

SAP Data Services
Document Version: 4.2 SP2 (14.2.2) - 2014-05-21

Content Objects What's New

Table of Contents

1	Introduction	3
1.1	About this guide.	3
1.2	Documentation set for SAP Data Services content objects.	3
1.3	SAP information resources.	4
1.4	Introduction to SAP Data Services 4.2 Content Objects.	4
2	SAP Data Services 4.2 SP2 features	6
2.1	Data Quality Management regional blueprints.	6
3	SAP Data Services 4.2 features	7
3.1	Data Quality Management regional blueprints.	7
3.2	Data Quality Management Match blueprints.	8
3.3	Data Quality Management product blueprints.	9
3.4	Data Quality Management custom functions.	10
3.5	Text Data Processing language blueprints.	10
3.6	Text Data Processing miscellaneous blueprints.	11

1 Introduction

1.1 About this guide

This *What's New* document highlights the new and enhanced blueprints and other content objects available for this release.

1.2 Documentation set for SAP Data Services content objects

You should become familiar with all of the pieces of documentation that relate to the SAP 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 Product Blueprints User's Guide</i>	Contains a list of available Data Quality Management product 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 Entity Extraction Dictionary File Generator User's Guide</i>	Contains instructions for installing and using the Excel spreadsheet to generate and compile dictionary XML files used by the Entity Extraction transform.
<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.
<i>Text Data Processing Miscellaneous Blueprints User's Guide</i>	Contains a list of available Text Data Processing Miscellaneous blueprints and instructions for downloading, configuring, and running them.

1.3 SAP information resources

A global network of SAP 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 support programs, as well as links to technical articles, downloads, and online forums. Consulting services can provide you with information about how SAP 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 can offer a training package to suit your learning needs and preferred learning style.
Product documentation http://help.sap.com/bods/	SAP product documentation.
Supported Platforms (Product Availability Matrix) https://service.sap.com/PAM	Get information about supported platforms for SAP Data Services. Use the search function to search for Data Services. Click the link for the version of Data Services you are searching for.
SAP Data Services Community Network http://scn.sap.com/community/data-services	Get online and timely information about SAP Data Services, including forums, tips and tricks, additional downloads, samples, and much more. All content is to and from the community, so feel free to join in and contact us if you have a submission.
Blueprints http://scn.sap.com/docs/DOC-8820	Blueprints for you to download and modify to fit your needs. Each blueprint contains the necessary SAP Data Services 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.

1.4 Introduction to SAP Data Services 4.2 Content Objects

Welcome to SAP Data Services 4.2 version 14.2.0 Content Objects.


Data Services overview

SAP 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 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 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 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.

2 SAP Data Services 4.2 SP2 features

2.1 Data Quality Management regional blueprints

The following is an enhancement to the regional blueprints provided for this release of Data Services.

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

Sample Global Address Cleanse transform for South Korea

A sample Global Address Cleanse transform configuration that is configured with best practice settings for cleansing address data in South Korea is available. This transform configuration can only be used with SAP Data Services version 4.2 SP2 or later.

3 SAP Data Services 4.2 features

3.1 Data Quality Management regional blueprints

The following are enhancements to the regional blueprints provided for this release of Data Services.

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

New status codes and information codes generated by Data Cleanse

The AddressDataCleanse blueprint in each regional blueprints package has been updated to output two new codes, and the CustomerValidation blueprint in each regional blueprints package has been updated to base validation rules to identify suspect data on the information code.

Single phone parser in Data Cleanse

All blueprints that cleanse phone data now use a single phone parser instead of a separate North American phone parser and international phone parser.

New option in Data Cleanse to remove diacritical characters

The AddressDataCleanse and CustomerValidation blueprints in the USA blueprints package enables the Remove Diacritical Characters option in the Data Cleanse transform, and no longer needs to use the CF_RemoveDiacriticalCharacters custom function.

New dynamic options in Data Cleanse

The AddressDataCleanse blueprint in the global blueprints package maps fields generated by the address cleansing process to Option_Country, Option_Region, and Option_Language in the Data Cleanse transform, in order for the transform to automatically determine the content domain sequence and output format based on data in the record.

Data Cleanse reports

The Data Cleanse transform now generates reports. The blueprints use a substitution parameter for the option in Data Cleanse just like in Global Address Cleanse, Match, and Geocoder. To view the reports, in the substitution parameter configuration, set `$$ReportsDataCleanse` to Yes.

New match fields

All blueprints that include matching now use the match-specific fields generated by the Global Address Cleanse and Data Cleanse transforms. This eliminates the need to include custom functions before the Match transform to further prepare data for the matching process, because the data in the match-specific fields is already optimized for matching.

Enhanced India blueprints package

The Global Address Cleanse transform has further enhanced the capabilities of cleansing India addresses. The matching process has been modified slightly based on the enhancements made to address cleansing.

3.2 Data Quality Management Match blueprints

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

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

New matching blueprints

The SuspectForReview blueprint has been replaced with a set of blueprints that illustrate how to use SAP Data Services and SAP Information Steward Match Review together to complete the cycle of identifying matching records, manually reviewing suspect match groups, and creating best records.

