



What's New

■ SAP BusinessObjects Data Services 4.1 Support Package 1 (14.1.1.0)

2012-12-12



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, Business Objects, 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.

2012-12-12

Contents

Chapter 1	Introduction	5
1.1	Introduction to SAP BusinessObjects Data Services 4.1 Support Package 1	5
1.2	SAP BusinessObjects information resources	6
Chapter 2	SAP BusinessObjects Data Services 4.1 features	7
2.1	Architecture	7
2.2	Usability	7
2.3	Enhanced SAP HANA support	9
2.4	Enhanced extraction capabilities for SAP Business Suite	
2.5	Sybase IQ loading performance enhancements	11
2.6	Connectivity: Big data and CDC	
2.7	Monitoring, administration, and installation	13
2.8	Functions and transforms	14
2.9	Text Data Processing	15
2.10	Data Quality	
2.10.1	Data Cleanse transform	17
2.10.2	Geocoder transform	19
2.10.3	Global Address Cleanse transform	21
2.10.4	USA Regulatory Address Cleanse transform	31
2.11	Documentation	

Introduction

1.1 Introduction to SAP BusinessObjects Data Services 4.1 Support Package 1

Welcome to SAP BusinessObjects Data Services 4.1 Support Package 1 (version 14.1.1.0). This *What's New* document highlights the new features available with this release, including those made available in the previous 4.1 (version 14.1.0.0) release.

For important information about this product release including installation notes, known issues, and fixed issues, see the SAP BusinessObjects Data Services Release Notes.

SAP offers other products that complement SAP BusinessObjects Data Services and provide additional SAP BusinessObjects enterprise information management solutions. These include:

- SAP BusinessObjects Data Federator
- SAP BusinessObjects Data Quality Management for Enterprise Applications
- SAP BusinessObjects Data Quality Management SDK
- SAP BusinessObjects Event Insight
- SAP BusinessObjects Information Steward
- SAP Master Data Governance for Embedded MDM
- SAP NetWeaver Business Warehouse (BW)
- SAP NetWeaver Master Data Management for Enterprise MDM
- SAP BusinessObjects Rapid Marts
- SAP BusinessObjects Text Analysis

For more information, see the SAP website or contact an SAP sales representative.

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.

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.

SAP BusinessObjects Data Services 4.1 features

2.1 Architecture

This version of Data Services includes several architecture improvements.

Adaptive processing server

Several SAP BusinessObjects Data Services processes are now run in the Adaptive Processing Server, which is hosted by SAP BusinessObjects Enterprise or Information platform services.

These processes include:

- RFC Server
- Administrator Service (used for log cleanup)
- View Data Service
- Metadata Browsing Service

Related Topics

Administrator's Guide: Configuring Metadata Browsing Service and View Data Service

2.2 Usability

This version of Data Services includes several enhancements that improve the user's experience.

Data ServicesWorkbench

The Data ServicesWorkbench is a new application that simplifies the migration of data and schema information between different database systems.

In previous versions, migrating data and schema information required you to create many dataflows in the Designer, with each dataflow reading from a single source table and writing to a template target table. This process could take days. In addition, incompatibilities between the source and target database types could require manual schema and data corrections.

The Data Services Workbench automates this migration process. Instead of creating many dataflows manually, you now provide connection information for the source and target databases and select the tables that you want to migrate. The Workbench automatically creates Data Services jobs, workflows,

and dataflows and imports them into a Data Services repository. You can execute and monitor these jobs from within the Workbench. In addition, the Workbench supports more advanced options such as bulk loading and delta loading.

Jobs created in the Workbench can be scheduled with the Data Services Management Console, and the generated objects can also be used as a starting point for further editing within the Data ServicesDesigner. For example, you might want to add more advanced transformations that are not available directly in the Workbench.

In this version of Data Services, the Workbench supports migration from Data Services-supported databases and SAP applications to SAP HANA, Sybase IQ, Sybase ASE, Oracle, Microsoft SQL Server, DB2, and Teradata targets.

For a complete list of supported sources and targets, as well as more information about using the Workbench to migrate data and schema information, see the *Workbench Guide*.

Design-Time Data Viewer

Previously in the SAP BusinessObjects Data ServicesDesigner, you could view data for only static sources in the object library such as tables, files, XML, COBOL copybooks, and corresponding sources and targets. To view data going through transforms, you had to use the Debugger.

Now, the Design-Time Data Viewer feature lets you view and analyze the input and output for a data set in real time as you design a transform. The data flow does not need to be complete or valid, although it must use a valid, accessible source that contains data.

You can display and configure Design-Time Data Viewer from the Debug menu.

For more information about Design-Time Data Viewer, see the Designer Guide.

Improved XML support through the XML Map transform

In previous versions of Data Services, you could process hierarchical data such as XML and IDocs through nesting and un-nesting within the output mapping of a Query transform. These methods present some challenges:

- Un-nesting data is a complicated process, particularly when the input schema is large and the output columns are from different schemas at different hierarchical levels.
- Un-nesting data is memory-intensive because the data needs to be transformed and copied.
- Converting from one hierarchical structure to another hierarchical structure is complicated. You must
 first un-nest the source structure into flat data, and then re-nest the flat data into the target hierarchical
 structure.

This version of Data Services adds a new XML_Map transform that simplifies support for hierarchical data structures such as XML and iDocs. The new XML_Map transform provides a simplified interface for nesting and un-nesting hierarchical data and converting from one hierarchical structure to another without an intermediate flattened-data step. The XML_Map transform supports one or more hierarchical or flat sources and produces one hierarchical or flat target, and supports common Query output schema concepts such as Distinct, Order By, Where, Group By, and Aggregation.

For more information about using the XML_Map transform, see the *Reference Guide*.

Locale Selector GUI is now available for UNIX/Linux

In previous versions UNIX and Linux users had to change the locale settings in the DSConfig.txt file. Now the locale setting can be changed in the Locale Selector GUI.

