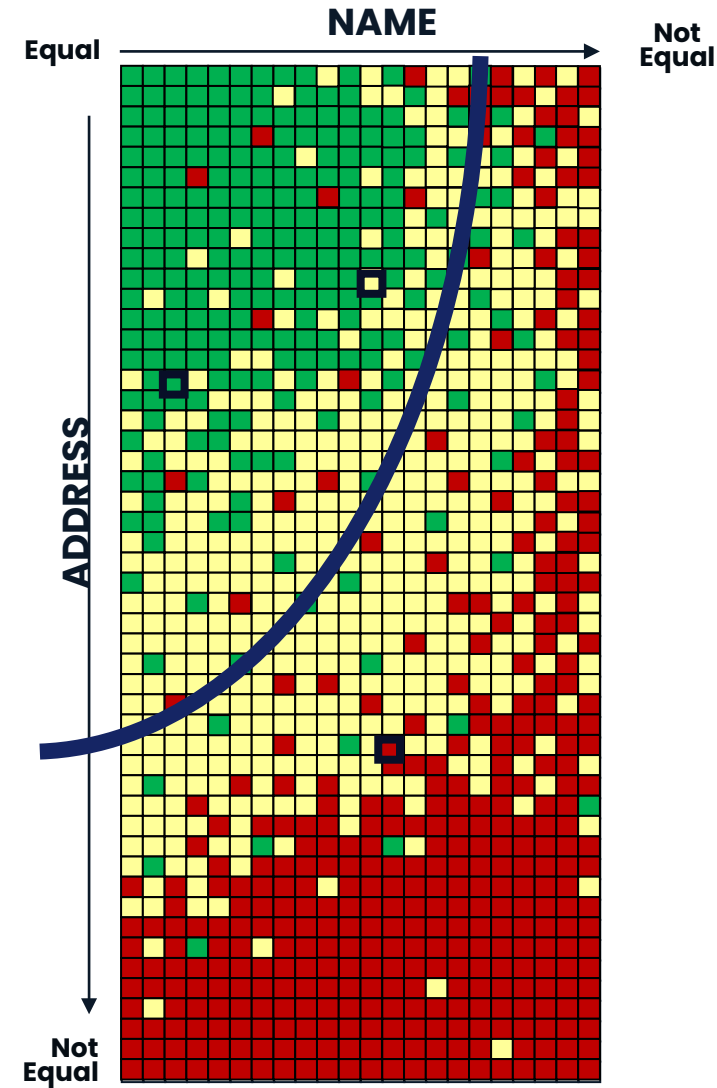
Users can prioritize at the field level for total control

We use pattern-based matching and AI

Record	First Name	Additional Names	Last Name	Street	ID	Rank String
Client	A	-	Taylor	Regent	142-99	
Potential link	Agnus	Reeves	Taylor	Regent	101-73	
Relationship	Close	Blank	Equal	Equal	Not Equal	CBEEN



Cutoff of 80% means an acceptable score



We use natural language processing

- ✓ We developed language rules called grammars
- ✓ Grammars allow Enlighten to process against any alphabet in any language
- ✓ Grammars identify the meaning and order of data
- ✓ Grammars can be added or changed without changing the core product



Enlighten[®]

Product Demo

Hector Cordova

Director, Professional Services & Consulting

www.innovativesystems.com

Demo – Data Profiler Statistics

Enlighten Data Profiler

Test Project A > Customer V1

Run Date: 12/01/2021 10:20:14 AM | Columns: 13 | File Size: 5.72 MB | Record Count: 33456 | Read Time: 02h 20m 49s

