SAP BusinessObjects

SAP BusinessObjects Data Services Content Objects Summary

- SAP BusinessObjects Data Services

2010-03-24

SAP

Copyright

2010-03-24

Contents

Overview

We've identified a number of common data quality 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 from the SAP Community Network website all of the blueprint packages or only the blueprints and other content that you think you will find useful. 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 *How to Contribute* 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.

Available blueprints

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

To see the contents of each blueprint, including jobs and custom functions, see the *Content Objects Summary*. For instructions on downloading and setting up the blueprints, see the *Content Objects User's Guide.*

Blueprint	Description
Data Quality Blueprints - Brazil	Contains sample jobs configured to illustrate best practice settings for common Data Quality use cases involving party data in Brazil.
Data Quality Blueprints - France	Contains sample jobs configured to illustrate best practice settings for common Data Quality use cases involving party data in France.
Data Quality Blueprints - Germany	Contains sample jobs configured to illustrate best practice settings for common Data Quality use cases involving party data in Germany.
Data Quality Blueprints - India	Contains sample jobs configured to illustrate best practice settings for common Data Quality use cases involving party data in India.
Data Quality Blueprints - Japan	Contains sample jobs configured to illustrate best practice settings for common Data Quality use cases involving party data in Japan.
Data Quality Blueprints - Mexico	Contains sample jobs configured to illustrate best practice settings for common Data Quality use cases involving party data in Mexico.
Data Quality Blueprints - USA	Contains sample jobs configured to illustrate best practice settings for common Data Quality use cases involving party data in the United States.
Data Quality Blueprints - USA Regulatory	Contains sample jobs configured to illustrate best practice settings for common Data Quality use cases involving party data in the United States, with regulatory address certification enabled.

Blueprint	Description
Data Quality Blueprints - Match	Contains miscellaneous jobs configured to illustrate best practice settings for specific Data Quality matching use cases.
Data Quality Blueprints - UDC	Contains sample jobs configured to illustrate parsing and standardizing of product data using Universal Data Cleanse (UDC).
Data Quality - UDC Bulk Load File	Contains a job that reads data from a multi-sheet Excel spreadsheet and generates a bulk load XML file that is ready to be bulk loaded into a custom Universal Data Cleanse dictionary.
Data Quality Custom Functions	Contains custom functions that perform additional manipulation of data that is not part of the functionality of Data Quality transforms, but are common functions that assist with the cleansing and matching of party data.
Data Assessment Truth Data Queries	Contains a job used by SAP BusinessObjects Data Insight to perform truth data queries.

2.1 Data Quality Blueprints – Brazil

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

This blueprints package includes the following jobs:

Job	Description
Job_DqBatchBrazil_AddressCleanse	Validates, cleanses, and standardizes address data.
Job_DqBatchBrazil_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
Job_DqBatchBrazil_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_DqBatchBrazil_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.

Job	Description
Job_DqBatchBrazil_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.
Job_DqBatchBrazil_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
Job_DqBatchBrazil_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.
Job_DqBatchBrazil_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
Job_DqRealTimeBrazil_AddressDataCleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
Job_DqRealTimeBrazil_AddressSuggestions	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_DqRealTimeBrazil_MatchConsumer	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.

2.2 Data Quality Blueprints – France

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

This blueprints package includes the following jobs:

Job	Description
Job_DqBatchFrance_AddressCleanse	Validates, cleanses, and standardizes address data.
Job_DqBatchFrance_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
Job_DqBatchFrance_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_DqBatchFrance_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.
Job_DqBatchFrance_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.
Job_DqBatchFrance_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
Job_DqBatchFrance_MatchConsumer Household	Performs hierarchical matching to identify matching consumers within the same household.
Job_DqBatchFrance_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
Job_DqRealTimeFrance_AddressData Cleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
Job_DqRealTimeFrance_AddressSuggestions	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_DqRealTimeFrance_MatchConsumer	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.