For more information accessing and configuring the UNIX/Linux Locale Selector, see the *Reference Guide*.

DSN-less and TNS-less connections

In previous versions, when you have a server group with multiple Job Servers running on different machines, you need to create a DSN or TNS entry on every machine. In addition, you need to modify the DSN or TNS entry on every machine if you want to change the connection parameters. This version provides server name connections (also known as DSN-less and TNS-less connections) that removes the need to define a DSN or TNS entry for creating a repository and data store. In addition, Data Services provides a Windows ODBC Driver Selector to configure drivers when using DSN-less connections to ODBC sources.

For more information, see the Administrator Guide.

Connection Manager is now available for UNIX/Linux

In previous versions, UNIX and Linux users had to perform a number of manual steps to add more database connections for sources and targets. Now the Connection Manager provides a graphical user interface through which you can do the following:

- Create, edit, or delete ODBC data sources and drivers after installation for DSN connections, configure the ds_odbc.ini file, and restart appropriate services.
- Create, edit, or delete ODBC drivers for DSN-less and TNS-less connections after installation, configure the ODBCINST file, and restart appropriate services.

For more information accessing and configuring the UNIX/Linux Connection Manager, see the *Administrator's Guide*.

2.3 Enhanced SAP HANA support

This release of Data Services includes several enhancements for integrating with the SAP HANA database.

SAP HANA repository support

Data Services now supports the SAP HANA database as a repository host. You are no longer required to use a separate database to host the Data Services repository.

For more information about using SAP HANA as a repository, see the *Installation Guide*.

SAP HANA performance improvements

Data Services automatically detects the SAP HANA target table type and updates the table accordingly for optimal performance.

For more information, see the *Performance Optimization Guide*: Bulk Loading and Reading.

Bulk updates enhancement

For improved performance when using changed-data capture or auto-correct load, Data Services now uses a transparent staging table for bulk loading to targets.

For more information, see the Performance Optimization Guide: Bulk Loading and Reading.

Support for stored procedures

Data Services now supports SAP HANA stored procedures. Scalar data types are supported for input and output parameters. Procedures can be called from a script or a Query transform as a new function call.

For more information, see the *Reference Guide*: Functions and Procedures.

2.4 Enhanced extraction capabilities for SAP Business Suite

This release includes the following enhancements for interoperating with SAP systems.

Data streaming in ABAP data flows

This version includes a new data transfer method option **RFC**, which lets you stream data from the source SAP system directly to the Data Services data flow process using RFC. This enhancement eliminates the need for intermediate data files and also provides the ability to adjust the data transfer package size.

For details, see the Supplement for SAP.

SAP table reader in regular data flows

Previously, reading large data volumes could result in an out-of-memory error on the SAP source side. The Data Services SAP table reader in regular data flows can now fetch data in batch from an SAP system. This new implementation allows the reader to process large volumes of data and mitigates out-of-memory errors. Also included are an **Array fetch size option**, which allows the data to be sent in chunks, avoiding large caches on the source side, and an **Execute in background (batch)** option, which lets you run the SAP table reader in batch mode (using a background work process) for time-consuming transactions.

For details, see the Supplement for SAP.

Parallel reading from business content extractors

Data Services now supports multithreading for SAP extractors for improved performance.

For more information, see the Performance Optimization Guide: Bulk Loading and Reading.

New ABAP functions for improved security

This version includes several enhancements to functions, procedures, and authorizations that provide secure integration with SAP systems. You can now at a more granular level assign authorizations that can limit which users can execute functions and which programs can be started.

In addition, the new secured functions are now delivered into a new /BODS/ namespace instead of the customer Z namespace.

For details, see the Supplement for SAP.

Note:

You must update the Data Services-supplied transport files and update the SAP authorizations to the 4.1 version. For details, see the *Supplement for SAP*.

SNC authentication and load-balancing support in SAP datastores

All SAP datastore types now include options for authentication using Secure Network Communications (SNC). SNC provides a more secure and efficient method for moving data from ABAP executable programs directly to the Data Services Job Server engine. New load-balancing options are also available for all SAP datastores.

For details, see the Supplement for SAP.

Parallel SAP connections for Open Hub tables

Data Services now supports multiple Open Hub connections to read multiple data packets from a BW system simultaneously for improved performance.

For more details, see the Supplement for SAP: "Open Hub Table source".

2.5 Sybase IQ loading performance enhancements

This version includes enhancements for bulk loading to Sybase IQ databases for improved performance.

Bulk updates enhancement

For improved performance when using changed-data capture or auto-correct load, Data Services now uses a transparent staging table for bulk loading to targets.

Binary data format support

Data Services now supports the Sybase IQ binary format for faster loading.

For more information, see the Performance Optimization Guide: Bulk Loading and Reading.

2.6 Connectivity: Big data and CDC

Hadoop integration

Data Services now offers connectivity to Hadoop sources including Hadoop distributed file systems (HDFS) and Hive data warehouses.

For more information, see the Reference Guide.

Microsoft SQL Server 2008 CDC support

For capturing changed data from Microsoft SQL Server 2008 sources, Data Services now supports the native methods Changed-Data Capture and Change Tracking in addition to the existing Replication Server method. The Changed-Data Capture and Change Tracking methods offer more flexibility and efficiency depending on your data processing needs.

For more information, see the *Designer Guide*, "Capturing Changed Data".

SuccessFactors adapter

Data Services now offers a SuccessFactors adapter. The SuccessFactors adapter is installed with every Data Services job server.

After creating an adapter instance and an adapter datastore, you can import a SuccessFactors database table to use as a source or a target in a Data Services dataflow.

For more information, see the "Using the SuccessFactors adapter" section in the Integrator's Guide.

Bulk loading using DataDirect Wire Protocol SQL Server ODBC driver

Use the DataDirect's Wire Protocol SQL Server ODBC driver bulk load feature to quickly insert and update a large number of records into a database. You don't need to use a separate database load utility because the bulk load feature is built into the driver. DataDirect drivers are included in the Data Services installation.

