



ORACLE

ORACLE

Oracle Enterprise Data Quality

Address Verification Essentials

Product Development

Address Verification in Oracle Enterprise Data Quality

What Does EDQ Address Verification Do?

- **Verifies that addresses exist, and to what level:**
 - For example, the street may exist, but there may be no record of the premise number within that street
- **Standardizes addresses:**
 - Adds missing information
 - Outputs a standard format
 - Where possible, corrects typos
- **Identifies addresses' locations (geocoding):**
 - Adds Latitude and Longitude co-ordinates
- **Searches:**
 - Returns addresses that are similar to the one you entered

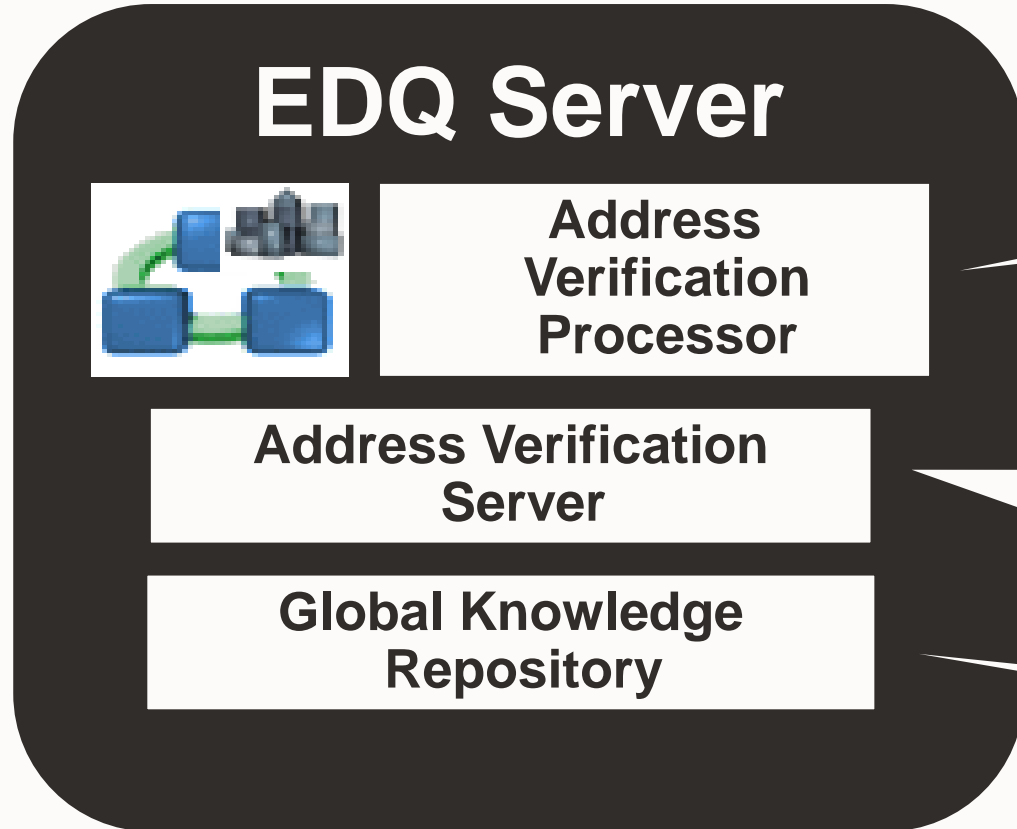
Why Verify and Standardize Addresses?

- **To reduce cost:**
 - Mailing discounts and first time delivery
- **To increase customer satisfaction:**
 - On time deliveries
- **To ensure accurate business and marketing intelligence.**
 - Examples:
 - Locate facilities near concentrations of customers
 - Reliable delivery targets and routes
 - Billing efficiency
- **To improve matching results**

Scope and Usage

- EDQ Address Verification:
 - Works worldwide
 - Runs in real-time and batch
 - Integrates with Enterprise Data Quality:
 - Integration is via EDQ Address Verification processor
 - Used in EDQ Customer Data Services Pack
 - Powerful Summarize Address Verification processor also provided

Three-Part Architecture



Can be dragged onto the EDQ Canvas. Enables you to configure inputs and options. Provided in EDQ distribution.

The brain of address verification. Consists of a set of libraries and a Java API. Must be purchased separately from Oracle.

Data library of worldwide addresses. Must be purchased from OEM partner, GBG Loqate.