Profiling Settings

Metric Grid | Chart View | Data Validation | Drilldown

Select and Copy All | Export Grid

No	Name	Uniqueness	Unique Count	Completeness	Row Count	Null Count	Blank Count	Pattern Count	Overall Data Type	Most Common Data Type	Minimum String	M
1	Customer Id	99.71%	33,360	100%	33,456	0	0	2	Alphanumeric	Alphanumeric	A10-E100750	D99-E1
2	Discount Code	0.02%	8	100%	33,456	0	0	1	Alphanumeric	Alphanumeric	A	Z
3	Forename	10.12%	3,386	100%	33,456	0	0	82	Alphanumeric	Alphanumeric	.	vijay
4	Surname	29.87%	9,992	97%	32,452	1,004	0	139	Alphanumeric	Alphanumeric	O'Connor	x
5	Email	99.71%	33,360	100%	33,456	0	0	19,218	Alphanumeric	Alphanumeric	-Fayek@Albany.medical.center.com	vijay.Ca
6	Telephone	99.2%	33,190	100%	33,456	0	0	83	Alphanumeric	Phone	() 312-441-9080	NOT PR
7	Company Name	82.62%	27,640	99.99%	33,451	5	0	11,414	Alphanumeric	Alphanumeric	1-800-Flowers.com, Inc.	yugas
8	First Order Date	40.35%	13,501	99.71%	33,360	96	0	2	Alphanumeric	Alphanumeric	01/01/1969	16/01/1
9	Sales Last Year	0.97%	324	100%	33,456	0	0	9	Decimal	Integer	-6000	990000
10	Addressline	77.18%	25,823	100%	33,455	1	0	4,942	Alphanumeric	Alphanumeric	# 256, ShenHe Qu QingNian Street	xyz

Showing 10 of 2

Demo – Data Profiler Drilldown

Enlighten Data Profiler

Test Project C > Customer V1

Run Date: 12/02/2021 1:02:15 PM | Columns: 13 | File Size: 5.72 MB | Record Count: 33456 | Read Time: 00h 00m 13s

