



Content Objects Summary

- SAP BusinessObjects Data Services 4.1 Support Package 1 (14.1.1)

2012-12-04

Copyright

© 2012 SAP AG. All rights reserved. SAP, R/3, SAP NetWeaver, Duet, PartnerEdge, ByDesign, SAP BusinessObjects Explorer, StreamWork, SAP HANA and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and other countries. Business Objects and the Business Objects logo, BusinessObjects, Crystal Reports, Crystal Decisions, Web Intelligence, Xcelsius, and other Business Objects products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Business Objects Software Ltd. Business Objects is an SAP company. Sybase and Adaptive Server, iAnywhere, Sybase 365, SQL Anywhere, and other Sybase products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Sybase, Inc. Sybase is an SAP company. Crossgate, m@gic EDDY, B2B 360°, B2B 360° Services are registered trademarks of Crossgate AG in Germany and other countries. Crossgate is an SAP company. All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary. These materials are subject to change without notice. These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

2012-12-04

Contents

Chapter 1	Overview.....	5
1.1	Documentation set for SAP BusinessObjects Data Services content objects.....	5
1.2	SAP BusinessObjects information resources.....	6
Chapter 2	Blueprint versions.....	9
Chapter 3	Available blueprints.....	11
Chapter 4	Data Quality Management regional blueprints.....	15
4.1	Data Quality Management Blueprints – Brazil.....	15
4.2	Data Quality Management Blueprints – China.....	16
4.3	Data Quality Management Blueprints – France	18
4.4	Data Quality Management Blueprints – Germany.....	19
4.5	Data Quality Management Blueprints – Global.....	21
4.6	Data Quality Management Blueprints – India.....	22
4.7	Data Quality Management Blueprints – Japan.....	23
4.8	Data Quality Management Blueprints – Mexico.....	24
4.9	Data Quality Management Blueprints – USA.....	26
Chapter 5	Data Quality Management Match blueprints.....	29
5.1	Data Quality Management Blueprints – Match.....	29
Chapter 6	Text Data Processing Language blueprints.....	31
6.1	Text Data Processing Blueprints – English.....	31
6.2	Text Data Processing Blueprints – German.....	32
6.3	Text Data Processing Blueprints – Japanese.....	33
Chapter 7	Text Data Processing Data Quality Management blueprints.....	35
7.1	Text Data Processing Blueprints – Data Quality Management.....	35
7.1.1	Data flow transform descriptions.....	36

Chapter 8	Other tools.....	39
8.1	Data Quality Management custom functions.....	39
8.2	Text Data Processing – Entity Extraction Dictionary File Generator	45
Chapter 9	Glossary.....	47
Index		53

Overview

We've identified a number of common scenarios that you are likely to perform with SAP BusinessObjects Data Services. For each scenario, we've included a blueprint that is already set up to solve the business problem in that scenario. Each blueprint contains the necessary project, jobs, data flows, file formats, sample data, template tables, and custom functions to run the data flows in your environment with only a few modifications.

You can download the blueprint packages from the SAP Community Network. On the website, we periodically post new and updated blueprints, custom functions, best practices, whitepapers, and other content. You can refer to this site frequently for updated content and use the forums to provide us with any questions or requests you may have. We've also provided the ability for you to upload and share any content that you've developed with the rest of the SAP BusinessObjects Data Services development community (for instructions on uploading content, see *Share Your Knowledge in SCN Topic Spaces* at <https://www.sdn.sap.com/irj/scn/submitcontent>).

Instructions for downloading and installing the content objects are also located on the SAP Community Network website.

1.1 Documentation set for SAP BusinessObjects Data Services content objects

You should become familiar with all of the pieces of documentation that relate to the SAP BusinessObjects Data Services blueprints and other content objects.

Document	What this document provides
<i>Content Objects Summary</i>	Lists all of the available blueprints and other content objects and the jobs and other objects that they contain.
<i>Content Objects What's New</i>	Highlights the new and enhanced blueprints and other content objects available for this release.
<i>Data Quality Management Custom Functions User's Guide</i>	Contains instructions for downloading and importing custom functions.
<i>Data Quality Management Match Blueprints User's Guide</i>	Contains a list of available Data Quality Management Match blueprints and instructions for downloading, configuring, and running them.

Document	What this document provides
<i>Data Quality Management Regional Blueprints User's Guide</i>	Contains a list of available Data Quality Management regional blueprints and instructions for downloading, configuring, and running them.
<i>Text Data Processing Data Quality Management Blueprints User's Guide</i>	Contains a list of available Text Data Processing Data Quality Management blueprints and instructions for downloading, configuring, and running them.
<i>Text Data Processing Language Blueprints User's Guide</i>	Contains a list of available Text Data Processing Language blueprints and instructions for downloading, configuring, and running them.

1.2 SAP BusinessObjects information resources

A global network of SAP BusinessObjects technology experts provides customer support, education, and consulting to ensure maximum information management benefit to your business.

Useful addresses at a glance:

Address	Content
<p>Customer Support, Consulting, and Education services</p> <p>http://service.sap.com/</p>	<p>Information about SAP Business User Support programs, as well as links to technical articles, downloads, and online forums. Consulting services can provide you with information about how SAP BusinessObjects can help maximize your information management investment. Education services can provide information about training options and modules. From traditional classroom learning to targeted e-learning seminars, SAP BusinessObjects can offer a training package to suit your learning needs and preferred learning style.</p>
<p>Product documentation</p> <p>http://help.sap.com/bods/</p>	<p>SAP BusinessObjects product documentation.</p>
<p>Supported Platforms (Product Availability Matrix)</p> <p>https://service.sap.com/PAM</p>	<p>Get information about supported platforms for SAP BusinessObjects Data Services.</p> <p>Use the search function to search for Data Services. Click the link for the version of Data Services you are searching for.</p>

Blueprint versions

The following table shows the version of the Data Quality Management and Text Data Processing blueprints that can be used for SAP BusinessObjects Data Services. The blueprint version is displayed on the SAP BusinessObjects Data Services Blueprints page of the SAP Community Network website.

Data Services version	Blueprint version	Blueprints available
4.1.1	4.1.1	Data Quality Management Text Data Processing
4.1	4.1	Data Quality Management
XI 4.0	XI 4.0	Data Quality Text Data Processing
XI 3.2	XI 3.2	Data Quality
XI 3.1	XI 3.0	Data Quality
XI 3.0	XI 3.0	Data Quality

Available blueprints

To help you compare the available blueprints and decide which to download, see the following table.

For instructions on downloading and setting up the blueprints and other tools, see the appropriate *User's Guide*.

Data Quality Management regional blueprints

Blueprint	Description
Data Quality Management Blueprints – Brazil	Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in Brazil.
Data Quality Management Blueprints – China	Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in China.
Data Quality Management Blueprints – France	Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in France.
Data Quality Management Blueprints – Germany	Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in Germany.
Data Quality Management Blueprints – Global	Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data when the data consists of multiple countries.
Data Quality Management Blueprints – India	Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in India.
Data Quality Management Blueprints – Japan	Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in Japan.
Data Quality Management Blueprints – Mexico	Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in Mexico.

Blueprint	Description
Data Quality Management Blueprints – USA	Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in the United States.

Data Quality Management Match blueprints

Blueprint	Description
Data Quality Management Blueprints – Match	Contains miscellaneous jobs configured to illustrate best practice settings for specific Data Quality Management matching use cases.