For more information, see the "Bulk loading using DataDirect's Wire Protocol SQL Server ODBC driver" section in the *Performance Optimization Guide*.

Continuous execution type workflow and Single execution type workflow

Data Services now supports two new workflow execution types, continuous and single.

The continuous execution type flow runs all its child data flows in a loop but keeps them in the memory for the next iteration. This avoids the need to run some of the common steps of execution such as connect to repository, parse/optimize/compile ATL, and open database connections. This will improve the speed of repeated data flow execution.

Note:

- Platform transforms with the exception of XML_Map cannot be used in a Continuous execution type flow.
- Data Integrator transforms with the exception of Data_Transfer, Table_Comparison, and XML_Pipeline cannot be used in a Continuous execution type flow.

The single execution type flow runs all its child data flows in one Operating System process. If the data flows are designed to run in parallel, then they are run in different threads instead of different processes. The advantage of single process is that it is possible to share resources such as database connections across multiple data flows.

Note:

A Single execution type flow can be only executed by a Continuous execution type flow. A Single execution type flow cannot call another Single execution type flow.

For more information the new options, see the Designer Guide.

2.7 Monitoring, administration, and installation

Support for Windows User Access Control (UAC)

Windows Vista and newer versions of Windows include a security feature known as User Access Control (UAC). In general, UAC addresses issues in system stability and security by running most applications with standard user rights instead of administrator-level rights, as in previous versions of Windows.

In this version of SAP BusinessObjects Data Services, UAC support has been improved so that standard users do not require elevated access rights in order to run Data Services applications such as the Designer.

Previous versions of Data Services stored all configuration options in a location that required elevated access rights to modify. Because of this, standard users could run applications such as the Designer only in elevated access mode.

The Data Services installation program now creates additional environment variables: DS_COMMON_DIR on all platforms, and DS_USER_DIR on Windows platforms.

User-specific configuration options have been moved to a new file in DS_USER_DIR, where standard users have full access by default. This allows a standard user to run applications such as the Designer without requiring elevated access.

Configuration options	Old location	New location	
Common options	<pre><link_dir>/bin/DSCon fig.txt</link_dir></pre>	<pre><ds_common_dir>/conf/DSCon fig.txt</ds_common_dir></pre>	
User-specific options	<pre><link_dir>/bin/DSCon fig.txt</link_dir></pre>	<pre><ds_user_dir>/conf/DSUserCon fig.txt</ds_user_dir></pre>	

Monitor log enhancements: Time-based sampling

In previous versions of Data Services, the performance data collected for the monitor log was based on the number of rows processed. For example, if your monitor rate was 1,000 rows, the performance data would be written to the log after each 1,000 rows have been processed. This sampling was performed individually by many different components, causing possible contention and performance issues when writing to the log file.

In this version of Data Services, performance data is now written to the monitor log at fixed intervals based on a timer. The new monitor log timer avoids potential file contention by centralizing log file accesses. The default sample rate is 30 seconds.

In addition, new performance data is written to the monitor log:

- The CPU utilization for each transform thread
- The number of rows in the input buffer and its used capacity (as a percentage)
- The number of input rows currently cached by each transform

Restriction:

The new monitor log columns are not visible in existing user interfaces such as the Designer and Management Console. If you want to view the new columns, open the monitor log file directly in an external editor.

SAP Solution Manager improvements: Alerting and Heartbeat

Heartbeat monitoring lets you use the SAP Solution Manager to check whether a component such as a Job Server or Access Server is up and running. You can also get information about real-time services for Access Servers. Alerts let you view critical errors in the SAP Solution Manager; in Data Services, Job Servers send alerts when a job fails.

For more information, see the Administrator's Guide: Integration with SAP and SAP Solution Manager.

IPv6 support

Data Services 4.1 supports IPv6.

Get a labeled object with filtering in the central repository library

To help organize and track the status of objects in your application, you can label objects.

The new filtering option for the Get by label operation allows you to filter (selectively change) environment-specific information in object definitions when working in a multi-user environment.

To use this option, open the central object library and select the highest level object you want to get. Right-click on the object, and select **Get By Label > With filtering**.

2.8 Functions and transforms

This version of Data Services includes enhancements to function support.

More pushdown of functions

For improved performance, the lookup_ext function now supports database pushdown when used in the column mapping, output schema, or SELECT WHERE clause of a Query transform.

For more information, see the "Functions and Procedures" section of the Reference Guide.

Enhanced functions for locale-specific case operations

We have added a new locale parameter to the upper(), lower(), and init_cap() functions to allow casing functionality for different languages.

Filtering updated rows in the Map_CDC_Operation transform

You can now drag and drop columns from the input schema to the **Define columns to filter updated rows** box to filter CDC updated rows.

Filtering comparison rows in the Table_Comparison transform

You can now use filtering expressions in the Table_Comparison transform to limit the rows that are compared to the input data set. Without filtering, all rows in the comparison table are considered for comparison against the input.

For more information about filtering comparison rows in the Table_Comparison transform, see "Transforms, Data Integrator transforms, Table_Comparison, Options" in the *Reference Guide*

2.9 Text Data Processing

Text Data Processing analyzes content and automatically identifies and extracts entities and facts in multiple languages. It includes a set of transforms that work together to improve the quality of your data. It has been enhanced with the following features:

Expanded language coverage

Language coverage has been expanded from 6 to 31.Text Data Processing can process text in 31 languages including several right-to-left languages.

Aligned entity type names for TDP and DQ

The entity type names used by the Entity Extraction transform in supported languages are aligned with Data Quality input field names so that users can use the same names in TDP and DQ transforms. For example, the **CITY** entity type name is no longer used in TDP. Instead, this entity type name is aligned to the **DQ LOCALITY** input field name.

Added Processing Timeout option

The Processing Timeout option governs the maximum amount of time spent processing a piece of text.

