Oracle® Enterprise Data Quality

Customer Data Services Pack Release Notes

Release 11g R1 (11.1.1.9)

E56081-01

April 2015

This document contains the release information for Oracle Enterprise Data Quality (EDQ) Customer Services Data Pack Release 11g R1 (11.1.1.9).

Oracle recommends you review its contents before installing, or working with the EDQ-CDS product.

1 New Features and Improvements

The following sections describe the new features introduced in each release, as well as, improvements:

- Section 1.1, "Release 11g R1 (11.1.1.9.0)"
- Section 1.2, "Release 11g R1 (11.1.1.7.4)"
- Section 1.3, "Release 11g R1 (11.1.1.7.3)"

1.1 Release 11g R1 (11.1.1.9.0)

This section addresses Release 11g R1 (11.1.1.9.0).

1.1.1 Search Mode Available in EDQ Address Cleaning Service

The CDS Address Clean service now supports Search mode. If the input parameter `mode` is set to `S`, the AddressClean service will search for address matches for each input address, and may return multiple results.

When using Search mode, the service will attempt to find the closest real address (in the installed Loqate data) for a partially input address. Search mode only supports searching for whole addresses. It is not suitable to return a subset of attributes based on partial input – for example it does not support the return of a list of postal codes (only) for a partial postal code input. [17518774]

See Address Search in *Oracle Enterprise Data Quality Customer Data Services Pack Installation Guide* for more information.

1.1.2 Enable Name-Only Matching Rules

To allow matches on sparse data, you may want to match individuals using name information only. Name-only match rules are now enabled by default in the CDS matching process. [18300496]

See Name-Only Matching in *Oracle Enterprise Data Quality Customer Data Services Pack Installation Guide* for more information.

1.1.3 Add Postal_Plus4_Code to CDS Address Cleaning Interface

The Address Clean web service now supports interface attributes for a primary and secondary postal code. If the existing attribute, `postalcode`, is blank, then the new attributes `postalcodeprimary` and `postalcodesecondary` are used.

The attribute `postalcodeprimary` is the first part of a 2-part postal code, for example the ZIP code of a US address, while the attribute `postalcodesecondary` is the second part of a 2-part postal code, for example the +4 part of a US address ZIP+4 code. [18167785]

See *Oracle Enterprise Data Quality Customer Data Services Pack Installation Guide* for more information.

1.1.4 Enhance JMS Triggers to Support Job Cancellation

Oracle Enterprise Data Quality allows you to cancel running jobs. The JMS trigger files and configuration files have been updated to support job cancellation. [18273284]

1.1.5 Improved Non-Latin Local Script Matching and Cross Script Matching

This feature improves matching on data that is written in a non-latin local script, and enables matching between two records with the same name, one written in non-latin characters (Chinese, Korean, Japanese) and the other in a latin script. Oracle Enterprise Data Quality now generates the script-to-script and script-latin maps for every script. [18241002]

1.1.6 Support Multiple Value Account Names in Individual Matching

The Individual Match service has been extended to support matching on multiple account names. A new attribute `accountnameid` has been added, which can be used to identify, in the case of a match, which account name matched. [16762149]

See *Oracle Enterprise Data Quality Customer Data Services Pack Installation Guide* for more information.

1.1.7 Match Results Can Show the Matched Email and Matched Phones

Match results now include the ids of matched email addresses and phone numbers, if the match rule included email address and/or phone number as matching attributes. [16762103]

See *Oracle Enterprise Data Quality Customer Data Services Pack Installation Guide* for more information.

1.2 Release 11g R1 (11.1.1.7.4)

This section addresses Release 11g R1 (11.1.1.7.4).

1.2.1 Improve Matching of Company Names with Initials and Acronyms

EDQ-CDS does not find a match for some entities where the entity name is an acronym, standardization or abbreviation. The following new entity match rule groups have been created: [17861632]

[E230] Standardized full name acronym exact
[E240] Full name without suffixes acronym exact
[E250] Full name without suffixes acronym contains

See *Oracle Enterprise Data Quality Customer Data Services Pack Matching Guide* for more details about these new rules

1.2.2 Improve Matching Involving Stronger Address Matches

EDQ-CDS does not find match for some entities where there is a strong address match and only a weak name match. The following new entity match rule groups have been created:

[E260] Entity name without suffixes loose typos
[E270] Full name without suffixes first token

The following new entity match rule has been created: [18048606]

[E900AA] Full Address Exact

1.3 Release 11g R1 (11.1.1.7.3)

This section addresses Release 11g R1 (11.1.1.7.3).