Profiling Settings

Metric Grid | Chart View | Data Validation | **Drilldown**

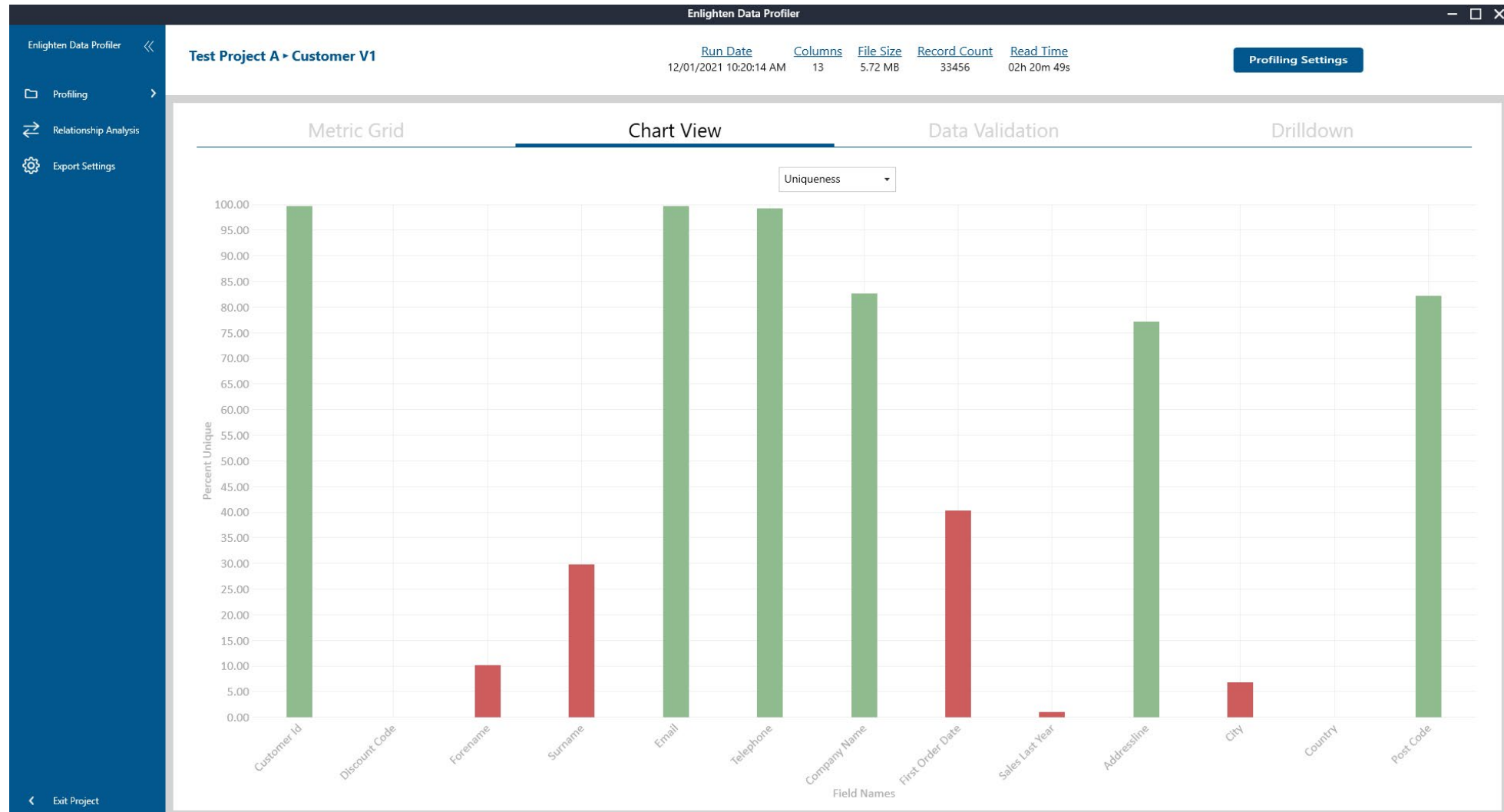
Select and Copy All | Export Grid

Unique Value Drill Down Summary for Addressline

Addressline	Count
NOT PROVIDED	51
None	35
c/o Logan Britton	15
Postbus 1	9
Postbus 20	8
Postbus 1800	8
148 Grenoble Road	6
135 Bishopthorpe Road	6
145 Annfield Rd	6
184 Well Lane	6

Showing 10 of 2583

Demo – Data Profiler Charts



Demo – Cleanse Grammar

Test Parser

Parse Reset

Grammar: US_Enlighten_grammar_2.0.ct

File Type: 825 Tab Delimited

File To Load: [Empty]

Test Parser Input: JOHN SMITH
123 MAIN ST
PITTSBURGH PA 15220
USA

Record 1 of 1

Token	Dictionaries	Standardizations
JOHN	Alpha2 AlphaWithVowel Name_APPENDAGE_SOFT4_PATTERN Dict_MaleFirstName	
SMITH	Alpha2 AlphaWithVowel Dict_UnknownCityNameUS Dict_UnknownGenderLastName	
123	CityUSA_ZIPCODE_INVALID_RULE CityUSA_ZIPCODE_INVALID5 CityUSA_ZIPCODE_INVALID4 StreetUSA_NUMERICAL_RD_NUM DescriptorId_SSN_3DIGIT StreetUSA_APARTMENTNUMBER_LINEBYSELF StreetUSA_APARTMENTNUMBER_LINEBYSELF_NUM13 StreetUSA_FRACTION_3DIGIT Numeric	
MAIN	Alpha2 AlphaWithVowel Name_APPENDAGE_SOFT4_PATTERN Dict_CommonStreetName Dict_UnknownCityNameUS Dict_WeakOrgWord	