Text Data Processing Language blueprints

Blueprint	Description
Text Data Processing Blueprints – English	Contains sample jobs configured to illustrate best practice settings for common Text Data Processing use cases involving unstructured text in the English language.
Text Data Processing Blueprints – German	Contains sample jobs configured to illustrate best practice settings for common Text Data Processing use cases involving unstructured text in the German language.
Text Data Processing Blueprints – Japanese	Contains a sample job configured to illustrate best practice settings for common Text Data Processing use cases involving multi-byte content in the Japanese language. It also showcases enriched support for core extraction pre-defined entity types.

Text Data Processing Data Quality Management blueprints

Blueprint	Description
Text Data Processing Blueprints – Data Quality Management	Contains sample jobs configured to illustrate the use of Text Data Processing in conjunction with Data Quality Management. It also helps you visualize the extracted concepts and sentiments using an SAP BusinessObjects BI 4.0 Universe and SAP BusinessObjects Web Intelligence reports.

Other tools

Blueprint	Description
Data Quality Management Custom Functions	Contains custom functions that perform additional manipulation of data that is not part of the functionality of Data Quality Management transforms, but are common functions that assist with the cleansing and matching of party data.
Text Data Processing – Entity Extraction Dictionary File Generator	An Excel spreadsheet with a macro that generates a dictionary source file based on the content in the spreadsheet and compiles the source file into a ready-to-use dictionary file for the Text Data Processing Entity Extraction transform.

Related Topics

- [Data Quality Management regional blueprints](#)
- [Data Quality Management Match blueprints](#)
- [Text Data Processing Language blueprints](#)
- [Text Data Processing Data Quality Management blueprints](#)
- [Other tools](#)

Data Quality Management regional blueprints

The following are the regional blueprints provided for common Data Quality Management use cases, and a list of the jobs and other objects that they contain.

For more information, see the *Regional Blueprints User's Guide*.

4.1 Data Quality Management Blueprints – Brazil

Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in Brazil.

This blueprints package includes the following jobs:

Job	Description
DqmBlueprintBrazil_AddressCleanse	Validates, cleanses, and standardizes address data.
DqmBlueprintBrazil_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
DqmBlueprintBrazil_CustomerValidation	Includes data quality validation rules to route records for remediation that fail the validation requirements.
DqmBlueprintBrazil_LoadDelta	Cleanses delta data and performs matching against a database, such as a CRM, to identify new records to be inserted and matching records to be updated in the database.
DqmBlueprintBrazil_LoadInitial	Performs an initial cleansing and removal of duplicates for a customer database, such as a CRM, storing the data in a way that provides for optimal matching in the future either with delta loads or with individual transactions.
DqmBlueprintBrazil_MatchAssociative	Cleanses party data and performs duplicate detection based on similar name and address, name and phone, or name and email, combining the three matching results to uncover hidden duplicates.
DqmBlueprintBrazil_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.

Job	Description
DqmBlueprintBrazil_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.
DqmBlueprintBrazil_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
DqmBlueprintBrazil_RtAddressDataCleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
DqmBlueprintBrazil_RtAddressSuggestions	Validates, cleanses, and standardizes address data given a partial address on input and drilling down by selecting from a pick list of valid addresses.
DqmBlueprintBrazil_RtMatchConsumer	Cleanses a single transaction and performs matching against a database, such as a CRM, to identify whether the record exists in the database.

This blueprints package includes the following additional objects:

- Two custom functions that prepare address data for optimal matching in order to identify accurate duplicate addresses in Brazil.
- A custom function that removes "noise" from firm data for optimal matching in order to identify accurate duplicate firms in Brazil.
- A custom function that formats Brazilian phone numbers.
- A custom function that provides control of the prename for which the gender of a person is likely, defaulting to "Sr." and "Sra."
- A custom function that converts the information code from the address cleansing process to a description in English.

4.2 Data Quality Management Blueprints – China

Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in China.

This blueprints package includes the following jobs:

Job	Description
DqmBlueprintChina_AddressCleanse	Validates, cleanses, and standardizes address data.
DqmBlueprintChina_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.

Job	Description
DqmBlueprintChina_CustomerValidation	Includes data quality validation rules to route records for remediation that fail the validation requirements.
DqmBlueprintChina_LoadDelta	Cleanses delta data and performs matching against a database, such as a CRM, to identify new records to be inserted and matching records to be updated in the database.
DqmBlueprintChina_LoadInitial	Performs an initial cleansing and removal of duplicates for a customer database, such as a CRM, storing the data in a way that provides for optimal matching in the future either with delta loads or with individual transactions.
DqmBlueprintChina_MatchAssociative	Cleanses party data and performs duplicate detection based on similar name and address, name and phone, or name and email, combining the three matching results to uncover hidden duplicates.
DqmBlueprintChina_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
DqmBlueprintChina_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.
DqmBlueprintChina_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
DqmBlueprintChina_RtAddressDataCleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
DqmBlueprintChina_RtMatchConsumer	Cleanses a single transaction and performs matching against a database, such as a CRM, to identify whether the record exists in the database.

This blueprints package includes the following additional objects:

- A custom function that removes "noise" from firm data for optimal matching in order to identify accurate duplicate firms in China.
- A custom function that removes "noise" from building names for optimal matching in order to identify accurate duplicate addresses in China.
- A custom function that formats Chinese phone numbers.
- A custom function that converts the information code from the address cleansing process to a description in Chinese.

4.3 Data Quality Management Blueprints – France

Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in France.

This blueprints package includes the following jobs:

Job	Description
DqmBlueprintFrance_AddressCleanse	Validates, cleanses, and standardizes address data.
DqmBlueprintFrance_AddressCleanseGeo	Validates, cleanses, and standardizes address data and appends latitude and longitude coordinates.
DqmBlueprintFrance_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
DqmBlueprintFrance_CustomerValidation	Includes data quality validation rules to route records for remediation that fail the validation requirements.
DqmBlueprintFrance_LoadDelta	Cleanses delta data and performs matching against a database, such as a CRM, to identify new records to be inserted and matching records to be updated in the database.
DqmBlueprintFrance_LoadInitial	Performs an initial cleansing and removal of duplicates for a customer database, such as a CRM, storing the data in a way that provides for optimal matching in the future either with delta loads or with individual transactions.
DqmBlueprintFrance_MatchAssociative	Cleanses party data and performs duplicate detection based on similar name and address, name and phone, or name and email, combining the three matching results to uncover hidden duplicates.
DqmBlueprintFrance_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
DqmBlueprintFrance_MatchConsumer Household	Performs hierarchical matching to identify matching consumers within the same household.
DqmBlueprintFrance_MatchCorporate Household	Performs hierarchical matching to identify matching contacts within the same organization.
DqmBlueprintFrance_RtAddressData Cleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
DqmBlueprintFrance_RtAddressSuggestions	Validates, cleanses, and standardizes address data given a partial address on input and drilling down by selecting from a pick list of valid addresses.

Job	Description
DqmBlueprintFrance_RtGeocodePoi	Provides a list of points of interest (in this example, ATMs) within a specified proximity of an input address.
DqmBlueprintFrance_RtMatchConsumer	Cleanses a single transaction and performs matching against a database, such as a CRM, to identify whether the record exists in the database.

This blueprints package includes the following additional objects:

- A custom function that removes "noise" from firm data for optimal matching in order to identify accurate duplicate firms in France.
- A custom function that formats French phone numbers.
- A custom function that provides control of the prename for which the gender of a person is likely, defaulting to "M." and "Mme".
- A custom function that converts the information code from the address cleansing process to a description in English.