1.3.1 Generic Batch Jobs

New generic batch jobs for clustering and matching have been added to the product. These jobs allow an external data source to easily be connected to the EDQ-CDS services with minimal configuration effort. The jobs can also be used as the basis for customization of the product. [SIE-271]

1.3.2 Matching Improvements

The following Matching improvements are delivered: [SIE-272, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, and 293]

- A new Elimination Rule for entities and individuals disabled by default, to eliminate Driver-Driver matches was added. It can be enabled in your Run Profile using the externalized property. For example:

  ```
  phase.*.process.Match\ -\ Entity.[ELIM010C]\ ELIMINATE\ MATCHES\ BETWEEN\ TWO\
  DRIVER\ RECORDS.entity_match_rules_enabled = true
  ```

- The entity and individual cluster limits are now configurable in your Run Profile using the externalized properties. For example:

  ```
  phase.*.process.Match\ -\ Entity.*.entity_match_cluster_comparison_limit =
  15000
  phase.*.process.Match\ -\ Entity.*.entity_match_cluster_group_limit = 0
  ```

- The full address match rules no longer use the value of the `subadminarea` attribute in their comparisons.
- With matching services, leading zeroes are stripped only on numeric `postalcodes` to avoid a numeric `postalcode` reinterpreted as a number by an external program where leading zeroes are automatically stripped. For example, Excel may reformat numeric `postalcodes` as a number by removing the leading zeroes. This is not enabled by default in the Run Profile. If there are any alpha characters present, the leading zeroes are not stripped.
- The address conflict entity rules, `Name[...]Address Conflict`, can be disabled using the externalized properties. For example:

  ```
  phase.*.process.Match\ -\ Entity.[E040V]\ Script\ full\ name\ without\
  suffixes\ exact\;\ address\ conflict.entity_match_rules_enabled = false
  ```

- New entity match rules, `Name[...]Address 1 Typo, City, Country`, have been added to the existing match rule groups. These rule ensure that the matching

weight is against the address line 1, city and country rather than the `postalcode`. This means that the `postalcode` can be different in any number of ways and a very similar `address1`, and exact `city` and `country` will be a match.

- The entity name standardization processes and processors have been changed to ensure that whitespace surrounding ampersands (&) in `name` fields are correctly normalized.
- A new entity cluster with the prefix, `NMA` (`Full Name Metaphone, Address No Numbers`), has been added.
- New entity match rules and group `[E125] Full name all words shorter with typos`, have been added. These rules allow typo-tolerant matching on full entity names where the shorter name is fully contained within the longer name.
- The Individual Cluster group with the prefix, `LMC`, has been changed from level 3 to level 2.

2 Bugs and Issues Resolved

The following sections describe the bugs and issues that are resolved in each release:

- Section 2.1, "Release 11g R1 (11.1.1.9.0)"
- Section 2.2, "Release 11g R1 (11.1.1.7.4)"
- Section 2.3, "Release 11g R1 (11.1.1.7.3)"

2.1 Release 11g R1 (11.1.1.9.0)

This section addresses Release 11g R1 (11.1.1.9.0).

2.1.1 Definition for Rule E125V is Inconsistent

The rule definition for the E125V rule in the Entity Match process now has the `Full name distilled WMC tolerant 2+` comparison set, the same as the other E125 rules. [17468960]

2.1.2 Defaults for Decision Keys in the Match Processors are Incorrect

For real-time performance reasons, the default decision keys settings (visible when you enable Match Review) have been changed to `Selection` with none selected for the Address, Entity and Individual Match processors. [17568554]

2.1.3 Incorrect IDs Sometimes Returned in Match

The IDs returned from match when using a blank ID on the driver record can be incorrect.

The match configuration has been updated to correctly prepare and return blank ID fields. [17582071]

2.1.4 Pipe Character in Input Data Causes Incorrect Results

When present in input data, the pipe character was incorrectly interpreted as a delimiter character causing incorrect results.