Added process logging for auditing and support resolution

The TEXT_ID input field supports tracing a piece of text back to its source when a processing error occurs.

Social Media entity type support

The Text Data Processing Entity Extraction transform now extracts Twitter handles and topics in English, French, German, Korean, and Spanish using the SOCIAL_MEDIA entity type.

Entity Extraction pushdown to Hadoop

Data Services can to pull directly from Hadoop using either Hive or HDFS and push data back to it as a target. Hadoop is used a lot to store social media data or other unstructured content. The Text Data

Processing Entity Extraction transform is the first transform to be pushed down to Hadoop as a Map Reduce function. This enhances performance when processing Hadoop content.

Processing of binary source documents

The Text Data Processing Entity Extraction transform now supports the processing of a number of binary source formats:

- Microsoft Word: 2003, 2007, and 2010 (Office Open XML)
- Microsoft PowerPoint: 2003, 2007, and 2010
- Microsoft Excel: 2003, 2007, and 2010
- Adobe PDF: 1.3 1.7
- Microsoft RTF: 1.8 and 1.9.1
- Microsoft Outlook E-mail Message: 2003, 2007, 2010
- Generic E-mail Message: ".eml" files
- Open Document Text, Spreadsheet, and Presentation: 1.0, 1.1, 1.2
- Corel WordPerfect: 6.0 (1993) X5 (2010)

The text content of such files can now be extracted when encountered. A new processing option, "Document Properties", governs whether certain additional information that might be present in the file, such as creation date and author name, should also be extracted.

Emoticon extraction support

The Text Data Processing Entity Extraction transform now extracts EMOTICON entities in English using the EMOTICON rules in the Voice of the Customer extraction pack.

Profanity extraction support

The Text Data Processing Entity Extraction transform now extracts PROFANITY entities in English using the PROFANITY dictionary in the Voice of the Customer extraction pack.

Expansion of Italian and Japanese extraction

The Text Data Processing Entity Extraction transform now supports more than 30 predfined entity types in Italian and Japanese.

Related Topics

- Designer Guide: Unstructured file formats
- Text Data Processing Language Reference Guide: Linguistic Analysis Language Feature Matrix
- Reference Guide: Processing options
- Reference Guide: Input fields

2.10 Data Quality

The set of transforms that work together to improve the quality of your data has been enhanced with the following features.

Global Suggestion Lists

Global Suggestion Lists will be used only to query data due to the introduction of the Global Address engine's Suggestion List option. To view changes to the available options and components, see the *Data Services Reference Guide*. For additional information on updating your current Global Suggestion List processes, see the *Data Services Upgrade Guide*.

Related Topics

- Reference Guide: Transforms, Data Quality Transforms, Global Suggestion Lists
- · Reference Guide: Data Quality transforms

2.10.1 Data Cleanse transform

The Data Cleanse solution has been significantly enhanced to include the following functionality:

New single SAP-supplied person and firm cleansing package

Support for the new single SAP-supplied person and firm cleansing package. The new cleansing package consolidates and improves on the previous regional cleansing packages. Utilizing a single multi-region cleansing package allows the Data Services Data Cleanse transform to cleanse your data across multiple regions or domains within a single data flow. Additionally you can specify the specific domain you want Data Cleanse to apply when formatting the output.

The new SAP-supplied cleansing package allows you to cleanse data across domain boundaries without needing to construct special parsing rules. It also consolidates all variations that previously existed in over 20 separate cleansing packages.

Note:

As a result of merging all domains, family names may be standardized to include diacritics when in the past they may not have. For example, the family name SANCHEZ was previously standardized in the Spanish cleansing package to "Sánchez" and in all other cleansing packages to "Sanchez". Now in the single cleansing package, this name will always be standardized to "Sánchez" regardless of the content domain sequence the user selects in the transform. If you do not want the name data to receive the diacritics, use the custom function CF_RemoveDiacriticalCharacters in a Query following Data Cleanse to remove the diacritics. Apply this function to the person fields and firm fields, since many firm names consist of person name data. To download this custom function, access http://www.sdn.sap.com/irj/boc/blueprints.

Note:

As a result of merging all domains, name prefixes referred to in the cleansing package as Pre_Given_Name and Pre_Family_Name will be standardized to lower case. For example, family names that were previously standardized to Maria Hernández del Rio and Jean Yves la Croix by some cleansing packages, and standardized to Maria Hernández Del Rio and Jean Yves La Croix by others, will now be standardized to Maria Hernández del Rio and Jean Yves la Croix regardless of the content domain sequence the user selects in the transform. If you want all of these name prefixes to be mixed case, use the custom function CF_PersonInitCapPrefix in a Query following Data Cleanse to change the first

letter of each name prefix to upper case. Apply this function only to the person name fields. To download this custom function, access http://www.sdn.sap.com/irj/boc/blueprints.

Two new options support the new cleansing package

Two new options are added to support the new single SAP-supplied person and firm cleansing package. The Content Domain Sequence option controls how Data Cleanse parses your domain-specific data based on the domain you select. And the Output Format option enables you to specify the domain for the output data. Based on the specified domain in the Output Format, Data Cleanse uses certain output fields and formats the data in those fields according to the regional standards.

Change to DQM SDK installation requirement

There is no longer a requirement to separately download the cleansing package for the DQM SDK install.

Support for Chinese data

The single SAP-supplied person and firm cleansing package supports cleansing Chinese data in both Simplified Chinese and Traditional Chinese.

Updates to Date options

The Date options have been enhanced and updated to support additional languages, including Chinese and Japanese.

International phone support

Improvements have been made to support international phone processing.

Social Security Number file

The U.S. Social Security Administration is no longer updating its Social Security Number algorithm. Thus SAP no longer provides an updated DRLSSN.dat file. The file dated July 5, 2011 is the final file that SAP will provide.

For more information about the Social Security Administration's new assignment process, see http://www.ssa.gov/employer/randomizationfaqs.html.