What AV Does: Parse and Verify

Input Data: "9070 South Rita Road Suit 1950 Tucson US"

	Parse	Verify
Premise Number	9070	✓
Thoroughfare Name	South Rita Road	✓
Sub-Building	Suit 1950	✓
Locality	Tucson	✓
Administrative Area		
Postal Code		
Country	US	✓

Step 1 – Parse (extract) elements of the address

Step 2 – Verify the elements against information in the Global Knowledge Repository data files

Change character set – transliterate if necessary

What AV Does: Standardize and Geocode

Input Data: "9070 South Rita Road Suit 1950 Tucson US"

Step 3 – Standardize: add missing information and use country postal regulations to format address correctly

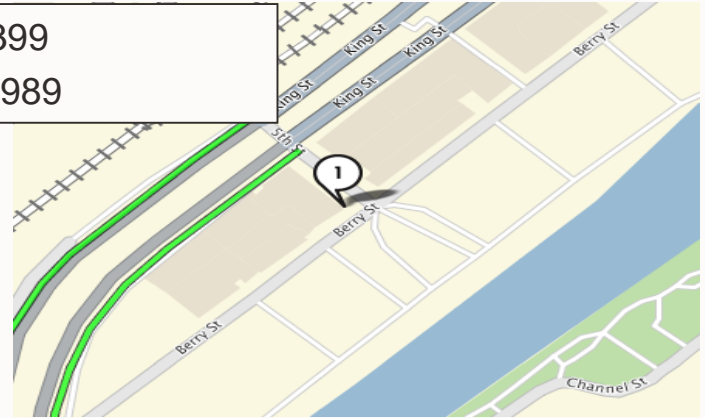
	Standardize
Premise Number	9070
Thoroughfare Name	South Rita Rd
Sub-Building	Ste 1950
Locality	Tucson
Administrative Area	AZ
Postal Code	85747
Country	USA

AV Whole Address
9070 South Rita Rd Ste 1950, Tucson AZ 85747

Step 4 – Find the location

Geocode

Latitude 32.204399
Longitude -110.907989



The Summarize Address Verification Processor

Sample data after Address verification

Summarize Address Verification

Match Score Summary

DASHBOARD address verification

Process

Results Browser

Job: [2] Summarize Address Verification Batch

Value	Count	%
4 - Premise	823	82.3%
5 - Delivery Point	175	17.5%
2 - Locality	2	0.2%

AV Verification Status AV Verification Level AV Raw Verification Level AV Parsing Status AV Lexicon Identifi

- **Provides succinct run-down of address verification results.**
 - Number of addresses verified to each level
 - Postal code statuses
 - and more...

AV is Used in the EDQ Customer Data Services Pack (CDS)

- The Customer Data Services Pack (CDS):
 - Enables EDQ to provide the Data Quality Protection component of a Customer Hub or Customer Relationship Management system
 - Includes an **Address Clean** service that uses Address Verification
- The Address Clean Service:
 - Verifies input address, returning a verification code and description
 - Corrects, standardizes and completes input addresses, provided the address was verified to a sufficient, configurable, level
 - Geocodes the address, returning latitude and longitude co-ordinates, with additional metadata

Configuration Options

Supplying the Addresses to be Verified

- You can input:
 - The whole address in a single attribute, or
 - Address elements in different attributes (Premise, Building, Thoroughfare, Locality, Administrative Area, Country etc.)
 - This can improve parsing and therefore verification results
- **Country should always be input**
 - If it is not populated you can use EDQ to derive a country value
- Reasonably tolerant of omissions / non-standard representations

The screenshot shows the 'Address Verification' software interface. It features a tabbed menu at the top with 'Attributes', 'Options', 'Dashboard', 'Notes', and 'Icon & Family'. The 'Attributes' tab is active, displaying a list of 'Available Attributes' on the left and a configuration panel on the right. The 'Available Attributes' list includes: ID (string), Title (string), FirstName (string), Surname (string), BuildingNumber (string), StreetName (string), City (string), State (string), ZIP (string), Country (string), Phone (string), eMail (string), Comments (string), Gender (string), DoB (string), DateMod (string), Added, Flags, and Incompatible. The configuration panel on the right maps these attributes to various address fields: Organization, Address lines, Double Dependent Locality, Dependent Locality, Locality (mapped to City (City)), Administrative Area (mapped to State (State)), Postal code (mapped to ZIP (ZIP)), Country (mapped to Country (Country)), Postal Box, Sub Building, Building, Premise (mapped to BuildingNumber (BuildingNumber)), and Thoroughfare (mapped to StreetName (StreetName)).