4.4 Data Quality Management Blueprints – Germany

Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in Germany.

This blueprints package includes the following jobs:

Job	Description
DqmBlueprintGermany_AddressCleanse	Validates, cleanses, and standardizes address data.
DqmBlueprintGermany_AddressCleanseGeo	Validates, cleanses, and standardizes address data, and appends latitude/longitude coordinates.
DqmBlueprintGermany_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
DqmBlueprintGermany_CustomerValidation	Includes data quality validation rules to route records for remediation that fail the validation requirements.
DqmBlueprintGermany_LoadDelta	Cleanses delta data and performs matching against a database, such as a CRM, to identify new records to be inserted and matching records to be updated in the database.

Job	Description
DqmBlueprintGermany_LoadInitial	Performs an initial cleansing and removal of duplicates for a customer database, such as a CRM, storing the data in a way that provides for optimal matching in the future either with delta loads or with individual transactions.
DqmBlueprintGermany_MatchAssociative	Cleanses party data and performs duplicate detection based on similar name and address, name and phone, or name and email, combining the three matching results to uncover hidden duplicates.
DqmBlueprintGermany_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
DqmBlueprintGermany_MatchConsumer Household	Performs hierarchical matching to identify matching consumers within the same household.
DqmBlueprintGermany_MatchCorporate Household	Performs hierarchical matching to identify matching contacts within the same organization.
DqmBlueprintGermany_RtAddressData Cleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
DqmBlueprintGermany_RtAddressSuggestions	Validates, cleanses, and standardizes address data given a partial address on input and drilling down by selecting from a pick list of valid addresses.
DqmBlueprintGermany_RtGeocodePoi	Provides a list of points of interest (in this example, ATMs) within a specified proximity of an input address.
DqmBlueprintGermany_RtMatchConsumer	Cleanses a single transaction and performs matching against a database, such as a CRM, to identify whether the record exists in the database.

This blueprints package includes the following additional objects:

- A custom function that removes "noise" from firm data for optimal matching in order to identify accurate duplicate firms in Germany.
- A custom function that formats German phone numbers.
- A custom function that provides control of the prename for which the gender of a person is likely, defaulting to "Hr." and "Fr."
- A custom function that converts the information code from the address cleansing process to a description in English.

4.5 Data Quality Management Blueprints – Global

Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data when the data consists of multiple countries.

This blueprints package includes the following jobs:

Job	Description
DqmBlueprintGlobal_AddressCleanse	Validates, cleanses, and standardizes address data.
DqmBlueprintGlobal_AddressCleanseGeo	Validates, cleanses, and standardizes address data, and appends latitude/longitude coordinates.
DqmBlueprintGlobal_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
DqmBlueprintGlobal_CustomerValidation	Includes data quality validation rules to route records for remediation that fail the validation requirements.
DqmBlueprintGlobal_LoadDelta	Cleanses delta data and performs matching against a database, such as a CRM, to identify new records to be inserted and matching records to be updated in the database.
DqmBlueprintGlobal_LoadInitial	Performs an initial cleansing and removal of duplicates for a customer database, such as a CRM, storing the data in a way that provides for optimal matching in the future either with delta loads or with individual transactions.
DqmBlueprintGlobal_MatchAssociative	Cleanses party data and performs duplicate detection based on similar name and address, name and phone, or name and email, combining the three matching results to uncover hidden duplicates.
DqmBlueprintGlobal_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
DqmBlueprintGlobal_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.
DqmBlueprintGlobal_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
DqmBlueprintGlobal_RtAddressData Cleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
DqmBlueprintGlobal_RtAddressSuggestions	Validates, cleanses, and standardizes address data given a partial address on input and drilling down by selecting from a pick list of valid addresses.

Job	Description
DqmBlueprintGlobal_RtGeocodePoi	Provides a list of points of interest (in this example, ATMs) within a specified proximity of an input address.
DqmBlueprintGlobal_RtMatchConsumer	Cleanses a single transaction and performs matching against a database, such as a CRM, to identify whether the record exists in the database.

This blueprints package includes the following additional objects:

- A custom function that removes "noise" from firm data for optimal matching in order to identify accurate duplicate global firms.

4.6 Data Quality Management Blueprints – India

Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in India.

This blueprints package includes the following jobs:

Job	Description
DqmBlueprintIndia_AddressCleanse	Validates, cleanses, and standardizes address data.
DqmBlueprintIndia_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
DqmBlueprintIndia_CustomerValidation	Includes data quality validation rules to route records for remediation that fail the validation requirements.
DqmBlueprintIndia_LoadDelta	Cleanses delta data and performs matching against a database, such as a CRM, to identify new records to be inserted and matching records to be updated in the database.
DqmBlueprintIndia_LoadInitial	Performs an initial cleansing and removal of duplicates for a customer database, such as a CRM, storing the data in a way that provides for optimal matching in the future either with delta loads or with individual transactions.
DqmBlueprintIndia_MatchAssociative	Cleanses party data and performs duplicate detection based on similar name and address, name and phone, or name and email, combining the three matching results to uncover hidden duplicates.
DqmBlueprintIndia_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.

Job	Description
DqmBlueprintIndia_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.
DqmBlueprintIndia_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
DqmBlueprintIndia_RtAddressDataCleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
DqmBlueprintIndia_RtAddressSuggestions	Validates, cleanses, and standardizes address data given a partial address on input and drilling down by selecting from a pick list of valid addresses.
DqmBlueprintIndia_RtMatchConsumer	Cleanses a single transaction and performs matching against a database, such as a CRM, to identify whether the record exists in the database.

This blueprints package includes the following additional objects:

- An India-specific Global Address Cleanse transform preconfigured with best practice settings and output fields to cleanse Indian address data.
- An India-specific Data Cleanse transform preconfigured with best practice settings and output fields to cleanse Indian person, title, firm, email, phone, and date data.
- Two custom functions that prepare address data for optimal matching in order to identify accurate duplicate addresses in India.
- A custom function that removes "noise" from firm data for optimal matching in order to identify accurate duplicate firms in India.
- A custom function that formats Indian phone numbers.
- A custom function that converts the information code from the address cleansing process to a description in English.

4.7 Data Quality Management Blueprints – Japan

Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in Japan.

This blueprints package includes the following jobs:

Job	Description
DqmBlueprintJapan_AddressCleanse	Validates, cleanses, and standardizes address data.
DqmBlueprintJapan_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.

Job	Description
DqmBlueprintJapan_CustomerValidation	Includes data quality validation rules to route records for remediation that fail the validation requirements.
DqmBlueprintJapan_LoadDelta	Cleanses delta data and performs matching against a database, such as a CRM, to identify new records to be inserted and matching records to be updated in the database.
DqmBlueprintJapan_LoadInitial	Performs an initial cleansing and removal of duplicates for a customer database, such as a CRM, storing the data in a way that provides for optimal matching in the future either with delta loads or with individual transactions.
DqmBlueprintJapan_MatchAssociative	Cleanses party data and performs duplicate detection based on similar name and address, name and phone, or name and email, combining the three matching results to uncover hidden duplicates.
DqmBlueprintJapan_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
DqmBlueprintJapan_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.
DqmBlueprintJapan_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
DqmBlueprintJapan_RtAddressData Cleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
DqmBlueprintJapan_RtMatchConsumer	Cleanses a single transaction and performs matching against a database, such as a CRM, to identify whether the record exists in the database.