One-to-one mapping option

The new one-to-one mapping option controls how the Date, Email, and Phone parsers output data in version 4.1.1. When the option is selected, the input data is placed in the corresponding output field, so that it is easier for you to see where the input data is located on output. For example, if Phone1 is blank and Phone2 contains data, the data is output into the Phone2 field. When this option is not enabled, the data is output to Phone1.

Note:

Due to the implementation of this feature, your output may look different because the data is parsed in a different order. This occurs even when this option is set to **No**.

The example shows the input data and how the data is output based on whether this option is set to **Yes** or **No**.

Note:

This example shows Date fields. The same example also applies to Phone and Email fields.

Field	Input data	Output data when option is No	Output data when option is Yes
Date1	<blank></blank>	1968/01/01	
Date2	1968/01/01	1968/02/02	1968/01/01
Date3	1968/02/02	1968/03/03	1968/02/02
Date4	<blank></blank>	1968/04/04	
Date5	1968/03/03	 	1968/03/03
Date6	1968/04/04	 	1968/04/04

Related Topics

- Designer Guide: Data Quality, Data Cleanse, About cleansing data
- · Reference Guide: Transforms, Data Quality transforms, Data Cleanse
- Reference Guide: Date options

2.10.2 Geocoder transform

The Geocoder transform has been enhanced with the following features.

New dynamic Option_Distance_Unit input field

The Option_Distance_Unit input field has been added that allows you to change the Geocoder transform's Distance Unit option setting after the transform is initialized, without having to terminate and reinitialize the transform. You can pass the new setting through the input field to the transform. The transform gets the updated setting from the input field and adjust its processing to use the new setting, before processing the incoming record. Dynamic setting values that are specified in the input field are only valid for that record and do not affect any subsequent record.

The value of the Option_Distance_Unit input field takes precedence over the value of the Distance Unit option. The value of the Distance Unit option is only used if the Option_Distance_Unit input field is not mapped or is invalid, blank, or NULL.

Improved synchronization between the Geocoder and Global Address Cleanse transforms

Two new output fields, Address_Line and Postcode, have been added to improve synchronization between the Geocoder transform output address components and the Global Address Cleanse transform input address components. This will enhance usability when you want to use the Geocoder transform to search for points-of-interest and then use the Global Address Cleanse transform to validate and standardize the address.

Closest house number latitude/longitude assignment

When you search a house number that is not in the reference directory, the latitude and longitude for the closest house number on the same street can be assigned. An information code will be generated to indicate that the assignment is not the input house number.

Example:

Input address: 100 main Street La Crosse WI 54650

If there are only two records in the geocoder reference data:

Record 1: 1-50 main Street La Crosse WI 54650

Record 2: 51-80 main Street La Crosse WI 54650

Previously, the Geocoder transform output the information code 005 to indicate that the house number "100" cannot be found in the reference data. Now, when the street level matches but the house number doesn't, you can get the closest latitude/longitude value for the house number. In the example, the house number "80" is the closest, and its latitude and longitude is returned.

Ranged house number latitude/longitude assignment

In some countries, house numbers can be a range. For example, in the address "SE17 2AL London Walworth Road 367-369", 367-369 is a ranged house number. For the United Kingdom and Australia, the Geocoder transform search now supports this type of house number and returns its latitude and longitude.

New country directory data

For this release, directory data is available for two additional countries: Australia and Switzerland.

The Geocoder transform is flexible enough to accept new country directory data immediately after the directory data is released. There is no need to wait for the next Data Services release to begin using new country directory data. At the time of this publication, the Geocoder transform uses directory data for the following countries:

- Australia
- Canada
- France
- Germany
- Switzerland
- United Kingdom
- United States

Check with your sales representative for a list of the most current country directories available.

Related Topics

- Reference Guide: Data Quality transforms, Dynamic transform settings
- Reference Guide: Data Quality transforms, Geocoder, Input fields
- · Reference Guide: Data Quality transforms, Geocoder, Output fields

2.10.3 Global Address Cleanse transform

New Dynamic input fields

The Global Address Cleanse transform now supports thirteen dynamic transform fields.

Dynamic transform settings allow the user to change a transform's settings after the transform is initialized, without having to terminate and reinitialize the transform. You can pass each new setting through an input field to the transform. The transform will retrieve an updated setting from the input field and adjust its processing to use the new setting, before processing the incoming record.

The settings a transform is initialized with are considered the transform's default settings. Dynamic setting values that are specified in the input fields are valid only for that record and do not affect any subsequent record. If the value specified for a given option is NULL or blank, then the record will be processed with the default setting for that option. If the dynamic setting is invalid, then the transform will log a warning and then use the default settings.

The dynamic input fields are:

- OPTION CANADA OUTPUT ADDRESS LANGUAGE
- OPTION GAC DUAL ADDRESS
- OPTION STANDARDIZATION ADDRESS LINE ALIAS
- OPTION STANDARDIZATION ASSIGN LOCALITY
- OPTION STANDARDIZATION CAPITALIZATION
- OPTION_STANDARDIZATION_CHARACTER_WIDTH_STYLE
- OPTION STANDARDIZATION DIRECTIONAL STYLE
- OPTION_STANDARDIZATION_OUTPUT_COUNTRY_LANGUAGE
- OPTION STANDARDIZATION POSTAL PHRASE STYLE
- OPTION STANDARDIZATION PRIMARY TYPE STYLE
- OPTION STANDARDIZATION REGION STYLE
- OPTION STANDARDIZATION SECONDARY DESCRIPTION STYLE
- OPTION STANDARDIZATION SECONDARY NUMBER STYLE

New information code for quick data entry

Quick data entry allows for address assignment with an input of only country and postcode (other input options are country, postcode, and primary range, or country, postcode, and secondary range). Quick data entry is currently supported for the United Kingdom and Netherlands. A new information code has been added to better explain quick data entry assignments.