- **MatchReview1:** This set of blueprints illustrates creating the best record in Data Services after the suspect match groups are reviewed in Information Steward Match Review.
- **MatchReview2:** This set of blueprints illustrates creating the best record automatically on high-confidence match groups in Data Services and manually creating the best record on suspect match groups in Information Steward Match Review in conjunction with reviewing the suspect match groups.

The Match blueprints package includes the following new jobs:

Job	Description
DqmBlueprintMatch_MatchReview1After	After review and remediation of suspect match groups formed in the MatchReview1Before job in Information Steward Match Review, this job creates best records by performing data consolidation within match groups.
DqmBlueprintMatch_MatchReview1Before	Identifies matching records that have similar firm and address data, and routes results to a staging table for the purpose of review and remediation of the suspect matches through SAP Information Steward Match Review.
DqmBlueprintMatch_MatchReview2After	Joins the staging table that contains the best records and non-matching records from the MatchReview2Before job, and the staging table that contains the results of the review and remediation, along with best record creation, through SAP Information Steward Match Review.
DqmBlueprintMatch_MatchReview2Before	Identifies matching records that have similar firm and address data, and separates high-confidence match groups and suspect match groups. High-confidence match groups are routed to receive best record creation by performing data consolidation within match groups, and the best records together with non-matching records are stored in a staging table. Suspect match groups are routed to a staging table for the purpose of review and remediation, along with best record creation, through SAP Information Steward Match Review.
DqmBlueprintMatch_MatchReviewPrep	Creates staging tables and a job status table, without inserting any data, in preparation for running the DqmBlueprint-Match_MatchReview1Before and BlueprintMatch_MatchReview2Before jobs. Note that this is simply a preliminary step to running the subsequent blueprints, since the best practice in production is not to create these tables with SAP Data Services.

3.3 Data Quality Management product blueprints

The following are new product blueprints provided for this release of Data Services.

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

New product blueprints package

There is a new product blueprints package that includes five jobs that cleanse data in different industries. Each comes with a custom cleansing package that can be imported into SAP Information Steward Cleansing Package Builder.

The blueprints package includes the following jobs:

Job	Description
DqmBlueprintProduct_Financial	Cleanses financial product data using a custom cleansing package.
DqmBlueprintProduct_Glove	Cleanses glove product data using a custom cleansing package, and performs duplicate detection based on the parsed product components.
DqmBlueprintProduct_Paper	Cleanses paper product data using a custom cleansing package.
DqmBlueprintProduct_Papier	Cleanses French paper product data using a custom cleansing package.
DqmBlueprintProduct_Pharma	Cleanses pharmaceutical product data using a custom cleansing package.

3.4 Data Quality Management custom functions

The following are enhancements to the custom functions provided for this release of Data Services.

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

Enhanced custom functions

The following custom functions have been enhanced:

- In the CF_AddressInfoCodeDescriptionCN, CF_AddressInfoCodeDescriptionEN, and CF_AddressInfoCodeDescriptionJP custom functions, we added descriptions to information codes 3210, 3300, and 5040.
- In the CF_AddressStatusCodeDescriptionCN, CF_AddressStatusCodeDescriptionEN, and CF_AddressStatusCodeDescriptionJP custom functions, we added descriptions to the fifth digit of the status code.

For more information about information codes and status codes, see the *SAP Data Services Reference Guide*.

3.5 Text Data Processing language blueprints

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

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

New Simplified Chinese jobs

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

Job	Description
TdpBlueprintZh_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 Simplified Chinese. The job processes a set of unstructured text files and extracts out-of-the-box entities from a set of unstructured text files.
TdpBlueprintZh_VOC	Demonstrates the usage of the Text Data Processing Entity Extraction transform configured with Simplified Chinese Voice of the Customer rules to extract sentiments from a set of unstructured text files.

3.6 Text Data Processing miscellaneous blueprints

The following is a new Text Data Processing miscellaneous blueprint provided for this release of Data Services. For more information, see the *SAP Data Services Text Data Processing Miscellaneous Blueprints User's Guide*.

New miscellaneous jobs

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

Job	Description
TdpBlueprintMisc_LanguageIdentification	Demonstrates usage of the Text Data Processing Entity Extraction transform configured to automatically identify input text language and perform extraction in that language. The job processes a set of unstructured text files in different languages and extracts out-of-the-box entities from them.
TdpBlueprintMisc_LinguisticAnalysis	Demonstrates usage of the Text Data Processing Entity Extraction transform configured with a <code>Para-Sent-Clause-NP-POS.fsm</code> rule file that helps you understand how your input text is parsed and passed to a custom rule that you write. It shows you how paragraphs, sentences, clauses, noun phrases, and individual tokens, as well as their parts of speech, are extracted.



www.sap.com/contactsap

© 2014 SAP AG or an SAP affiliate company. All rights reserved.
No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. The information contained herein may be changed without prior notice.

Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors. National product specifications may vary. 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.

SAP 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.

Please see <http://www.sap.com/corporate-en/legal/copyright/index.epx> for additional trademark information and notices.