Collapse All Nodes

- JOHN SMITH 123 MAIN ST PITTSBURGH PA 15220 USA (Root)
- JOHN SMITH 123 MAIN ST PITTSBURGH PA 15220 USA (GrammarEnI)
- JOHN SMITH 123 MAIN ST PITTSBURGH PA 15220 USA (GrammarEnI_R_IDV_BLOCK)
- JOHN SMITH (GrammarEnI_NAME_IDV_BLOCK_VERYSTRONG)
- JOHN SMITH (GrammarEnI_LINE_1_IDV)
- JOHN SMITH (IndividualLines)
- JOHN SMITH (Name_STRONG)
- JOHN SMITH (NameWestern_STRONG)
- JOHN SMITH (NameWestern_STRONG_WRAPPER_SINGLE)
- JOHN SMITH (NameWestern_STRONG_NAME)
- JOHN SMITH (NameWestern_STRONG_FIRSTLAST)
- JOHN SMITH (NameWestern_STRONG_FIRSTLAST_M)
- JOHN SMITH (NameWestern_STRONG_FIRSTLAST_M_NAME)
- JOHN (NAME_Full)
- JOHN (FirstName)
- JOHN (NameWestern_GIVENNAME_M_RULE)
- JOHN (DICT_NAME_WESTERN_GIVENNAME_MALE)
- JOHN (Dict_MaleFirstName)
- SMITH (NAME_Full)
- SMITH (LastName)
- SMITH (NameWestern_SURNAME_HARD_RULE)
- SMITH (DICT_NAME_WESTERN_SURNAME)
- SMITH (Dict_UnknownGenderLastName)
- (EOL)
- 123 MAIN ST (GrammarEnI_STREET_IDV_BLOCK_VERYSTRONG)
- 123 MAIN ST (GrammarEnI_STREET_IDV_LINE_VERYSTRONG)
- (EOL)
- PITTSBURGH PA 15220 USA (GrammarEnI_CITY_IDV_BLOCK_STRONG)
- PITTSBURGH PA 15220 (GrammarEnI_CITY_IDV_LINE_STRONG)
- (EOL)
- USA (GrammarEnI_CITY_IDV_LINE_STRONG)
- (EOL)

Demo – Cleanse Project with Foreign Data

The screenshot displays the 'Test Cleanse Project' application. At the top, there are 'Test' and 'Reset' buttons. Below them, the 'Cleanse Project' dropdown is set to 'US_RET_CL Working'. There are radio buttons for 'Manual Entry' (selected) and 'Upload File'. The 'Test Field Inputs' table is as follows:

Field	Occurs	Value
Name Address	1	HÉCTOR (****.,&*^@) MAGA#A-Pérez
Name Address	2	C/O Junior Conceição
Name Address	3	#7 123, MAIN STREET.
Name Address	4	pittsburgh=/(/% pennsylvania 15220
Name Address	5	USA
Name Address	6	
Name Address	7	
Name Address	8	

The right-hand pane shows the parsed elements for the input data:

- In X Ref Number(3) '03'
- In X Ref Number(4) '04'
- In X Ref Number(5) '05'
- Name Information
 - Name Line(1) 'HÉCTOR MAGAÑA-PÉREZ'
 - Name Parsed Elements
 - First Name(1) 'HÉCTOR'
 - First Name(2) 'JUNIOR'
 - Last Name(1) 'MAGAÑA-PÉREZ'
 - Last Name(2) 'CONCEIÇÃO'
 - Output Gender(1) 'M'
 - Output Gender(2) 'M'
- Organization Information
 - Organization Parsed Elements
- Street Information
 - Street Line(3) '# 7 123 MAIN ST'
 - Street Parsed Elements
 - House Number(3) '123'
 - Street Name(3) 'MAIN'
 - Street Identifier(3) 'ST'
 - Subaddress Type(3) '#'
 - Subaddress Number(3) '7'
- Urbanization Information
 - Urbanization Parsed Elements
- Municipality Information
 - Municipality Parsed Elements
- City Information
 - City Line(4) 'PITTSBURGH PA 15220'
 - City Parsed Elements
 - City Name(4) 'PITTSBURGH'
 - State Code(4) 'PA'
 - Postal Code(4) '15220'
 - Country(5) 'USA'
- Other Information
 - Other Line(2) 'C/O JUNIOR CONCEIÇÃO'
 - Other Line(5) 'USA'
- User_Group
 - User Link A Occurs
 - User Link B Occurs