This blueprints package includes the following additional objects:

- A custom function that removes "noise" from firm data for optimal matching in order to identify accurate duplicate firms in Japan.
- A custom function that formats Japanese phone numbers.
- A custom function that converts the information code from the address cleansing process to a description in Japanese.

4.8 Data Quality Management Blueprints – Mexico

Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in Mexico.

This blueprints package includes the following jobs:

Job	Description
DqmBlueprintMexico_AddressCleanse	Validates, cleanses, and standardizes address data.
DqmBlueprintMexico_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
DqmBlueprintMexico_CustomerValidation	Includes data quality validation rules to route records for remediation that fail the validation requirements.
DqmBlueprintMexico_LoadDelta	Cleanses delta data and performs matching against a database, such as a CRM, to identify new records to be inserted and matching records to be updated in the database.
DqmBlueprintMexico_LoadInitial	Performs an initial cleansing and removal of duplicates for a customer database, such as a CRM, storing the data in a way that provides for optimal matching in the future either with delta loads or with individual transactions.
DqmBlueprintMexico_MatchAssociative	Cleanses party data and performs duplicate detection based on similar name and address, name and phone, or name and email, combining the three matching results to uncover hidden duplicates.
DqmBlueprintMexico_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
DqmBlueprintMexico_MatchConsumer Household	Performs hierarchical matching to identify matching consumers within the same household.
DqmBlueprintMexico_MatchCorporate Household	Performs hierarchical matching to identify matching contacts within the same organization.
DqmBlueprintMexico_RtAddressData Cleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
DqmBlueprintMexico_RtAddressSuggestions	Validates, cleanses, and standardizes address data given a partial address on input and drilling down by selecting from a pick list of valid addresses.
DqmBlueprintMexico_RtMatchConsumer	Cleanses a single transaction and performs matching against a database, such as a CRM, to identify whether the record exists in the database.

This blueprints package includes the following additional objects:

- A Mexico-specific Global Address Cleanse transform preconfigured with best practice settings and output fields to cleanse Mexican address data.

- A custom function that removes "noise" from firm data for optimal matching in order to identify accurate duplicate firms in Mexico.
- A custom function that formats Mexican phone numbers.
- A custom function that provides control of the prename for which the gender of a person is likely, defaulting to "Sr." and "Srta."
- A custom function that converts the information code from the address cleansing process to a description in English.

4.9 Data Quality Management Blueprints – USA

Contains sample jobs configured to illustrate best practice settings for common Data Quality Management use cases involving party data in the United States.

This blueprints package includes the following jobs:

Job	Description
DqmBlueprintUSA_AddressCleanse	Validates, cleanses, and standardizes address data.
DqmBlueprintUSA_AddressCleanseGeo	Validates, cleanses, and standardizes address data, and appends latitude/longitude coordinates.
DqmBlueprintUSA_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
DqmBlueprintUSA_CustomerValidation	Includes data quality validation rules to route records for remediation that fail the validation requirements.
DqmBlueprintUSA_LoadDelta	Cleanses delta data and performs matching against a database, such as a CRM, to identify new records to be inserted and matching records to be updated in the database.
DqmBlueprintUSA_LoadInitial	Performs an initial cleansing and removal of duplicates for a customer database, such as a CRM, storing the data in a way that provides for optimal matching in the future either with delta loads or with individual transactions.
DqmBlueprintUSA_MatchAssociative	Cleanses party data and performs duplicate detection based on similar name and address, name and phone, or name and email, combining the three matching results to uncover hidden duplicates.
DqmBlueprintUSA_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
DqmBlueprintUSA_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.

Job	Description
DqmBlueprintUSA_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
DqmBlueprintUSA_RtAddressDataCleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
DqmBlueprintUSA_RtAddressSuggestions	Validates, cleanses, and standardizes address data given a partial address on input and drilling down by selecting from a pick list of valid addresses.
DqmBlueprintUSA_RtGeocodePoi	Provides a list of points of interest (in this example, ATMs) within a specified proximity of an input address.
DqmBlueprintUSA_RtMatchConsumer	Cleanses a single transaction and performs matching against a database, such as a CRM, to identify whether the record exists in the database.

This blueprints package includes the following additional object:

- A custom function that removes "noise" from firm data for optimal matching in order to identify accurate duplicate firms in the United States.

Data Quality Management Match blueprints

The following are the Match blueprints provided for common Data Quality Management use cases, and a list of the jobs and other objects that they contain.

For more information, see the *Match Blueprints User's Guide*.

5.1 Data Quality Management Blueprints – Match

Contains miscellaneous jobs configured to illustrate best practice settings for specific Data Quality Management matching use cases.

This blueprints package includes the following jobs:

Job	Description
DqmBlueprintMatch_AddrBlankIfFirm	Illustrates how to configure a data flow to use a specific match option that identifies two firm records as a match even though they have different addresses, in which one is a street address and the other is a postal address.
DqmBlueprintMatch_AddrIgnoreFirmIfName	Illustrates how to configure a data flow to use a specific match option that identifies two firm records as a match even though they have different firm names, as long as the two records have matching person name and address data.
DqmBlueprintMatch_AddrResidentRrNoBox	Illustrates how to configure a data flow to use a specific match option that considers multiple records with data such as "Current Resident, RR 2" as non-matching records.
DqmBlueprintMatch_AddrSteetAndRrOrBox	Illustrates how to configure a data flow to use a specific match option that identifies two records as a match, as long as either the street or rural addresses match or the postal addresses match.
DqmBlueprintMatch_MatchDNB	Inputs a data source and identifies matches to Dun & Bradstreet (DNB) data, enriching matching records with DUNS data.
DqmBlueprintMatch_MatchDNBPrep	Illustrates cleansing and preparing Dun & Bradstreet (DNB) data, outputting it to a database that is prepared for optimal matching.

Job	Description
DqmBlueprintMatch_SuppressDMA	Inputs a customer data list and identifies matches to Direct Marketing Association (DMA) data, suppressing matching records from the resulting mailing list.
DqmBlueprintMatch_SuppressDMAPrep	Illustrates cleansing and preparing Direct Marketing Association (DMA) data, outputting it to a database that is prepared for optimal matching.
DqmBlueprintMatch_SuspectsForReview	Inputs a data source and identifies matching records that have similar firm and address data, and routes suspect records to a table that may be reviewed for remediation through Match Review in Information Steward.
DqmBlueprintMatch_SuspectsForReview Prep	Creates a suspects table and a job status table, without inserting any data, in preparation for running the DqmBlueprintMatch_SuspectsForReview job. Note that this is simply a preliminary step to running the subsequent blueprint, since the best practice in production is not to create these tables with Data Services.

This blueprints package includes the following additional objects:

- Two custom functions that prepare address data for optimal matching in order to identify accurate duplicate addresses in a multi-country data source.
- A custom function that removes "noise" from firm data for optimal matching in order to identify accurate duplicate firms in a multi-country data source.
- A custom function that forms words commonly found in firm names in varying formats into a base word in order to identify accurate duplicate firms.
- A custom function that converts characters with diacritics into a character without diacritics in order to identify accurate duplicate firms.
- A custom function that removes special characters in order to identify accurate duplicate firms.
- A custom function that identifies the best content domain sequence to use for person and firm cleansing based on the country of the address.

Text Data Processing Language blueprints

The following are the blueprints provided for common Text Data Processing use cases, and a list of the jobs and other objects that they contain.

For more information, see the *Text Data Processing Language Blueprints User's Guide*.