Setting the Address Verification Processor's Options

Address Verification

Attributes Options Dashboard Notes Icon & Family

Processing mode: Verify (Best Match)

Maximum number of results:

Geocode? Yes No

Output address separator: Pipe (|)

Output script: Native

Output Casing: Title

Return field status flags? Yes No

Additional options

<< Show externalized options

OK Cancel

Verify (Best Match) – returns the best single match against AV data files.

Verify (Allow Multiple Results) – returns best single match if possible, or multiple matches if ambiguous.

Search – returns multiple matches.

Geocode - controls whether or not to return Latitude and Longitude coordinates .

Output script – controls which writing system should be used for the output. ('Native' returns the output in its original script).

Setting the Address Verification Processor's Options (Continued)

Address Verification

Attributes Options Dashboard Notes Icon & Family

Processing mode: Verify (Best Match)

Maximum number of results: []

Geocode? Yes No

Output address separator: Pipe (|)

Output script: Native

Output Casing: Title

Return field status flags? Yes No

Additional options

<< Show externalized options

OK

Output Casing – controls the letter case for output fields. Select from 'Upper', 'Title' (default - the first character of each word is upper case, the others are lower case), or 'Lower'.

Return field status flags? – If set to Yes, an additional flag will be generated for each output attribute, with a numeric value indicating the verification status of each. See EDQ's online help for more information.

Additional options – enables you to set any other AV option (see Loqate support website for full details).

Setting the Processing Mode

- **There are three options:**

- Verify (Best Match)

- One-to-one: Checks the reference data and returns the best match
 - Answers the question: which elements of this address exist?

- Verify (Allow Multiple Results)

- Attempts to verify the input address one-to-one, but allows Address Verification to return multiple possible results if the input address has an ambiguous verification result. Results are returned as arrays.
 - You can configure the maximum number of results to be returned

- Search

- One-to-many: checks the reference data and returns multiple matches in arrays
 - You can configure the maximum number of results to be returned
 - Answers the question: what addresses similar to this one exist?

av.AccuracyCode	av.
V44-I44-P5-096	1
P22-I44-P6-100	1
V44-I44-P7-100	1
U00-I44-P2-100	1

V=verified
P=partially verified
U=unverified

av.Address	av.CountryName
{1401 Ponce Leon Blvd 4533 33146 Salzkotten}	{Germany}
{1402 Ponce Leon Blvd 4533 33146 Salzkotten}	{Germany}
{1403 Ponce Leon Blvd 4533 331...	{Germany}
{1445 Ponce Leon Blvd 4533 33146 Salzkotten}	{United States}

Understanding the Results

What information does the Processor Return?

- Verified addresses and address elements.
- Flags indicating verification status and accuracy.

The screenshot shows a tree view of 'Output Attributes' with the following structure:

- Output Attributes:
 - Added (51)
 - av.Address
 - av.CountryName
 - av.DeliveryAddress
 - av.Contact
 - av.Function
 - av.Department
 - av.Organization
 - av.OrganizationName
 - av.OrganizationType
 - av.Unmatched
 - av.Latitude
 - av.Longitude
 - av.GeoDistance
 - Flags (4)
 - av.AccuracyCode
 - av.ResultCount
 - av.MatchScore
 - av.GeoAccuracy

Red arrows point from the following text to the corresponding attributes in the tree:

- Standardized whole address. (points to av.Address)
- Elements of address parsed into own attributes. (points to av.Contact)
- Latitude and Longitude returned if Geocode option is selected (points to av.Latitude and av.Longitude)
- Accuracy codes output as flags (points to av.AccuracyCode, av.ResultCount, and av.MatchScore)

Address Verification Processor Results

Results Browser

Job: Address Verification - File Latest Run: 10-Oct-2014 13:35:44 - 13

Viewing 100 of 1,000 sample records (2,000)