2.3 Data Quality Blueprints – Germany

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

This blueprints package includes the following jobs:

Job	Description
Job_DqBatchGermany_AddressCleanse	Validates, cleanses, and standardizes address data.
Job_DqBatchGermany_AddressCleanseGeo	Validates, cleanses, and standardizes address data, and appends latitude/longitude coordinates.
Job_DqBatchGermany_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
Job_DqBatchGermany_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_DqBatchGermany_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.
Job_DqBatchGermany_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.
Job_DqBatchGermany_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
Job_DqBatchGermany_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.
Job_DqBatchGermany_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
Job_DqRealTimeGermany_AddressDataCleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
Job_DqRealTimeGermany_AddressSuggestions	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
Job_DqRealTimeGermany_MatchConsumer	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.

2.4 Data Quality Blueprints – India

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

This blueprints package includes the following jobs:

Job	Description
Job_DqBatchIndia_AddressCleanse	Validates, cleanses, and standardizes address data.
Job_DqBatchIndia_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
Job_DqBatchIndia_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_DqBatchIndia_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.
Job_DqBatchIndia_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.
Job_DqBatchIndia_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.

Job	Description
Job_DqBatchIndia_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.
Job_DqBatchIndia_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
Job_DqRealTimeIndia_AddressDataCleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
Job_DqRealTimeIndia_AddressSuggestions	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_DqRealTimeIndia_MatchConsumer	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.

2.5 Data Quality Blueprints – Japan

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

This blueprints package includes the following jobs:

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

Job	Description
Job_DqBatchJapan_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_DqBatchJapan_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.
Job_DqBatchJapan_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.
Job_DqBatchJapan_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
Job_DqBatchJapan_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.
Job_DqBatchJapan_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
Job_DqRealTimeJapan_AddressDataCleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
Job_DqRealTimeJapan_MatchConsumer	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.

2.6 Data Quality Blueprints – Mexico

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

This blueprints package includes the following jobs:

Job	Description
Job_DqBatchMexico_AddressCleanse	Validates, cleanses, and standardizes address data.
Job_DqBatchMexico_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
Job_DqBatchMexico_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_DqBatchMexico_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.
Job_DqBatchMexico_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.
Job_DqBatchMexico_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
Job_DqBatchMexico_MatchConsumer Household	Performs hierarchical matching to identify matching consumers within the same household.
Job_DqBatchMexico_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
Job_DqRealTimeMexico_AddressData Cleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
Job_DqRealTimeMexico_AddressSuggestions	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_DqRealTimeMexico_MatchConsumer	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.

2.7 Data Quality Blueprints – USA

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

This blueprints package includes the following jobs:

Job	Description
Job_DqBatchUSA_AddressCleanse	Validates, cleanses, and standardizes address data.
Job_DqBatchUSA_AddressCleanseGeo	Validates, cleanses, and standardizes address data, and appends latitude/longitude coordinates.
Job_DqBatchUSA_AddressDataCleanse	Cleanses address, name, title, firm, email, phone, and date data.
Job_DqBatchUSA_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_DqBatchUSA_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.
Job_DqBatchUSA_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.
Job_DqBatchUSA_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.
Job_DqBatchUSA_MatchConsumerHousehold	Performs hierarchical matching to identify matching consumers within the same household.
Job_DqBatchUSA_MatchCorporateHousehold	Performs hierarchical matching to identify matching contacts within the same organization.
Job_DqRealTimeUSA_AddressDataCleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
Job_DqRealTimeUSA_AddressSuggestions	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
Job_DqRealTimeUSA_MatchConsumer	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.

2.8 Data Quality Blueprints – USA Regulatory

Contains sample jobs configured to illustrate best practice settings for common Data Quality use cases involving party data in the United States, with regulatory address certification enabled.