Oracle Enterprise Data Quality Customer Data Services Pack Matching Guide documents that for the interface fields which do not accept multiple values (identified in the tables), you should avoid using the pipe character, or double-pipe character. [17714016]

2.1.5 Address Clean Search Mode Does Not Return Organizations

To avoid duplicate search results (one with an organization at an address, and another with just the address), the output returned from Search results has been adjusted by adding in the organization output from Address Verification to the address fields in the output. [17741377]

2.1.6 Address Cleaning Strips Out Organization

The latest address cleaning functionality uses the "Delivery Address" to create the return address. This is convenient because it strips out the locality and postal code. However, it also stripped out the organization, if there was one.

The Address Clean service has been updated to add the organization, if there is one, back into the return address. [16431169]

2.2 Release 11g R1 (11.1.1.7.4)

This section addresses Release 11g R1 (11.1.1.7.4).

2.2.1 EDQ-CDS Is Not Matching Some Company Names with Abbreviations

EDQ-CDS does not find a match for some entities where the entity name is an acronym and some or all of the acronym appears in the entity phrase strip list data.

All 2, 3, and 4 letter space-separated acronyms have been removed from the Entity Full Latin Phrase Strip List by deactivating these entries in the initialization project and removing these rows from the `.jmp` files. [18237187]

2.3 Release 11g R1 (11.1.1.7.3)

This section addresses Release 11g R1 (11.1.1.7.3).

2.3.1 Use Provided Country Value in Preference to Default Code in Address Cleaning Service

The Address Cleaning service has been changed so that it will always use the country code you provide rather than the default country code in the Run Profile. If you do not provide a country code or it cannot be mapped to the standard country codes in EDQ-CDS, then EDQ-CDS attempts to derive country from the address data provided. When this is not possible, the default country code in the Run Profile is used. [16969621]

2.3.2 Entity Script to Latin Maps and Strip Lists Have Duplicates

The reference data entities with duplicates have been corrected as follows: [16743518]

- Ent Name - International Script Phrase Strip List — Removed the duplicate for Thailand and changed the entries that had multiple data sources so that only one entry exists with pipe delimited data source values rather than multiple rows in the data.
- Ent Name - International Script Token Strip List — Changed entries that contained multiple data sources to have only one entry with pipe delimited data source values rather than multiple rows in the data.
- Ent Name - International Script to Latin Phrase Map — De-duplicated the arabic script value that maps to both General People's Congress (GPC) and GP by only including the GPC entry (GPC is from Saudi Arabia and GP is from Iran.)

- Ent Name - International Script to Latin Token Map — Removed the duplicate entry, mapping ЗАТ to 'PLC', in the Russian data source.

2.3.3 Russian Character Replace Map Should Not Contain Both Upper and Lower Case

The Russian character replace map in the Transliterate processor had both the upper and lower case versions of the Cyrillic characters. When this was used with the 'ignore case' option in the replace text, it caused EDQ display a duplicate keys error. The Transliterate processor has been changed to map the three Cyrillic characters that had different versions for the upper and lower case to the character given for the lower case versions. [16622146]

3 Related Documents

For more information, see the following documents in the Oracle Enterprise Data Quality documentation set:

- *Oracle Enterprise Data Quality Installation Guide*
- *Oracle Enterprise Data Quality Architecture Guide*
- *Oracle Enterprise Data Quality Customer Data Services Siebel Integration Guide*
- *Oracle Enterprise Data Quality Customer Data Services Siebel Connector Installation Guide*

See the latest version of this and all documents in the Oracle Enterprise Data Quality Documentation website at

http://download.oracle.com/docs/cd/E48549_01/index.htm

4 Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at
http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc.

Access to Oracle Support

Oracle customers that have purchased support have access to electronic support through My Oracle Support. For information, visit
http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info or visit
http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs if you are hearing impaired.

Oracle Enterprise Data Quality Customer Data Services Pack Release Notes, Release 11g R1 (11.1.1.9)
E56081-01

Copyright © 2013, 2015, Oracle and/or its affiliates. All rights reserved.

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use

in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate failsafe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information about content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services unless otherwise set forth in an applicable agreement between you and Oracle. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services, except as set forth in an applicable agreement between you and Oracle.