Demo – Cleanse Project

Test Cleanse Project

Test Reset

Cleanse Project
US_RET_CL Working

Manual Entry Upload File

Field	Occurs	Value
Name Address	1	Mr. John And Mrs. Mary A. Smith SAVINGS ACCOUNT 891729381293
Name Address	2	DBA SMITH ASSOCIATES
Name Address	3	CARE OF TOM JACKSON ACCOUNTANT
Name Address	4	BUILDING 18
Name Address	5	2890 north 5th street
Name Address	6	los angeles ca 90134
Name Address	7	
Name Address	8	

Name Line(1) 'JOHN SMITH'
Name Line(2) 'MARY A SMITH'
Name Parsed Elements
Title(1) 'MR'
Title(2) 'MRS'
First Name(1) 'JOHN'
First Name(2) 'MARY A'
First Name(5) 'TOM'
Last Name(1) 'SMITH'
Last Name(2) 'SMITH'
Last Name(5) 'JACKSON'
Name Appendage(5) 'ACCOUNTANT'
Conjunction(1) 'AND'
Output Gender(1) 'M'
Output Gender(2) 'F'
Output Gender(5) 'M'
Organization Information
Organization Line(4) 'SMITH ASSOCIATES'
Organization Parsed Elements
Street Information
Street Line(6) 'BLDG 18'
Street Line(7) '2890 N 5TH ST'
Street Parsed Elements
PreDirectional(7) 'N'
House Number(7) '2890'
Street Name(7) '5TH'
Street Identifier(7) 'ST'
Building Name(6) 'BLDG'
Building Number(6) '18'
Urbanization Information
Urbanization Parsed Elements
Municipality Information
Municipality Parsed Elements
City Information
City Line(8) 'LOS ANGELES CA 90134'
City Parsed Elements
City Name(8) 'LOS ANGELES'
State Code(8) 'CA'
Postal Code(8) '90134'

Demo – Address Validation

PostLocate Home About Contact

[< Back](#)

Input Data

Innovative System

Name of a household, company, location, etc.

790 holliday

Address line 1

Address line 2

Address line 3

Wabash, PA

City, State and/or ZIP code

Response Data

INNOVATIVE SYSTEMS INC

Output Line 1

790 HOLIDAY DR

Output Line 2

PITTSBURGH PA 15220-8127

Output Line 3

Hit Details

Return Code: NormalHit
Country Used: NormalHit
PostalCode: Missing
City: Equal
State: Equal
PreDirection: Equal
Street: Check
StreetType: Missing
PostDirection: Equal
HouseNumber: Exact
Subaddress: Missing
SubaddressType: Missing
Firm: LooseMatch
Locality: Equal

Firm: INNOVATIVE SYSTEMS INC
HouseNumber: 790
City: PITTSBURGH
VanityCity: WABASH
State: PA
PostalCode: 152208127
AddressLine1: 50
AddressLine2: 48
CityLine: 51
StreetName: HOLIDAY
StreetType: DR
RecordType: Firm

Full Search Info

Output Details

Line 0: INNOVATIVE SYSTEMS INC
Line 1: 790 HOLIDAY DR
Line 2: PITTSBURGH PA 15220-8127
Latitude: 40.425852N
Longitude: 80.061619W
GeocodeType: streetlevel

CensusTract: 469000
CensusBlock: 3016
MSA: 6280
CBSA: 38300
MCD: 31256
PlaceCode: 31256
CountyFIPSCode: 003
StateFIPSCode: 42
CountyName: ALLEGHENY