This blueprints package includes the following jobs:

Job	Description
Job_DqBatchUSAReg_AddressCleanse	Validates, cleanses, and standardizes address data.
Job_DqBatchUSAReg_AddressCleanseGeo	Validates, cleanses, and standardizes address data, and appends latitude and longitude coordinates.
Job_DqBatchUSAReg_AddressData Cleanse	Cleanses address, name, title, firm, email, phone, and date data.
Job_DqBatchUSAReg_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_DqBatchUSAReg_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.
Job_DqBatchUSAReg_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.
Job_DqBatchUSAReg_MatchConsumer	Cleanses consumer data and performs duplicate detection based on similar name and address using fuzzy matching techniques.

Job	Description
Job_DqBatchUSAReg_MatchConsumer Household	Performs hierarchical matching to identify matching consumers within the same household.
Job_DqBatchUSAReg_MatchCorporate Household	Performs hierarchical matching to identify matching contacts within the same organization.
Job_DqRealTimeUSAReg_AddressData Cleanse	Cleanses a single transaction that contains address, name, title, firm, email, phone, and date data.
Job_DqRealTimeUSAReg_MatchConsumer	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.

2.9 Data Quality Blueprints – Match

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

This blueprints package includes the following jobs:

Job	Description
Job_DqBatchMatch_MatchDNB	Inputs a data source and identifies matches to Dun & Bradstreet (DNB) data, enriching matching records with DUNS data.
Job_DqBatchMatch_MatchDNBPrep	Illustrates cleansing and preparing Dun & Bradstreet (DNB) data, outputting it to a database that is prepared for optimal matching.
Job_DqBatchMatch_SuppressDMA	Inputs a customer data list and identifies matches to Direct Marketing Association (DMA) data, suppressing matching records from the resulting mailing list.
Job_DqBatchMatch_SuppressDMAPrep	Illustrates cleansing and preparing Direct Marketing Association (DMA) data, outputting it to a database that is prepared for optimal matching.

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.

2.10 Data Quality Blueprints – UDC

Contains sample jobs configured to illustrate parsing and standardizing of product data using Universal Data Cleanse.

This blueprints package includes the following jobs:

Job	Description	Example
Job_DqBatchUDC_FinancialProductsEN	Parses and standardizes financial product data given a product description.	**Input data:** 5.40% 3 MONTH $50K CD RIVER CITY COMMUNITY CREDIT UNION **Output data:** Financial Institution: River City Community Credit Union Financial Term: 3-month Financial Product: CD Yield: 5.40% Minimum Deposit: $50,000

Job	Description	Example
Job_DqBatchUDC_GloveProductsEN	Parses and standardizes glove product data in two different sources and performs duplicate detection between them.	**Input data:** GLOVE THRML-LND GRIP LADIES ACRYLIC LATEX COATED SMALL KNIT WRIST KINCO INTL INC **Output data:** UNSPSC: 46181504 Vendor: KINCO Product Category: GLOVE Size: SMALL Type: GENERAL UTILITY Class: WOMENS Material: ACRYLIC Color: BLACK Palm Type: LATEX Cuff: KNIT WRIST Lining: LINED
Job_DqBatchUDC_PaperProductsEN	Parses and standardizes paper product data given a product description (English data).	**Input data:** SPR00812 Copy Paper, 92 Bright, 20 Lb, 11 x 17, 500/PK, White **Output data:** Vendor: Smart Papers Description: Copy Paper Size: 11"x17" Color: White Quantity: 500/pack Brightness: 92 Br. Weight: 20# Item Number: SPR00812

Job	Description	Example
Job_DqBatchUDC_PaperProductsFR	Parses and standardizes paper product data given a product description (French data).	**Input data:** Ramette de 500 f. A4 80g REY papier d'imprimeur (Carton de 5) bleu **Output data:** Couleur: Bleu Format: A4 Carton: Carton de 5 Ream Quantity: 500 f./ramette Description: papier d'imprimeur Grammage: 80 g/m² Fabricant: REY
Job_Dq BatchUDC_PharmaProductsEN	Parses and standardizes pharmaceutical product data given a product description.	**Input data:** CLOXACILLIN SODIUM POWDER FOR SUSPENSION ORAL -- 250 MG PER 5 ML **Output data:** Drug Name: Cloxacillin Sodium Potency: 250 MG PER 5 ML Dosage: PWD F/SUSP Route Administered: ORAL

2.11 Data Quality Blueprints – UDC Bulk Load File

Contains a job that reads data from a multi-sheet Excel spreadsheet and generates a bulk load XML file that is ready to be bulk loaded into a custom Universal Data Cleanse dictionary.

