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

**EDQ Server**

- **Address Verification Processor**

  - 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.

- **Address Verification Server**

- **Global Knowledge Repository**

  - 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"

<table>
<thead>
<tr>
<th></th>
<th>Parse</th>
<th>Verify</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premise Number</td>
<td>9070</td>
<td>✔</td>
</tr>
<tr>
<td>Thoroughfare Name</td>
<td>South Rita Road</td>
<td>✔</td>
</tr>
<tr>
<td>Sub-Building</td>
<td>Suit 1950</td>
<td>✔</td>
</tr>
<tr>
<td>Locality</td>
<td>Tucson</td>
<td>✔</td>
</tr>
<tr>
<td>Administrative Area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postal Code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>US</td>
<td>✔</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Standardize</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Premise Number</td>
<td>9070</td>
</tr>
<tr>
<td>Thoroughfare Name</td>
<td>South Rita Rd</td>
</tr>
<tr>
<td>Sub-Building</td>
<td>Ste 1950</td>
</tr>
<tr>
<td>Locality</td>
<td>Tucson</td>
</tr>
<tr>
<td>Administrative Area</td>
<td>AZ</td>
</tr>
<tr>
<td>Postal Code</td>
<td>85747</td>
</tr>
<tr>
<td>Country</td>
<td>USA</td>
</tr>
</tbody>
</table>

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

- 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
Setting the Address Verification Processor’s Options

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)

- **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.

- **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’.

- **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?

V=verified  
P=partially verified  
U=unverified
Understanding the Results
What information does the Processor Return?

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

Standardized whole address.

Elements of address parsed into own attributes.

Latitude and Longitude returned if Geocode option is selected.

Accuracy codes output as flags.
## Address Verification Processor Results

<table>
<thead>
<tr>
<th>City</th>
<th>State</th>
<th>ZIP</th>
<th>Country</th>
<th>BuildingNumber</th>
<th>StreetName</th>
<th>av.AccuracyCode</th>
<th>av.Address</th>
<th>av.Longitude</th>
<th>av.Latitude</th>
<th>av.GeoAccuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>KENOSHA</td>
<td>WI</td>
<td>53144</td>
<td>USA</td>
<td>12721</td>
<td>12th Street</td>
<td>V44-H4-P7-100</td>
<td>12721 12th St</td>
<td>Kenosha WI 53144-7510</td>
<td>-87.96263</td>
<td>42.63958</td>
</tr>
<tr>
<td>WOODINVILLE</td>
<td>WA</td>
<td>98072</td>
<td>USA</td>
<td>17301</td>
<td>133rd Avenue NE</td>
<td>V44-H4-P7-100</td>
<td>17301 133rd Ave</td>
<td>Woodinville WA 98072-8553</td>
<td>-122.16223</td>
<td>47.7529</td>
</tr>
<tr>
<td>BELLEVUE</td>
<td>WA</td>
<td>98025</td>
<td>USA</td>
<td>4002</td>
<td>134th Avenue SE</td>
<td>V44-H4-P7-100</td>
<td>4002 134th Ave</td>
<td>Bellevue WA 98025-1316</td>
<td>-122.16195</td>
<td>47.57455</td>
</tr>
<tr>
<td>WOODINVILLE</td>
<td>WA</td>
<td>98072</td>
<td>USA</td>
<td></td>
<td>144th Avenue NW</td>
<td>V44-H4-P7-100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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

- Verification status and match levels
- Parsing Status and match levels
- Postcode status
- Match score

Could the address be found in the Global Knowledge Repository?
Was it possible to parse (extract) the elements of the address?
To what extent was the postal code verified?
How similar were the input and output addresses?
Verification Status and Match Levels

**av.AccuracyCode**

V44-I44-P7-100

- **V**: Verified
- **P**: Partially Verified
- **U**: Unverified
- **A**: Ambiguous
- **C**: Conflict
- **R**: Reverted to Input

**Verification status**
- V

**Verification level after standardization**
- 4

**Verification level before standardization**
- 4

**Verification level after standardization**:
- 5: Delivery Point (post box or sub-building)
- 4: Premise (premise or building)
- 3: Thoroughfare
- 2: Locality
- 1: Administrative Area
- 0: None
Parsing Status and Lexicon and Context Match Levels

**av.AccuracyCode**

**V44-I44-P7-100**

- Parsing status:
  - I: Identified andParsed
  - U: Unable to Parse

**Identification match – lexicon**
- 5: Delivery Point (post box or sub-building)
- 4: Premise (premise or building)
- 3: Thoroughfare
- 2: Locality
- 1: Administrative Area
- 0: None

**Identification match – context**
Postal Code Status

V44-I44-P7-100

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

100

av. Accuracy Code
V44-I44-P7-100

- 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**

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