Further Info

AliasFlag: NotAnAlias
BaseAlternateFlag: BaseMatch
AbbrevPrefflag: Not-Applicable
UniqueZipFlag: 48
DeliveryPoint: 90
DeliveryPointCheckDigit: 51
CarrierRoute: C056
LacsFlag: 32
LacsLinkCode: Blank
SuiteLinkCode: NotChecked
ELOTAscDesc: Neither
StateFIPSCode: 42

CountyFIPSCode: 003
CountyName: ALLEGHENY
LowStreetNumber: 790
HighStreetNumber: 790
OddEvenFlag: even
LowSubaddressNumber: 11
HighSubaddressNumber: 11
SubaddressFromHouse: false

Demo – Match Simulator

Match Simulation

Setup Results

Viewing results for: MR ABDUL REZA SHAHLAEE Match MR ABDUL REZA SHAHLAEE Match

Run Results

- Match Potential
- Name Matching
- Match String Status

It's a Match

Match String **C2 E C1 E E E** Match Score **88.31%**

The two records are considered a Match because the resulting **Match String** was deemed worthy of review

Match Field	Client Data	List Data	Match Attribute
Significant Name 1 Reason: Similarity: 80.00% Match Attribute: C2 Show Less ^ <input checked="" type="checkbox"/> Close (C1): At least 65.00% similar <input checked="" type="checkbox"/> Very Close (C2): At least 80.00% similar <input type="checkbox"/> Allow Match with Extensions	ABDUL	ABDOL	VERY CLOSE
Significant Name 2	REZA	REZA	EQUAL
Significant Name 3 Reason: Similarity: 71.43% Match Attribute: C1 Show Less ^ <input checked="" type="checkbox"/> Close (C1): At least 65.00% similar <input checked="" type="checkbox"/> Very Close (C2): At least 75.00% similar <input checked="" type="checkbox"/> Allow Match with Extensions Match with Extensions treated as Close (C1)	SHAHLAE	SHAHLAI	CLOSE
Significant Name 4			EQUAL
Significant Name 5			EQUAL
Significant Name 6			EQUAL

Break down what makes *ABDUL* and *ABDOL* "Very Close" by looking at how the Name field's **Similarity** percentage was classified in the setup stage

Demo – Match Simulator Details

Match Simulation

Setup Results

Run Results

- Match Potential
- Name Matching
- Match String Status

Client Data

Input Name Mr. Abdul-Reza Shahlaee
Parsed Name MR ABDUL REZA SHAHLAEE
Dropped MR

List Data

Input Name ABDOL REZA SHAHLAI
Parsed Name ABDOL REZA SHAHLAI
Dropped

Edit Distance Threshold

Edit Distance Threshold 2
Threshold Met ✔

Show Less ▾

A minimum of two words must be within the edit distance threshold for the records to be further compared.
Words that are shorter than 3 characters long must be exactly equal regardless of the edit distance threshold.

Original Tokens

	ABDOL	REZA	SHAHLAI
ABDUL	1	5	6
REZA	5	0	6
SHAHLAEE	7	7	2

Common, inconsequential words are dropped so that they do not bring back unnecessary results

Demo – Search Example

[Live Demo](#)

Match Simulation

Setup Results

Screening Configuration: DowJonesTest_Final_Continued | v1

Screening Type: Individual Organization

Simulation Type: Search Manual

Client Data: Name: Mr. Abdul-Reza Shahlaee

Search Parameters: Lists: Select All, OFAC Specially Designated Nationals, World-Check, Dow Jones Watchlist, Dow Jones - Test File 2020

Search

15 results were brought back using this Screening Configuration. Each result has a corresponding **Match String/Match Score** that show the quality of the Match

Match String	Match Score	Name
C2 E C1 E E E	88.31%	ABDOL REZA SHAHLAI
C2 E C1 E E E	88.31%	ABDOL-REZA SHAHLAI
C1 E B E E E	86.41%	REZA SHAHLAI
C1 E B E E E	86.41%	REZA SHAHLAI
C1 E B E E E	86.41%	REZA SHAHLAI

Showing 15 results

Simulate

Simulate, step-by-step, how the final Match/No-Match determination was made

Enlighten[®]

Enlighten[®] Data Quality for CRM

What sets Enlighten apart?