City	State	ZIP	Country	BuildingNumber	StreetName	av.AccuracyCode	av.Address	av.Longitude	av.Latitude	av.GeoAccuracy
KENOSHA	WI	53144	USA	12721	12th Street	V44-I44-P7-100	12721 12th St Kenosha WI 53144-7510	-87.96263	42.63958	P4
WOODINVILLE	WA	98072	USA	17301	133rd Avenue NE	V44-I44-P7-100	17301 133rd Ave Ne Woodinville WA 98072-8563	-122.16223	47.7529	P4
BELLEVUE	WA	98006	USA	4002	134th Avenue SE	V44-I44-P7-100	4002 134th Ave Se Bellevue WA 98006-1316	-122.16195	47.57455	P4
WOODINVILLE	WA	98072	USA	19561	144th Avenue NE	V44-I44-P7-100	19561 144th Ave Ne Woodinville WA 98072-8424	-122.148	47.7529	P4

Inputs

Outputs

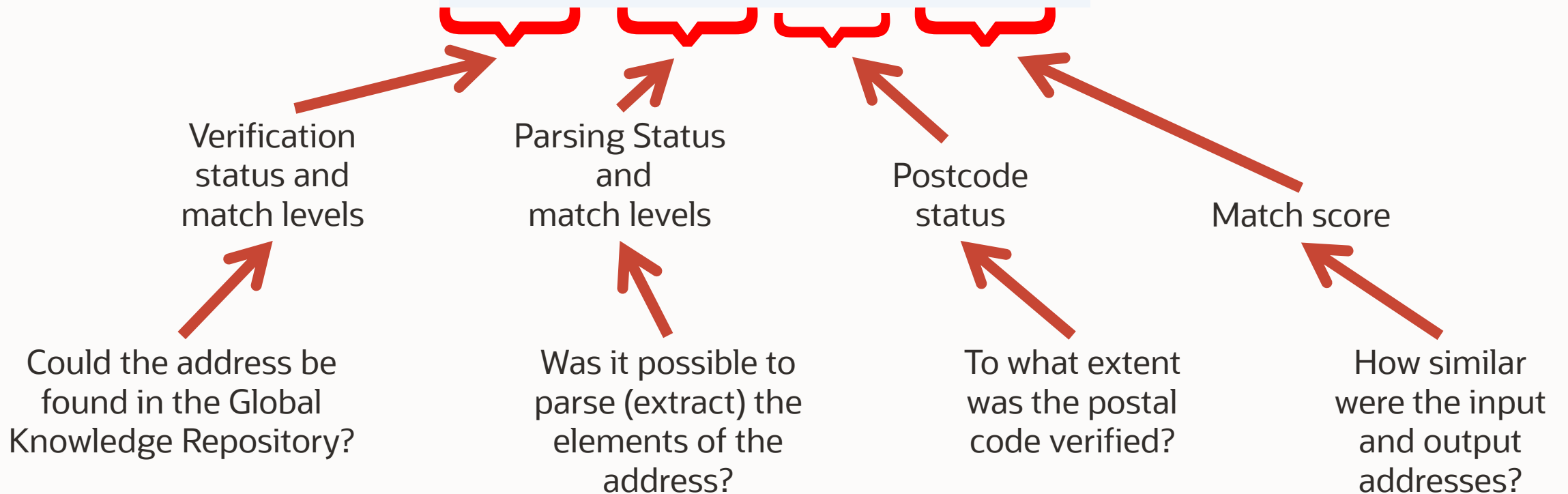
How Much Should you Trust your Address Data?

- **The Address Verification processor generates flags for each address. There are two particularly important flags:**
 - AV.AccuracyCode
 - Tells you
 - Whether and to what extent the address was verified
 - How much it was changed during standardization
 - AV.GeoAccuracy
 - Tells you
 - How the longitude and latitude were determined
 - To what level they are likely to be accurate