This blueprints package includes the following job:

Job	Description
Job_DqBatchUDC_GenerateBulkLoadFile	Reads data from a multi-sheet Excel spreadsheet and generates a bulk load XML file that is ready to be bulk loaded into a custom Universal Data Cleanse dictionary.

2.12 Data Quality Custom Functions

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

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_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_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 process in English.	Converts "SC200" to "Data Quality corrected the following address components: region, locality, primary type".

Custom function	Description	Example
CF_AddressStatusCodeDescriptionJP	Generates a list of the address elements changed in the Data Quality 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_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".
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_JapaneseHiraganaToKatakana	Converts Hiragana to Katakana.	Converts "さとう" to "サトウ".
CF_JapaneseHiraganaToRomaji	Converts Hiragana to Romaji, using the Hepburn system of Romanization.	Converts "さとう" to "ｓａｔｏｕ".
CF_JapaneseKatakanaToHiragana	Converts Katakana to Hiragana.	Converts "サトウ" to "さとう".
CF_JapaneseKatakanaToRomaji	Converts Katakana to Romaji, using the Hepburn system of Romanization.	Converts "サトウ" to "ｓａｔｏｕ".
CF_PhoneFormatBR	Formats Brazilian phone numbers.	+55 (11) 3074-2404
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_PrenameConvertDA	Generates the Danish prenames Hr. and Fr.	Hr. Børge Jensen Fr. Susanne Petersen

Custom function	Description	Example
CF_PrenameConvertDE	Generates the German prenames Hr. and Fr.	Hr. Hans Müller Fr. Anne Katrin Schmid
CF_PrenameConvertES	Generates the Spanish prenames Sr. and Srta.	Sr. Juan Fernández Srta. Ana Luisa Torres
CF_PrenameConvertFR	Generates the French prenames M. and Mme.	M. Jean Claude Rousseau Mme Evelyne Breton
CF_PrenameConvertHU	Generates the Hungarian prenames Úr and Úrnő.	Nagy Benci Úr Szabó Annabella Julia Úrnő
CF_PrenameConvertIT	Generates the Italian prenames Sig. and Sig.ra.	Sig. Antonio Boscolo Sig.ra Giuseppina Francesca Romanò
CF_PrenameConvertNL	Generates the Dutch prenames dhr. and mevr.	dhr. Hans Budjhawan mevr. Petronella IJpenberg
CF_PrenameConvertPL	Generates the Polish prenames Pan and Pani.	Pan Piotr Kowalski Pani Maria Magdalena Wisniewska
CF_PrenameConvertPT	Generates the Portuguese prenames Sr. and Sra.	Sr. João Lopes Sra. Renata Macedo
CF_PrenameConvertSV	Generates the Swedish prenames hr and fr .	hr Erik Åström fr. Karin Lindberg
CF_RemoveDiacriticalCharacters	Converts characters with diacritical characters to the closest equivalent character without diacritics.	Converts "beschränkter" to "beschrankter".

Custom function	Description	Example
CF_RemoveProfanityEN	Removes English profanity words.	Converts "@&%!# Automaker Corp." to "Automaker Corp."
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".

2.13 Data Assessment Truth Data Queries

SAP BusinessObjects Data Insight offers a feature by which you can perform truth data queries that allow you to verify address data by checking them against address data directories. In SAP BusinessObjects Data Services, you run a workflow that verifies the address data from the truth data queries that you created in SAP BusinessObjects Data Insight.

This blueprints package includes the following job:

Job	Description
jb_ data_assessment_truth_data_global_address_DataServices	Used by SAP BusinessObjects Data Insight to perform truth data queries.

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 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 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.

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.

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.

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.

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 operationaland 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.

Index