Information Code	Description	Engine
3300	The postcode only lookup returned multiple primary names.	Global Address

La Poste Ratification

With this release, the Global Address Cleanse transform can perform certified processing of French addresses for mailing purposes.

Setting the "Disable Certification" option to **No** enables certified processing. When performing certified processing for mailing purposes:

- All punctuation is removed except for firm data.
- Accented characters have their accents removed (only A-Z and 0-9 are allowed).
- The address is returned in 6 lines.

Certified Australia options

The Australia processing options meet all requirements for AMAS 2012 certification.

Certified Canada options

The Canada processing options meet all requirements for SERP 2012 certification. Canada Post has extended the transition period for excluded large urban business addresses to January 2013. This allows mailers more time to correct these addresses before they are considered invalid and count against the Statement of Accuracy score. No further extensions will be provided after January 2013.

Certified New Zealand options

The New Zealand processing options meet all requirements for SendRight 2012 certification.

Suggestion List Address Range option

In order to maintain consistency across all engines in the Global Address Cleanse transform, the Address Range option was added to the Global Address engine Suggestion List options.

Option	Description
	Specifies a span around the input primary address range for which to return suggestions. By using this option, you can limit the suggestions returned to be within a few blocks of your input. For example, assume you entered 500 for this value. Then, you submit the following street address:
Address Range	1000 Pine St.
	Suggestions would be returned in a range from 750 to 1250 Pine Street.
	If you don't want to limit the ranges returned in suggestions, type 0. The maximum value is 5000.

Suggestion List support for Japan and China

The Global Address Cleanse Suggestion List options now support processing of Japanese and Chinese addresses.

Improved standardization format for Austria, Germany, Liechtenstein, and Switzerland

The EMEA address assignment logic has been improved to meet local, cultural standards, and business requirements. The Standardization option, Use Local Primary Type Style, has been added to enable use of the Primary Type Style that is present in the reference data for Austria, Germany, Liechtenstein, and Switzerland.

Option	Description
Use Local Primary Type Style	Specifies whether to use the type style for primary address components that is present in the address data. Setting this option to Yes ignores the Primary Type Style option.
	Yes: Uses the Primary Type Style that is present in the address data.
	No : Uses the Primary Type Style specified in the Primary Type Style option.

New and updated output fields for several countries

New and updated output components are available in the Global Address Cleanse transform to support additional address information for Austria, Belgium, Canada, France, Germany, Liechtenstein, Poland, Spain, and Switzerland. The following tables describe the fields as they are used by the supported country data.

Field name	Description	Engine
Additional_Info1	Austria: Includes the PAC code of the currently valid address when you choose to preserve the alias address on output.	C, G
	Belgium: Includes the NIS code.	
	Canada : The official 13-character abbreviation of the city name, or the full spelling if the city name is less than 13 characters (including spaces).	
	For ACE Canada users, this maps to the CACE_CITY13 output field.	
	France: Includes the INSEE code.	
	Germany : Includes a portion of the German freightcode (Frachtleitcode).	
	Liechtenstein : Includes the postal service district (Botenbezirke) when it is available in the data.	
	Poland: Includes the district name (powiat).	
	Spain: Includes the INE 91 section code.	
	Switzerland : Includes the postal service district (Botenbezirke) when it is available in the data.	
Additional_Info2	Austria: Includes the City ID (OKZ).	C, G
	Canada : The official 18-character abbreviation of the city name, or the full spelling if the city name is less than 18 characters (including spaces).	
	For ACE Canada users, this maps to the CACE_CITY18 output field.	
	Germany: Includes the District Code.	
	Liechtenstein: Additional postcode.	
	Poland: Includes the community name (gmina).	
	Spain: Includes the INE Street code.	
	Switzerland: Additional postcode.	
Additional_Info3	Austria: Includes the Pusher-Leitcode (parcel).	G
	Germany: Includes the German City ID (ALORT).	
	Spain: Includes the INE Town code.	

Field name	Description	Engine
Additional_Info4	Austria: Includes the Pusher-Leitcode (letter). Germany: Includes the German street name ID (StrSchl).	G
Additional_Info5	Austria: Includes the SKZ Street Code (7-digit). Germany: Includes the discount code for the freightcode.	G
Additional_Info6	Austria : Includes the corner-house identification (1-digit). The value for a corner house is 1 .	G
Delivery_Point	Austria: Includes the PAC code, which is a unique identifier assigned by the Austrian postal authority.	G

Improvements in processing mixed-case reference data

With improved assignment logic and the availability of mixed-case reference data for Austria, Germany, and Switzerland/Liechtenstein, the Global Address Cleanse transform returns the data cased in the format contained in the data for that country when the "Capitalization" option is set to **MIXED**.

Improved assignment for Chinese addresses

The Global Address engine now supports Pinyin fuzzy logic to correct Kanji misspellings when processing Chinese addresses. Here are some examples where Pinyin fuzzy logic is used to correct Kanji misspellings:

Example 1:

Input	
上海市 埔 东新区	

Postcode	Region	Locality1	Locality2	Primary Name	Primary Num- ber
201203	上海 市	上海 市	浦东新区		

Example 2:

Input 上海汇金房地有限公司 上海赵家 浜路 1000 号

Postcode	Region	Locality1	Locality2	Primary Name	Primary Number	Firm
200030	上海 市	上海 市	徐汇区	肇嘉 浜路	1000 号	上海汇金房地 有限公司

Example 3:

Input
北京奥斯达商贸有限公司
北 山 环中路 20 号双秀公园东门
100088 中国北京

Postcode	Region	Locality1		Primary Name	Primary Range	Primary Number Ex- tra	Firm
100088	北京市	北京市	海淀区	北三环中路	20 号	双秀公园东 门	北京奥斯达 商贸有限公 司

New output fields to support India

The following components are added to support address processing for India.