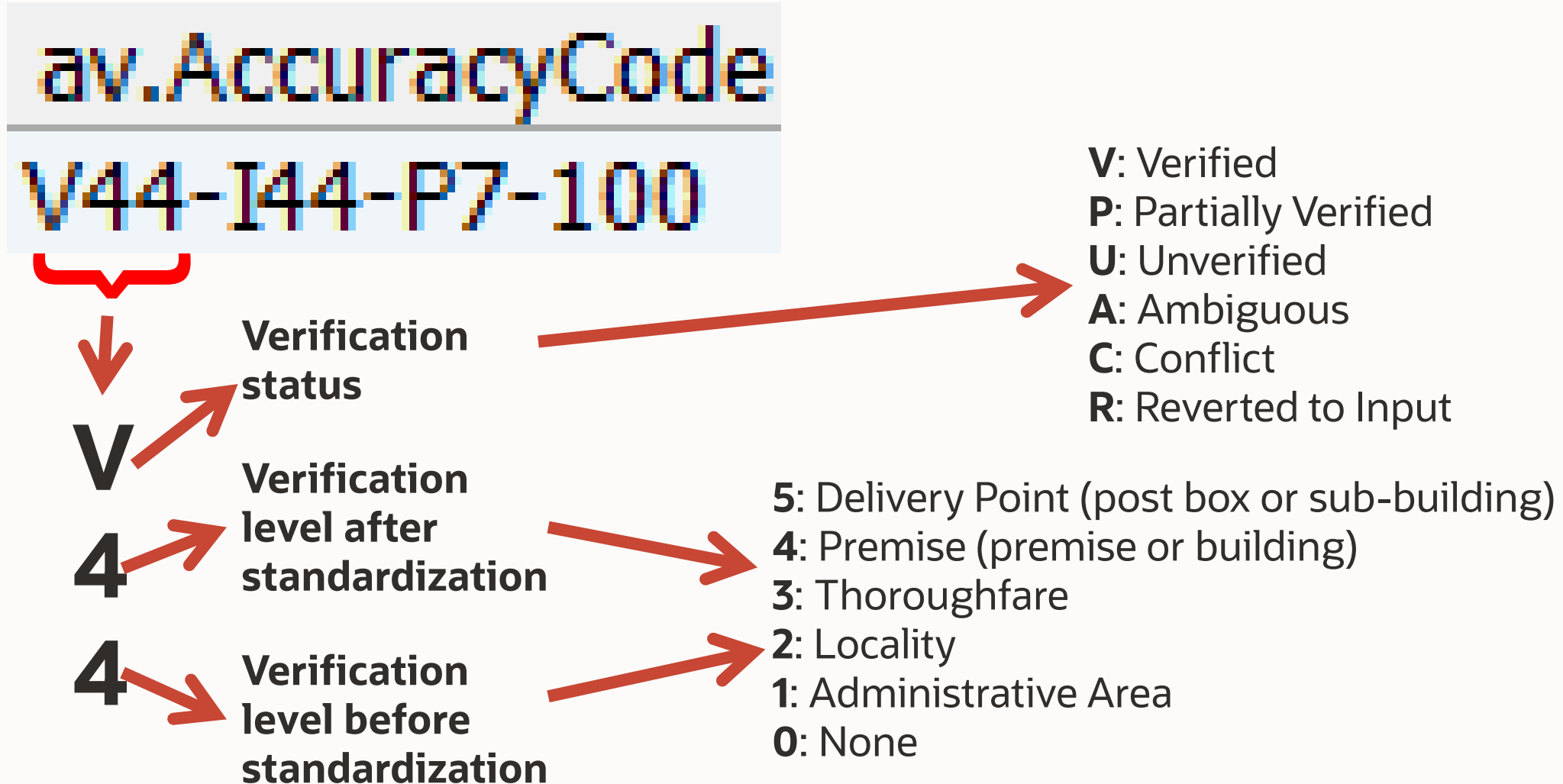
The AV Accuracy Code: Four Segments

av.AccuracyCode

V44-I44-P7-100



Verification Status and Match Levels



Parsing Status and Lexicon and Context Match Levels

av.AccuracyCode
V44-I44-P7-100

Parsing
status

I: Identified and Parsed
U: Unable to Parse

I

Identification
match -
lexicon

5: Delivery Point (post box or sub-building)

4

4: Premise (premise or building)