6.1 Text Data Processing Blueprints – English

Contains sample jobs configured to illustrate best practice settings for common Text Data Processing use cases involving unstructured text in the English language.

This blueprints package includes the following jobs:

Job	Description
TdpBlueprintEn_Basic	Demonstrates the basic usage of the Text Data Processing Entity Extraction transform with all default options except the language option, which is set to English. The job processes a set of unstructured text files and extracts out-of-the-box entities from the documents.
TdpBlueprintEn_Binary	Demonstrates the extraction of entities from binary documents, such as Word and PDF documents, as well as binary document properties, such as title, author, and creation date.
TdpBlueprintEn_DictionaryGenerate	Demonstrates how to generate a Text Data Processing Entity Extraction dictionary from an Excel source file containing the dictionary entries. The job creates and compiles the dictionary.
TdpBlueprintEn_DictionaryUse	Demonstrates the usage of a dictionary in the Text Data Processing Entity Extraction transform. The job uses the dictionary generated by the TdpBlueprintEn_DictionaryGenerate job. It processes a set of unstructured text files and extracts the entities specified in the dictionary as well as other out-of-the-box entities from the documents.

Job	Description
TdpBlueprintEn_Rule	Demonstrates the usage of rules in the Text Data Processing Entity Extraction transform. The Entity Extraction transform is configured with the English enterprise rules with the Rule File option set to <code>.\TextAnalysis\languages\english-tf-ent-xxx.fsm</code> . The job also uses the dictionary generated by the TdpBlueprintEn_DictionaryGenerate job. It processes a set of unstructured text files and extracts the events recognized by the rules, and extracts the entities specified in the dictionary as well as other out-of-the-box entities from the documents.
TdpBlueprintEn_RtBasic	Demonstrates the usage of a Text Data Processing Entity Extraction transform in a simple real-time job.
TdpBlueprintEn_RtRule	Demonstrates the usage of a Text Data Processing Entity Extraction transform configured with enterprise rules in a real-time job that manipulates the output using a Query transform to link extraction output rows together.
TdpBlueprintEn_Twitter	Demonstrates the usage of a User-Defined transform using Python to retrieve unstructured text, such as tweets, from a public API and process it with the Text Data Processing Entity Extraction transform.

6.2 Text Data Processing Blueprints – German

Contains sample jobs configured to illustrate best practice settings for common Text Data Processing use cases involving unstructured text in the German language.

This blueprints package includes the following job:

Job	Description
TdpBlueprintDe_Basic	Demonstrates the basic usage of the Text Data Processing Entity Extraction transform with all default options except the language option, which is set to German. The job processes a set of unstructured text files and extracts out-of-the-box entities from the documents.

6.3 Text Data Processing Blueprints – Japanese

Contains a sample job configured to illustrate best practice settings for common Text Data Processing use cases involving multi-byte content in the Japanese language. It also showcases enriched support for core extraction pre-defined entity types.

This blueprints package includes the following job:

Job	Description
TdpBlueprintJa_Basic	Demonstrates the basic usage of the Text Data Processing Entity Extraction transform with all default options except the language option, which is set to Japanese. The job processes a set of unstructured text files and extracts out-of-the-box entities from the documents.

Text Data Processing Data Quality Management blueprints

The following is a blueprint provided to illustrate the use of Text Data Processing in conjunction with Data Quality Management.

For more information, see the *Text Data Processing Data Quality Management Blueprints User's Guide*.

7.1 Text Data Processing Blueprints – Data Quality Management

Contains sample jobs configured to illustrate the use of Text Data Processing in conjunction with Data Quality Management. It also helps you visualize the extracted concepts and sentiments using an SAP BusinessObjects BI 4.0 Universe and SAP BusinessObjects Web Intelligence reports.

Note:

The Text Data Processing Blueprints - Data Quality Management blueprint is packaged with an SAP BusinessObjects BI 4.0 Universe and SAP BusinessObjects Web Intelligence reports for visualizing the output of the TdpBlueprintDqm_VocMatch and TdpBlueprintDqm_VocMatchIQ jobs. To see the reports, you must have SAP BusinessObjects BI 4.0 installed.

This blueprints package includes the following jobs:

Job	Description
TdpBlueprintDqm_AddressDataCleanse	Demonstrates the use of the Text Data Processing Entity Extraction transform in conjunction with the Data Quality Management Address Cleanse and Data Cleanse transforms. The job uses the Entity Extraction transform to process a set of unstructured text files and extract person-firm information. The job then formats the person-firm data so that it can be passed as input to the Data Quality Management transforms, which cleanse any address, name, firm, email, and phone data found in the documents.

Job	Description
TdpBlueprintDqm_VocMatch	For Microsoft SQL Server. Demonstrates the use of the Text Data Processing Entity Extraction transform in conjunction with the Data Quality Management Match transform. The job uses the Entity Extraction transform to process a set of unstructured text files that contain feedback about a vehicle using the voice of the customer rules to extract any positive and negative sentiments. The job filters the output to separate sentences in the documents, sentiments organized by type, and topics about which the sentiments are expressed. The job writes output to a Microsoft SQL Server database. The Match transform groups similar topics together. You can use the SAP BusinessObjects BI 4.0 Universe and SAP BusinessObjects Web Intelligence reports packaged with the blueprint package to visualize and analyze the output data created by the job.
TdpBlueprintDqm_VocMatchIQ	For Sybase IQ. Demonstrates the use of the Text Data Processing Entity Extraction transform in conjunction with the Data Quality Management Match transform. The job uses the Entity Extraction transform to process a set of unstructured text files that contain feedback about a vehicle using the voice of the customer rules to extract any positive and negative sentiments. The job filters the output to separate sentences in the documents, sentiments organized by type, and topics about which the sentiments are expressed. The job writes output to a Sybase IQ database. The Match transform groups similar topics together. You can use the SAP BusinessObjects BI 4.0 Universe and SAP BusinessObjects Web Intelligence reports packaged with the blueprint package to visualize and analyze the output data created by the job.

7.1.1 Data flow transform descriptions

The following table describes each of the transforms in the order that they appear in the TdpBlueprintDqm_VocMatch and TdpBlueprintDqm_VocMatchIQ data flows.

Transform	Description
Base_EntityExtraction	<p>Generates sentiments, nouns (which represent sentiment topics), and individual sentences from each document in the TdpDqmln SampleData source. Sentiments are extracted by using the voice of the customer rules, which are shipped with Data Services in [\$\$SamplesInstall]\TextAnalysis\languages\english-tf-voc-request.fsm and \$\$SamplesInstall\TextAnalysis\languages\english-tf-voc-sentiment.fsm. Nouns are extracted using the [\$\$SamplesInstall]\Tutorial Files\Text Data Processing Samples\Data Quality\VocMatch\Rules\noun.fsm rule file that is provided in the blueprint. Sentences are extracted using the [\$\$SamplesInstall]\Tutorial Files\Text Data Processing Samples\Data Quality\VocMatch\Rules\sentence.fsm rule file, that is provided in this blueprint.</p>
AddFileID	<p>Adds a unique identifier to each input file that will be processed by the Base_EntityExtraction transform.</p>
AddUUID	<p>Adds a unique identifier to each row output by the Base_EntityExtraction transform. The identifier enables the SAP BusinessObjects Interactive Analysis Desktop reports to accurately aggregate counts.</p>
RouteEntity	<p>Separates the entities of interest based on type into three different paths within the data flow and discards all other entities. The Sentences path stores the SENTENCE entities in the TDP_BLUEPRINTS_DQM_VOCMATCH_SENTENCES template table. The Feedbacks path takes the StrongPositiveSentiment, WeakPositiveSentiment, StrongNegativeSentiment, WeakNegativeSentiment, MajorProblem, MinorProblem, ContactRequest, and GeneralRequest entities and adds a high-level feedback type description column, and then stores the data in the TDP_BLUEPRINTS_DQM_VOCMATCH_FEEDBACK template table. The description is added by the AddFeedbackTypeDesc transform. The description is used by the SAP BusinessObjects BI 4.0 Universe and SAP BusinessObjects Web Intelligence reports to group the FEEDBACK into three high-level groupings of Positive, Negative, and Request. The Topics path takes the NOUN entities and performs topic clustering using a combination of Match, Associate, and other transforms to produce a set of high quality top-level topics for use in the reports. The top-level topics are stored in the TDP_BLUEPRINTS_DQM_VOCMATCH_TOPIC_GROUPS template table, and all topics with reference to the top-level topics are stored in the TDP_BLUEPRINTS_DQM_VOCMATCH_TOPICS template table.</p>

