



Content Objects What's New

- 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	Introduction.....	5
1.1	Documentation set for SAP BusinessObjects Data Services content objects.....	5
1.2	SAP BusinessObjects information resources.....	5
1.3	Introduction to SAP BusinessObjects Data Services 4.1 SP1 Content Objects.....	6
Chapter 2	SAP BusinessObjects Data Services 4.1 SP1 features.....	9
2.1	Data Quality Management regional blueprints.....	9
2.2	Data Quality Management Match blueprints.....	10
2.3	Data Quality Management custom functions.....	11
2.4	Text Data Processing Language blueprints.....	12
2.5	Text Data Processing Data Quality Management blueprints.....	13
Index		15

Introduction

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.
<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
Customer Support, Consulting, and Education services http://service.sap.com/	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.
Product documentation http://help.sap.com/bods/	SAP BusinessObjects product documentation.
Supported Platforms (Product Availability Matrix) https://service.sap.com/PAM	Get information about supported platforms for SAP BusinessObjects Data Services. Use the search function to search for Data Services. Click the link for the version of Data Services you are searching for.

1.3 Introduction to SAP BusinessObjects Data Services 4.1 SP1 Content Objects

Welcome to SAP BusinessObjects Data Services 4.1 version 14.1.1 Content Objects. This *What's New* document highlights the new blueprints and other content objects available for this release.

Data Services overview

SAP BusinessObjects Data Services delivers a single enterprise-class solution for data integration, data quality, data profiling, and text data processing that allows you to integrate, transform, improve, and deliver trusted data to critical business processes. It provides one development UI, metadata repository, data connectivity layer, run-time environment, and management console—enabling IT organizations to lower total cost of ownership and accelerate time to value. With SAP BusinessObjects Data Services, IT organizations can maximize operational efficiency with a single solution to improve data quality and gain access to heterogeneous sources and applications.

Data Services Content Objects 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 *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.

SAP BusinessObjects Data Services 4.1 SP1 features

2.1 Data Quality Management regional blueprints

The following are new and enhanced regional blueprints provided for this release of Data Services.

For more information, see the *SAP BusinessObjects Data Services Regional Blueprints User's Guide*.

New China blueprints package

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

Enhanced India blueprints package

The Data Services Global Address Cleanse transform has enhanced the capabilities of cleansing India addresses, so the jobs in the India blueprints package now take advantage of this by including the best practice fields for India addresses. The matching process has also been updated based on the enhancements made to address cleansing.

Customer validation

Each regional blueprints package has a new job that includes data quality validation rules to route records for remediation that fail the validation requirements.

New global SAP-supplied person and firm cleanse package

A new global SAP-supplied person and firm cleansing package, which consolidates the variations that previously existed in over 20 separate regional cleansing packages, is now provided for the Cleansing Package Builder.

The multinational jobs in the Data Services blueprints have been updated to use a single Data Cleanse transform instead of one transform for each language.

Removed USA Regulatory blueprints

This version of the blueprints does not include a USA Regulatory blueprints package, because certification requirements require certain reference files that have stringent rules that prevent users from using the blueprints.

2.2 Data Quality Management Match blueprints

The following are new and enhanced Match blueprints provided for this release of Data Services.

For more information, see the *SAP BusinessObjects Data Services Match Blueprints User's Guide*.

New matching jobs

The Match blueprints package includes the following new 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_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.

2.3 Data Quality Management custom functions

The following are new and enhanced custom functions provided for this release of Data Services.

For more information, see the *SAP BusinessObjects Data Services Custom Functions User's Guide*.

New custom functions

The following custom functions have been added:

Custom function	Description	Example
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_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.
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_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_PhoneFormatCN	Formats Chinese phone numbers.	+86 (0)21 66861350, +86 18957205580

Enhanced custom function

The following custom function has been enhanced:

Custom function	Description	Example
CF_RemoveDiacriticalCharacters	Converts letters with diacritical characters to the equivalent letter(s) without diacritics.	Converts "beschränkter" to "beschraenkter".

The following table describes how the affected diacritical characters are converted. All other characters are converted the same as in the previous release.

Previous conversion	Current conversion
ä to a	ä to ae
Ä to A	Ä to AE
ö to o	ö to oe
Ö to O	Ö to OE
U to U	U to UE
ü to u	ü to ue

2.4 Text Data Processing Language blueprints

The following are new and enhanced Text Data Processing blueprints provided for this release of Data Services.

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

New English jobs

The Text Data Processing Language – English blueprints package includes the following new jobs:

Job	Description
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_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.

New Japanese blueprints package

The new Text Data Processing Language – Japanese blueprints package contains a sample job configured to show how multi-byte content can be processed with the Text Data Processing Entity Extraction transform. It also showcases enriched support for core extraction pre-defined entity types.

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.

2.5 Text Data Processing Data Quality Management blueprints

The following is the enhanced Text Data Processing Data Quality Management blueprint provided for this release of Data Services.

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

Enhanced TdpBlueprintDqm_AddressDataCleanse job

The TdpBlueprintDqm_AddressDataCleanse job has been enhanced to use the new English language entity type ADDRESS2, which provides better input from the Entity Extraction transform and results in improved address cleansing output.

The Data Cleanse transform has also been modified to use the PERSON_FIRM global cleansing package.

Index

B

blueprints
 Match 10
 regional 9
 Text Data Processing 12
 Text Data Processing Data Quality
 Management 13

C

custom functions 11

F

functions, custom 11

M

Match
 blueprints 10

R

regional
 blueprints 9

T

Text Analysis
 see Text Data Processing 12, 13
Text Data Processing
 blueprints 12
Text Data Processing Data Quality
 Management
 blueprints 13