3: Thoroughfare

4

Identification
match -
context

2: Locality

1: Administrative Area

0: None

Postal Code Status

av.AccuracyCode
V44-I44-P7-100

P7

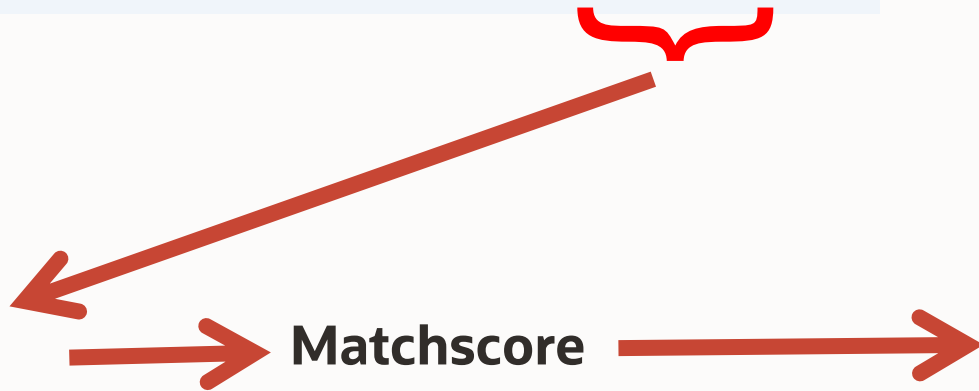
**Postcode
status**

- P8:** PostalCodePrimary and PostalCodeSecondary verified
- P7:** PostalCodePrimary verified, PostalCodeSecondary added or changed
- P6:** PostalCodePrimary verified
- P5:** PostalCodePrimary verified with small change
- P4:** PostalCodePrimary verified with large change
- P3:** PostalCodePrimary added
- P2:** PostalCodePrimary identified by lexicon
- P1:** PostalCodePrimary identified by context
- P0:** PostalCodePrimary empty

Matchscore

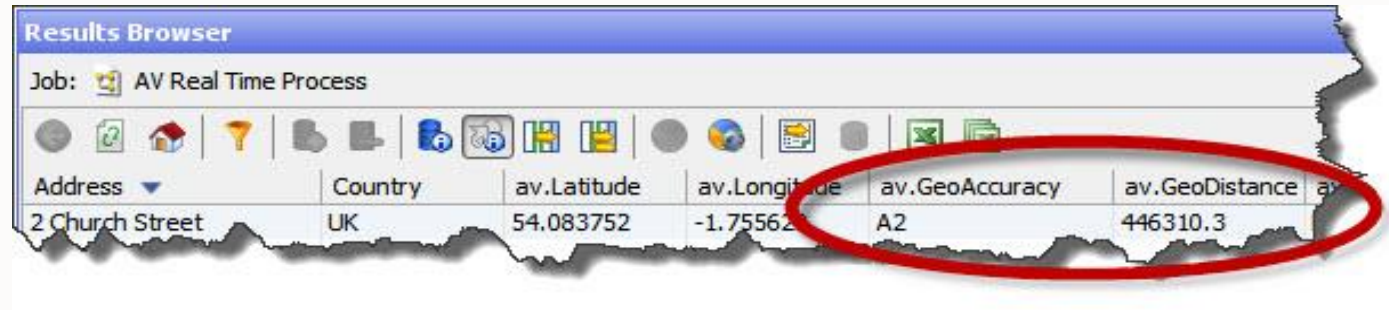
```
av.AccuracyCode  
V44-I44-P7-100
```

100



- Similarity between input data and output data
- Expressed as a percentage
- 100% = complete similarity to the verification level (ignores added data)

Geo Accuracy Code and Geo Distance



Address	Country	av.Latitude	av.Longitude	av.GeoAccuracy	av.GeoDistance
2 Church Street	UK	54.083752	-1.75562	A2	446310.3

Geo Accuracy Code: Geocoding Status

- **P:** Point – a single geocode was found matching the input address
- **I:** Interpolated – a geocode was interpolated from the input address's location in a range
- **A:** Average - multiple candidate geocodes were found to match the input address, and an average of these was returned
- **U:** Unable to geocode - not able to generate geocode for input address

Geo Accuracy Code: Geocoding Level

- **5:** Delivery Point (PostBox or SubBuilding)
- **4:** Premise (Premise or Building)
- **3:** Thoroughfare
- **2:** Locality
- **1:** AdministrativeArea
- **0:** None

Geo Distance

Radius of accuracy in meters – indicates likely maximum distance between geocode and physical location

Using the Flags

- **The AV flags provide important management information about**
 - The level to which you can trust your address data
 - The level to which you can trust the latitude and longitude
- **The flags may influence the way in which you use your address data, including in downstream EDQ processing**

Enterprise Data Quality Address Verification Self-Paced Training

- **Address Verification Self-Paced Training is available from the Oracle Learning Library:**

https://apex.oracle.com/pls/apex/f?p=44785:24:17531673927993:PRODUCT:::P24_CONTENT_ID,P24_PREV_PAGE,P24_PROD_SECTION_GRP_ID:10244,141,3019

Our mission is to help people
see data in new ways, discover
insights, unlock endless possibilities.