Transform	Description
AddCustomSortingLength	Adds a new column to the schema that calculates a new length to be used for sorting rows during matching. Sorting on the LENGTH column alone produces topic groups that are either too coarse with poor recall or too many fine-grained topic groups, so the sorting length is designed to move all of the NOUNS that were shorter than ten characters to the bottom of the sort while allowing other NOUNS to be sorted by ascending length.
Topic_TightMatch	Performs matching on word similarity where the match score is high. It also utilizes the Approx substrig adjustment score option to find matches between NOUNS, such as "cupholders", "seat cupholders", and "passenger cupholders".
RouteTopic	Separates the topics that have been identified as subordinates in a group from the master and unique topics in the Topic_TightMatch transform. The subordinates are already high quality matches and using them in the subsequent match transforms instead undoes the work of the Topic_TightMatch transform when the AssociateTightAndMediumMatch transform is executed. The SubordinateTopics path adds in columns with default values to match the new columns that are added in the other path by the Topic_MediumMatch transform. This is done so that a merge of the two paths can be performed later. The UniqueOrMasterTopics path executes a looser match process, Topic_MediumMatch, to combine topics such as "car" and "cars".
Topic_MediumMatch	Introduces break groups based on the STANDARD_FORM column using the first two characters to prevent poor matches such as "interior" and "exterior" from occurring due to the lower Match score setting.
MergeTopic	Merges the two paths back together.
AssociateTightAndMediumMatch	Combines the data from the two different Match transforms into a composite set of match groups. The output is then stored in the TDP_BLUEPRINTS_DQM_VOCMATCH_TOPICS and TDP_BLUEPRINTS_DQM_VOCMATCH_TOPIC_GROUPS template tables. Before storing the output in the TDP_BLUEPRINTS_DQM_VOCMATCH_TOPIC_GROUPS template table, the FilterTopicGroups ensures that only the master and unique topic groups are stored. This separate table avoids the need to create a derived table in the Universe.

Other tools

The following are other tools provided for common SAP BusinessObjects Data Services use cases, and a list of objects that they contain.

8.1 Data Quality Management custom functions

Contains custom functions that perform additional manipulation of data that is not part of the functionality of Data Quality Management transforms, but are common functions that assist with the cleansing and matching of party data.

For more information, see the *Custom Functions User's Guide*.

This package includes the following custom functions:

Custom function	Description	Example
CF_AddressInfoCodeDescriptionEN	Generates a description in English for the Global Address Cleanse information code.	Converts "3010" to "Locality, region, and postcode are valid. Unable to match primary name to directory".
CF_AddressInfoCodeDescriptionCN	Generates a description in Chinese for the Global Address Cleanse information code.	Converts "3010" to "路名或者小区不存在于参照数据中".
CF_AddressInfoCodeDescriptionJP	Generates a description in Japanese for the Global Address Cleanse information code.	Converts "3010" to "入力された丁目、番地が照合結果、不合".
CF_AddressMatchPrepBuildingName	Removes characters to generate a building name field for the matching process.	Removes words such as "号" and "大厦" from Chinese building names.

Custom function	Description	Example
CF_AddressMatchPrepFloorUnit	Concatenates the floor and unit numbers to generate a secondary number field for the matching process.	For an address that includes Floor 4 and Unit 10, generates a field that contains "4 10".
CF_AddressMatchPrepNumberField	Removes characters to generate a primary or secondary number field for the matching process.	For a field that contains "5 - 7", generates a field that contains "5 7".
CF_AddressStatusCodeDescriptionEN	Generates a list of the address elements changed in the Data Quality Management process in English.	Converts "SC200" to "Data Quality corrected the following address components: region, locality, primary type".
CF_AddressStatusCodeDescriptionJP	Generates a list of the address elements changed in the Data Quality Management process in Japanese.	Converts "SC400" to "Data Quality が下記の住所構成要素を訂正：都道府県, 市区町村, 丁目・番地".
CF_ContentDomainSequence	Generates a dynamic Content Domain Sequence based on the country for the Data Cleanse PERSON_FIRM cleansing package.	Generates "EN_US GLOBAL" when the country is "US", generates "DE GLOBAL" when the country is "DE", generates "PT_BR GLOBAL" when the country is "BR", and so on.
CF_FirmFormBaseWordsEN	Converts words in English firm names to a base form to generate a field to form break groups for the matching process.	Converts variations such as Consultancy, Consultant, Consultants, Consultation, and Consulting to the base word "Consult".

Custom function	Description	Example
CF_FirmRemoveNoise	Uses the language-specific FirmRemoveNoise functions to remove noise words from multinational firm data to generate a firm field for the matching process.	Removes words such as "Inc." and "and" for countries where firm names are primarily in English, removes words such as "GmbH" and "und" for countries where firm names are primarily in German, removes words such as "S.A. de C.V." and "y" for countries where firm names are primarily in Spanish, removes words such as "株式会社" and "特例有限会社" for countries where firm names are primarily in Japanese, and so on.

Custom function	Description	Example
CF_FirmRemoveNoiseAR (Middle Eastern) CF_FirmRemoveNoiseBG (Bulgarian) CF_FirmRemoveNoiseCS (Czech) CF_FirmRemoveNoiseDA (Danish) CF_FirmRemoveNoiseDE (German) CF_FirmRemoveNoiseEL (Greek) CF_FirmRemoveNoiseEN (English) CF_FirmRemoveNoiseES (Spanish) CF_FirmRemoveNoiseET (Estonian) CF_FirmRemoveNoiseFI (Finnish) CF_FirmRemoveNoiseFR (French) CF_FirmRemoveNoiseHU (Hungarian) CF_FirmRemoveNoiseIT (Italian) CF_FirmRemoveNoiseJA (Japanese, both in Latin script and in Japanese scripts) CF_FirmRemoveNoiseMS (Malay) CF_FirmRemoveNoiseNL (Dutch) CF_FirmRemoveNoiseNO (Norwegian) CF_FirmRemoveNoisePL (Polish) CF_FirmRemoveNoisePT (Portuguese) CF_FirmRemoveNoiseRO (Romanian) CF_FirmRemoveNoiseSK (Slovak) CF_FirmRemoveNoiseSL (Slovenian) CF_FirmRemoveNoiseSR (Serbian) CF_FirmRemoveNoiseSV (Swedish) CF_FirmRemoveNoiseZH (Chinese)	Removes language-specific noise words from firm data to generate a firm field for the matching process.	