Field name	Description
Area_Name1	An industrial area such as RIICO INDUSTRIAL AREA.
Building_Name2	The building name for the address, which in some countries is used in place of the primary number.
Point_Of_Reference1-2	A well known place or easily visible location to help locate an address. For example, Opposite to Citibank ATM.
Region2	Either the Region2_Name or Region2_Symbol based on the standardization option Region Style.
Region2_Description	Region2 Description.

Field name	Description
Region2_Full	Includes Region2 and Region2_Description.
Region2_Name	The fully spelled out Region2 name.
Region2_Symbol	The two- or three-character representation of the region2 name.

Standardization option value name changes

The following Secondary Number Style option value names have been changed to more accurately describe the options.

- Dashed has been changed to Attached.
- Trailing has been changed to Unattached.

Discrete input support for house number

The house number can now be entered in a separate multiline input field.

This feature is also available for the USA Regulatory Address Cleanse transform.

Example 1:

Input		
Multiline1: 38/3		
Multiline2: Gerhart-Hauptmann-Straße		
Locality1: Innsbruck		
Postcode: 6020		
Country: AT		

Example 2:

nput address
Multiline1: 332
Multiline2: Front St S
City: La Crosse
State: WI
Zip Code: 54601
Country: US

German abbreviated street format

You can now output abbreviated German street addresses. Use the Street Name Style option to specify how it is output.

Option	Description		
Street Name Style	Specifies the format for street data for addresses. This option applies to addresses in Germany and the Netherlands. Preserve: Preserves the street data format as it was input. Short: Outputs street data in the format preferred by the postal authority. For the Netherlands, this returns a street address with a maximum of 24 characters in mixed case. Note: To use the short street name style, the Address Line Alias option must be set to Convert. Note: The Capitalization option in the Global Address Cleanse transform overrides the Street Name Style option.		

German locality addition information

The Global Address Cleanse transform now outputs consistent German locality addition information. Two new Standardization options, Include Locality Addition and Locality Name Style, have been added to enable use of the city addition information.

Option	Description		
Include Locality Addition	Specifies whether the Locality1_Full output field contains both the Locality1_Name and the Locality1_Addition information.		
	Yes: Includes both the locality and the locality addition information.		
	No: Does not include the locality addition.		
	Preserve: Includes locality addition if found on input. This is the default setting.		
	Note: The Locality Name Style option in the Global Address Cleanse transform overrides this option. If the Locality Name Style option is set to Short, the Locality1_Full field will not contain the locality addition information.		
Locality Name Style	Specifies the format for locality data in the Locality1_Name output field for addresses.		
	This option applies to German addresses.		
	Preserve : Preserves the locality data format as it was input. This is the default setting.		
	Short : Outputs locality data in the abbreviated version, if available in the reference data.		
	Note: To use the short locality name style, the Address Line Alias option must be set to Convert.		
	Note: This option overrides the Include Locality Addition option.		

A new output field, Locality1_Addition, contains the locality addition information. The existing Locality1_Name field has changed to include only the locality information and not the addition. For example, for Freiberg am Neckar, Freiberg is output in the Locality1_Name field and am Neckar is output in the Locality1_Addition field. The Locality1_Full output field may or may not include locality addition information depending on the setting of the new Include Locality Addition option.

Output field name	Description
Locality1_Addition	Additional locality information.

Support missing leading zero in German postal codes

The Global Address Cleanse transform now supports German postal codes that are missing a leading zero (for example, 01234).

Improved UK address assignment

UK address assignment has been improved with the following enhancements:

- **Region correction**: The Global Address Cleanse transform outputs the correct region that matches the input city.
- Local locality aliases: The Global Address Cleanse transform supports a limited number of additional local locality aliases, which are short and local names. For example, for the locality GORLESTON, the alias is GORLESTON ON SEA. London aliases are not currently supported.
- **Firm matching**: Address assignment of UK firms disallows two-word firm matches in some cases when one of the words is a poor match, but the rest matches well.

Postal code and reverse postal code lookup

This feature is integrated with the existing Global Suggestion List. When an input address record is assigned to a locality, a lastline suggestion list may be returned that contains the optional postcodes. This feature is supported by the Global Address engine, not the US or Canada engines.

Full country support for China

The Global Address Cleanse transform can now cleanse address line data for the entire country of China, so that you can parse, validate, and correct Chinese address information beyond the major cities and provinces.

Address-line validation for Mexico

The Global Address Cleanse transform now supports address-line data for Mexico, in addition to last-line support.

Mixed scripts in one Global Address Cleanse transform

Previously, if your data included both CJK (Chinese, Japanese, and Korean) script records and Latin script records, you had to map the input fields differently, and route the records through two separate Global Address Cleanse transforms. In this release, you can now map international addresses (both Latin and CJK scripts) in the same input formats, so that you can use one Global Address Cleanse transform to cleanse your address data.

Expanded major locality support for country identification

Country identification has been improved by adding localities to the dictionary. The additional localities are used to determine the country when the input data does not specify it.

2.10.4 USA Regulatory Address Cleanse transform

NCOALink certification

This release addresses two issues relating to NCOALink certification so that you can successfully complete self-certification with the USPS. If you are an NCOALink licensee, you should install these fixes at your earliest convenience. The changes are required to pass NCOALink certification testing.

In the USPS NCOALink certification stage 2 test file:

- An input record containing the postname "V" is now processed correctly.
- A specific address record that was incorrectly assigned a return code of 14 is now assigned the correct return code of 16.

CASS Cycle N

The USA Regulatory Address Cleanse transform in this version of SAP BusinessObjects Data Services contains updates to CASS Cycle N. It also contains other USPS regulatory updates including additional benefits for DSF2 walk sequencing.

This version of the transform has been recertified for CASS Cycle N to support the USPS CASS Cycle N rule change and rule clarification. These changes provide greater flexibility for determining what is included on address lines with regard to SuiteLink secondary address information and invalid secondary address information.