Dictionary

- ✓ Millions of words, phrases, and patterns
- ✓ Technology continues to evolve, now in its 14th generation with Enlighten
- ✓ Finding and correcting data is faster and more accurate
- ✓ No need for continual testing and tuning

Pattern Matching

- ✓ Greater accuracy than percentage-based approaches
- ✓ Maintains field-level matches for precision
- ✓ One match string for each use case
- ✓ Total control over priorities and ranking

Language Grammars

- ✓ Linguistic-based grammars for each language
- ✓ Universal grammar for any language
- ✓ Processes data based on the language used
- ✓ No need for custom coding

What sets Innovative Systems apart?

- ✓ **Experience:** Our team has deep knowledge of working with customer-centric data
- ✓ **Technology:** We recruit from top universities and remain a leader in computational linguistics.
- ✓ **Analysts:** We are a recognized leader in technology and customer satisfaction
- ✓ **Flexibility:**
 - ✓ Software: On-premise, SaaS/Cloud, or Hybrid
 - ✓ We can train you to use the software or we can do the work for you (DQaaS)
- ✓ **Volume:** Global data centers process 190 billion+ records annually; meet privacy & GDPR requirements
- ✓ **Access:** 24/7 support
- ✓ **Talent:** Timely access to staff with deep technical expertise