Custom function	Description	Example
CF_FirmStandardizeWordsEN	Performs data standardization to English firm words additional to Data Cleanse.	Converts a common misspelling such as "Internacional" to "International".
CF_GeoInfoCodeDescriptionEN	Generates a description in English for the Geocoder information code.	Converts "004" to "The input data is insufficient or incorrect to match the reference data".
CF_PersonInitCapPrefix	Capitalizes the first letter of each person name prefix.	Converts "van Helsing" to "Van Helsing", and converts "de los Angeles" to "De Los Angeles".
CF_PhoneFormatBR	Formats Brazilian phone numbers.	+55 (11) 3074-2404
CF_PhoneFormatCN	Formats Chinese phone numbers.	+86 (0)21 66861350, +86 18957205580
CF_PhoneFormatDE	Formats German phone numbers.	+49 (0)30 8959760 +49 (0)711 7317020
CF_PhoneFormatFR	Formats French phone numbers.	+33 (0)1 41 92 70 74
CF_PhoneFormatIN	Formats Indian phone numbers.	+91 (0)33 2283 4487 +91 (0)4347 233 465
CF_PhoneFormatJP	Formats Japanese phone numbers.	+81 3-5655-7650 +81 862-54-4877
CF_PhoneFormatMX	Formats Mexican phone numbers.	+52 (55) 56-69-03-70 +52 (614) 429-61-15
CF_PrenomeConvertDA	Generates the Danish prenames Hr. and Fr.	Hr. Børge Jensen Fr. Susanne Petersen
CF_PrenomeConvertDE	Generates the German prenames Hr. and Fr.	Hr. Hans Müller Fr. Anne Katrin Schmid

Custom function	Description	Example
CF_PrefixConvertES	Generates the Spanish prenames Sr. and Srta.	Sr. Juan Fernández Srta. Ana Luisa Torres
CF_PrefixConvertFR	Generates the French prenames M. and Mme.	M. Jean Claude Rousseau Mme Evelyne Breton
CF_PrefixConvertHU	Generates the Hungarian prenames Úr and Úrnő.	Nagy Benci Úr Szabó Annabella Julia Úrnő
CF_PrefixConvertIT	Generates the Italian prenames Sig. and Sig.ra.	Sig. Antonio Boscolo Sig.ra Giuseppina Francesca Romanò
CF_PrefixConvertNL	Generates the Dutch prenames dhr. and mevr.	dhr. Hans Budjhawan mevr. Petronella IJpenberg
CF_PrefixConvertPL	Generates the Polish prenames Pan and Pani.	Pan Piotr Kowalski Pani Maria Magdalena Wisniewska
CF_PrefixConvertPT	Generates the Portuguese prenames Sr. and Sra.	Sr. João Lopes Sra. Renata Macedo
CF_PrefixConvertSV	Generates the Swedish prenames hr. and fr.	hr. Erik Åström fr. Karin Lindberg
CF_RemoveDiacriticalCharacters	Converts letters with diacritical characters to the equivalent letter(s) without diacritics.	Converts "beschränkter" to "beschraenkte".
CF_RemoveProfanityEN	Removes English profanity words.	Converts "@&%!# Automaker Corp." to "Automaker Corp."

Custom function	Description	Example
CF_RemoveSpecialCharacters	Removes special characters from a string.	Converts "C-H-R Automaker Corp. (Group)" to "CHR Automaker Corp Group".
CF_RemoveSpecialCharactersSpace	Removes special characters from a string and leaves a space in their place.	Converts "C-H-R Automaker Corp. (Group)" to "C H R Automaker Corp Group".

8.2 Text Data Processing – Entity Extraction Dictionary File Generator

An Excel spreadsheet with a macro that generates a dictionary source file based on the content in the spreadsheet and compiles the source file into a ready-to-use dictionary file for the Text Data Processing Entity Extraction transform.

For more information, see the *Text Data Processing Entity Extraction Dictionary File Generator User's Guide*.

Glossary

Address Cleanse

Transforms that produce a correct and complete standardized form of an input address. The transform can also assign codes for postal automation and append other useful address information.

address line

A line of data in an address that contains the primary and, possibly, secondary address. The primary address contains components such as the primary range, primary name, directionals (post- and pre-), and the suffix. The secondary address normally contains components such as the unit designator and the secondary range.

association matching

A method of matching that combines the results of two or more Match transforms by using the Associate transform. Association matching is used to find duplicates based on multiple different match criteria (for example, based on Name+Address and then SSN+DOB) and bring them together.

A common use for association matching is to identify customers who have multiple residences. Examples of such customers could include students and snowbirds.

batch

Executes one job or a series of jobs all at one time. After batch processing begins, it continues until it is done or until an error occurs.

batch job

The unit of work that can be scheduled independently for execution by the Administrator. Jobs are special work flows that can be scheduled for execution, but cannot be called by other work flows or jobs.

blueprint

A sample Data Quality Management job that can be used by SAP BusinessObjects Data Services without modification. Each blueprint contains the necessary project, jobs, data flows, file formats, sample data, template tables, and custom functions to run the data flows

bulk loading

A software-based mechanism that moves large amounts of data into a database to achieve optimal performance. Bulk loading is faster than traditional INSERT statements. This mechanism supports compression, blocking, and buffering to optimize transfer times.

business rules

1. Settings within your Data Quality Management transforms that explain how you want to process your data. These include things like telling the Global Address Cleanse transform how to case output data, or setting up match criteria for a matching process.
2. Business rules can also be used to group validation rules from Validation transforms for display in the Data Validation reports in the Management Console.

case-sensitive

Pertaining to the differentiation between upper-case and lower-case letters. A case-sensitive program differentiates between upper-case and lower-case letters when evaluating a text string.

content object

Objects, including blueprints and custom functions, that you can import into SAP BusinessObjects Data Services to view as examples or modify and use for your business needs.

custom function

A script that you create to evaluate or make calculations on input values and produce a return value.

Data Cleanse

A transform that identifies and isolates specific parts of mixed data, and then standardizes the data based on information stored in the parsing dictionary, business rules defined in the rule file, and expressions defined in the pattern file.

data flow

A reusable object containing steps to define the transformation of data from source to target. Data flows are called from inside a work flow or job. You can pass information into or out of data flows using parameters.

data quality

A set of transforms that work together to improve the quality of your data by cleansing, enhancing, matching and consolidating data elements. The transforms include Address Cleanse, Data Cleanse, and Match.

data validation

Defining rules to which correct data should conform. In Data Services, you define these rules in the Validation transform. You can separate data that passes the validation rules from failed data.

datastore

A logical channel connecting Data Services to a source or target application. Different datastore types include database, application, web service, and adapters. The datastore definition typically includes the name and location of the database as well as user authentication information. Data Services uses a datastore definition to qualify a table name wherever a table is indicated in a diagram or expression. You can access the datastore definition through the object library.

delimited flat file

A data file in which each column value is separated by a delimiter, such as a comma, semicolon, tab, space, and so on. Each row starts a new line.

delimiter

Data Services has three types of delimiters: column, row, and text (character string). To separate columns, a delimiter can be a tab, semicolon, comma, space, or any character sequence. To separate rows of data, a delimiter can be a {new line} or any other character sequence. To denote the start and end of a character string, a delimiter can be single quotation marks ('), double quotation marks ("), or {none}.

delta load

Extracts only data that has changed since the last time a refresh cycle was performed.