Note:

Data Services 4.0 (14.0.1.0 and beyond) is fully certified for CASS Cycle N. Therefore, you are not required to upgrade to this version of Data Services 4.1 (14.1.1.0) to maintain the CASS Cycle N certification. If you have already installed Data Services 4.0 (14.0.1.0 or beyond) and you decide that you will not use the features detailed in this document, you do not have to upgrade to this version. However, SAP recommends that you upgrade.

CASS Cycle N certified engines

The table below contains the CASS Cycle N certified engine versions and associated Data Services releases and whether it contains the USPS CASS rule change and clarification detailed in this document.

Data Services release (version)	CASS Cycle N certified	CASS version	Includes USPS CASS Cycle N rule change and clarification
3.2 (12.2.3.1 – 12.2.3.3)	Yes	8.03.00.N	No
3.2 (12.2.3.4 and be- yond)	Yes	8.04.00.N	Yes
4.0 (14.0.1.0 – 14.0.2.1)	Yes	8.05.00.N	No
4.0 (14.0.2.2 and be- yond)	Yes	8.06.00.N	Yes
4.1 (14.1.0.0 and be- yond)	Yes	8.06.00.N	Yes

Note:

If you are using Data Services 3.2, consider upgrading to Data Services 4.0 or 4.1, because Data Services 3.2 reaches end of mainstream support after March 31, 2013.

CASS self-certified users: If you previously CASS Cycle N self-certified with a CASS version of 8.03.00.N or 8.05.00.N, you must re-self-certify with a CASS 8.04.00.N or 8.06.00.N version of Data Services.

DSF2 walk sequence improvement

The DSF2 Walk Sequencer transform uses DSF2 data to assign sequence numbers for all addresses that are DSF2-confirmed delivery points. With this release, the number of walk-sequenced records may increase (thus improving your postal discounts) because the software now sequences addresses with DSF2 status codes of D, S, and Y. Previously the software sequenced addresses only for records with a DSF2 status of Y.

CASS Cycle N rule change and rule clarification

The CASS engine in this version of Data Services has been updated and recertified to support a CASS Cycle N rule change and rule clarification. The changes were announced after the assignment engine in Data Services 4.0 (support packs 1 and 2) had received CASS Cycle N certification by the USPS.

Rule change

The USPS announced a CASS Cycle N rule change that states the secondary information returned by SuiteLink does not need to appear in the address line but must always be reflected in the ZIP+4. For information about how to control this behavior, see the "New Add Firm Match Secondary option" section.

Rule clarification

The USPS clarified a CASS Cycle N rule that requires that the invalid secondary address information is retained when a SuiteLink match is made. The USPS clarified that the invalid secondary address information does not have to remain on the standardized address line. For information about controlling this behavior, see the "Retain invalid secondary when a SuiteLink match is made" section.

New Add Firm Match Secondary option

The original CASS Cycle N requirement for SuiteLink affected CASS customers who do not want to update their address lines with SuiteLink information because of industry requirements or customer requests. In response to this situation, the USPS updated the SuiteLink rule, stating that adding the SuiteLink-supplied secondary address information to the address is optional but that the SuiteLink secondary address information must be reflected in the ZIP+4.

To provide the most flexibility for posting SuiteLink information to the address line, SAP BusinessObjects has included a new option in the USA Regulatory Address Cleanse transform named Add Firm Match Secondary (the default setting is Yes) located in the "Standardization Options" group.

If you use SuiteLink and do not want to update your address lines with SuiteLink secondary information, you can set the new option Add Firm Match Secondary to No. The software updates the ZIP+4 to reflect the SuiteLink secondary information, but does not update the address line with the SuiteLink secondary information. The software standardizes the address line based on your standardization style settings in the job setup.

Retain invalid secondary when a SuiteLink match is made

The CASS Cycle N requirement to retain the invalid secondary address information has been clarified to state that the invalid secondary address information does not have to remain on the address line, but it must be retained in a field that you can access if you need to.

You can control this behavior by using the Include Unused Address Line Data option in the "Standardization Options" group. This option controls whether the software outputs the unused address line data in the address line.

The Primary_Secondary_Address field represents the complete, standardized primary and secondary address (no remainder data), and it is output as if the option Include Unused Address Line Data is set to No.

Invalid secondary standardization

When the input secondary address does not match with directory data, but SuiteLink finds a match to a different secondary address, the invalid input secondary address is output to various fields. The following list describes the factors that determine how the software outputs the invalid secondary address information:

- The type of information that it contains (for example, unit designator versus pound unit designator).
- The type of format that is used (discrete versus multiline).
- The location in the input address (on a separate line or on the address line).

2.11 Documentation

Several changes have been made to the documentation set that is provided with SAP BusinessObjects Data Services in recent releases. This list is a reminder of those changes.

- We have added a new *Master Guide*, which provides an overview of the application, its components, and scenarios for planning and designing your system landscape. Information about SAP BusinessObjects Information Steward is also provided in this guide.
- We have added a new *Workbench Guide*, which provides information about how to use the Workbench to migrate data and database schema information between different database systems.
- Release documentation:
 - The Release Summary has been renamed to What's New.
 - Resolved issues in the Release Notes have been moved to a separate Fixed Issues document.
- The Getting Started Guide and Advanced Development Guide have been removed and most of their content has been moved to the new Administrator's Guide. The Administrator's Guide contains information about administrative tasks such as monitoring, lifecycle management, security, and so on. It has also been added to the Technical Manuals, which is a compilation of the core Data Services documents.

Multi-user content in the *Advanced Development Guide* has been moved to the *Administrator's Guide* and *Designer Guide*.

- The Management Console: Metadata Reports Guide and Management Console: Administrator Guide have been combined into the Management Console Guide.
- The Installation Guide for Windows and Installation Guide for UNIX have been removed from the Technical Manuals. All post-installation, administrative content in the installation guides has been moved to the Administrator's Guide.
- The Integrator's Guide has been added to the Technical Manuals.
- The Salesforce.com Adapter Guide has been renamed to Supplement for Salesforce.com and added to the Technical Manuals.