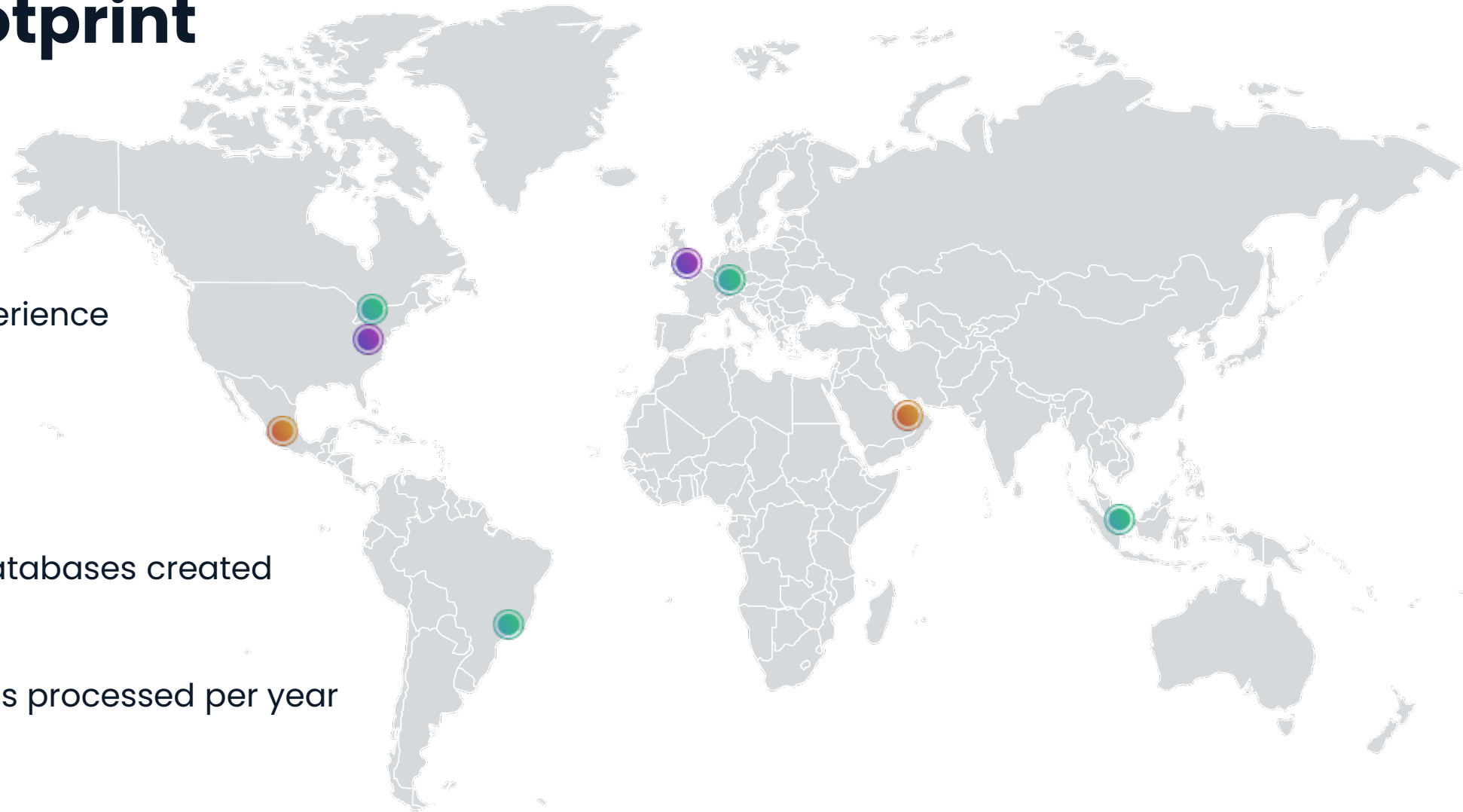
Global Footprint

50+ Years of experience

65+ Countries

1,000s Customer databases created

190+ Billion records processed per year



Locations in: Pittsburgh | London | Dubai | Frankfurt | Mexico City | São Paulo | Singapore | Toronto

Enlighten[®]

Appendix

mott@innovativesystems.com

www.innovativesystems.com

What Type of Data Can Be Used in Grammars?

- **Dictionaries:**

First names, last names, common organization and address terms
(any collection of words in any language)



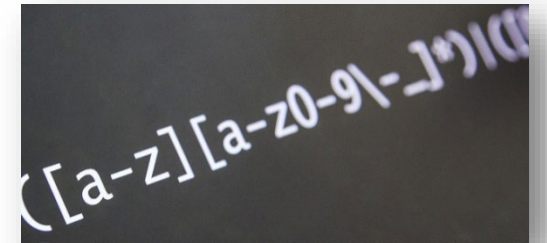
- **Postal standards:**

Country-specific or regional data from any available database
(public, license, or self-owned)



- **Regular expressions:**

Email, telephone, national ID numbers, and part numbers
(any data with a defined structure)



- **Common domains**

Public: email, website, countries, area codes
Organizational: invalid or dummy data



- **Et Cetera**

Anything that can be represented either as a collection of elements or as a pattern of text characters in any language

What do our grammars deliver?

- ✓ Cleansing and Matching **in any language**
 - ✓ Support for special characters such as accent marks, etc. (*Unicode ready*)
 - ✓ The ability to use any field and any number of fields in matching
- ✓ **Custom parsing** to meet unique data needs
 - ✓ Specific requirements by country, industry, client, or type of data
- ✓ Ability to cleanse data **beyond name and address**
 - ✓ Telephone, email, date of birth, tax number, national ID number, etc.
- ✓ Ability to cleanse data beyond **customer or party domain**
 - ✓ e.g., Product, Inventory
- ✓ A **“starter set”** of grammars and configuration options
 - ✓ For rapid deployment

漢

דְּנִיֵּאלָהּ

Núñez

ПРИВЕТ

ΤΥΦΧΨΩ

שלום

和

Innovative Systems Data Quality Products



Enlighten[®]
Data Quality



PostLocate[®]
Address Validation



FinScan[®] Compliance
Sanctions and PEP Screening

Selected Customers

Banking/Financial



Insurance



Manufacturing/Retail



NGOs



World Learning



Enlighten[®]

Questions?

mott@innovativesystems.com

www.innovativesystems.com