Designer

A graphical user interface that allows you to design and test Data Services jobs.

diacritical character

A character that contains an accent, dieresis (umlaut), tilde, cedilla, or other distinguishing marks (for example, ä or Ç). You can choose to have standardized data with these types of characters. The application uses the Latin-1 code page for assigning these accents.

discrete field

Input or output data that has separate fields for each piece of information, such as addresses and names.

discrete format

Input source format in which pieces of data are parsed down to nearly the most distinct level. For example, a "first name" field would be discrete, whereas a "name" field that could contain first, middle, or last name information would not be discrete.

Entity Extraction

The process of discovering and presenting specific entities and facts that occur in unstructured text.

flat file

A flat file is a file containing records, generally one record per line. Fields may have a fixed width with padding, or be delimited by tabs, commas (CSV), or other characters. There are no structural relationships. The data is "flat" like a sheet of paper, rather than to more complex models such as a relational database.

function

A program that operates on values that are passed to it. Data Services functions are available through a function wizard in a script, conditional, or Query transform. Data Services also gives you access to functions provided by the DBMS you are using. In addition, you can define your own functions using the Data Services scripting language.

fuzzy match

Finding approximate matches to a pattern in a string.

Geocoder

A transform that identifies and appends geographic information, such as latitude and longitude, to address data.

household matching

Multiple levels of matching. An example of consumer householding involves identifying records with matching address data (residence level), and then within those identifying records with matching person data (consumer level). An example of corporate householding

involves identifying records with matching firm and address data (corporate level), and then within those identifying records with matching person data (contact level).

import

The process of acquiring information for the Data Services repository. Import the following kinds of information into Data Services:

- The metadata for source and target databases
- Descriptions and code for user-defined and DBMS functions and transforms
- ATL or XML files with definitions of Data Services objects that were previously exported out of a another Data Services repository.

job

The unit of work that can be scheduled independently for execution by the Administrator. Jobs are special work flows that can be scheduled for execution, but cannot be called by other work flows or jobs.

linguistic analysis

Natural-language processing (NLP) capabilities, such as segmentation, stemming, and tagging, among other things. Entity Extraction analyzes unstructured text, in multiple languages from any text data source, and automatically identifies and extracts key entity types, including people, dates, places, organizations, or other information, from the text.

Match

A transform that identifies duplicate records at multiple levels within a single pass for individuals, households, or corporations within multiple tables or databases and consolidates them into a single source.

matching record

A group of records found to be matches based on the criteria and business rules you choose. The records do not necessarily have the same data.

noise words

Words that exist in a firm name that distort matching results. A custom function can remove noise words on a record-by-record basis depending on its country of origin. For example, it is common for variations such "AT&T" and "AT and T" to be considered 100% similar. After removing noise words, the Match transform evaluates them both as "AT T" which makes them 100% similar. In Germany, GmbH is a noise word. After removing noise words, "BMW GmbH" and "BMW" match as 100% similar.

party data

Customer or consumer data that describes the individuals, groups of people, and legal entities that you do business with, including name, phone, email, and address.

real-time job

A group of objects (data flows, work flows, conditionals, scripts, and so forth) that execute on-demand as a "request-response" system. You design real-time jobs in the Designer, then configure them as real-time services and associate them with an Access Server in the Administrator, where they are started, managed and monitored. When a real-time service receives a request from a caller, it processes the request and returns a reply.

reference file

A file of address data used by Data Services to match, assign, standardize, and verify addresses. Reference files are also referred to as postal directories. These files have a .dir extension.

SAP BusinessObjects Data Insight

Software that allows you to monitor, analyze, and report on the quality of information contained in the data marts, data warehouses, ERP systems, packaged applications, and any other data stored in databases.

sentiment

A person's feelings about concepts, places, actions, items, and so on; for example, a product, company, service, or person. SAP BusinessObjects Data Services text data processing software includes rules that are designed to extract sentiments.

substitution parameter

A text string "alias" that you can use within your job and transforms. You define a substitution parameter and its value in a substitution parameter configuration. Then, at runtime, that parameter is replaced with its value anywhere it is used in your job.

suggestion lists

Normally, when an address cleansing transform looks up an address in the postal directories, it finds one matching record. Sometimes, due to incomplete information, there may be two or more records (or suggestions) in the postal directories that could possibly be the correct record. Suggestion lists provide you with a list of "matching" addresses, so that you can choose which is the best address.

text data processing

Software that lets you perform linguistic analysis of and extraction of content from unstructured text.

topic

In linguistic analysis, represents what a sentiment or request is about.

truth data query

Allows you to validate your address data by checking it against address data directories.

Universal Data Cleanse transform

A transform that lets you parse and manipulate operational and product data using dictionaries and rules that you create to meet your specific needs.

USA Regulatory Address Cleanse

An Address Cleanse transform that identifies, parses, validates, and corrects USA address data according to the U.S. Coding Accuracy Support System (CASS). This transform can create the USPS Form 3553 and output many useful codes to your records. You can also run in a non-certification mode as well as produce suggestion lists.

voice of the customer content

A set of rules in SAP BusinessObjects Data Services text data processing software that address requirements for extracting customer sentiments and requests. You can use this content to retrieve specific information about your customers' needs and perceptions when processing and analyzing text.

Index

A

about blueprints 5
available blueprints 11

B

blueprints
 about 5
 available 11
 Match 29
 regional 15
 Text Data Processing 31
 Text Data Processing Data Quality Management 35
 versions 9
Brazil Data Quality Management Blueprints 15

C

China Data Quality Management Blueprints 16
custom functions 39, 45

D

Data Quality Custom Functions 45
Data Quality Management blueprints list of 11
Data Quality Management Blueprints – Brazil 15
Data Quality Management Blueprints – China 16
Data Quality Management Blueprints – France 18
Data Quality Management Blueprints – Germany 19
Data Quality Management Blueprints – Global 21
Data Quality Management Blueprints – India 22
Data Quality Management Blueprints – Japan 23
Data Quality Management Blueprints – Match 29
Data Quality Management Blueprints – Mexico 24

Data Quality Management Blueprints – USA 26
Data Quality Management custom functions 39
Data Quality Management Text Data Processing Blueprints 35

E

English Text Data Processing Blueprints 31
Excel spreadsheet 39

F

France Data Quality Management Blueprints 18
functions, custom 39, 45

G

German Text Data Processing Blueprints 32
Germany Data Quality Management Blueprints 19
Global Data Quality Management Blueprints 21
glossary 47

I

India Data Quality Management Blueprints 22

J

Japan Data Quality Management Blueprints 23
Japanese Text Data Processing Blueprints 33

M

Match blueprints 29
Match Data Quality Management Blueprints 29

Mexico Data Quality Management Blueprints 24

R

regional blueprints 15

S

SAP BusinessObjects Data Services Blueprints versions 9
spreadsheet 39

T

TdpBlueprintDqm_VocMatch data flow 36
TdpBlueprintDqm_VocMatchIQ data flow 36
Text Analysis
 see Text Data Processing 31, 35
Text Data Processing blueprints 31
Text Data Processing Blueprints – Data Quality Management 35
Text Data Processing Blueprints – English 31
Text Data Processing Blueprints – German 32
Text Data Processing Blueprints – Japanese 33
Text Data Processing Data Quality Management blueprints 35
tools 39
transforms 36

U

USA Data Quality Management Blueprints 26

V

versions 9

