BusinessObjects™ Data Quality XI for Informatica®

User's Guide

Data Quality XI for Informatica 11.5.1.0

January 2007

Contact information

Contact us on the Web at http://www.firstlogic.com/customer

If you find any problem with this documentation, please report it to Business Objects in writing at documentation@businessobjects.com.

Patents

Business Objects owns the following U.S. patents, which may cover products that are offered and sold by Business Objects: 5,555,403, 6,247,008 B1, 6,578,027 B2, 6,490,593 and 6,289,352.

Third-party contributors

Business Objects products in this release may contain redistributions of software licensed from third-party contributors. Some of these individual components may also be available under alternative licenses. A partial listing of third-party contributors that have requested or permitted acknowledgments, as well as required notices, can be found at: http://www.businessobjects.com/thirdparty

Contents

Preface

About this guide

This guide provides a general overview of BusinessObjects Data Quality XI for Informatica as well as specific information for installing and integrating this product into your Informatica environment. Some sections in this guide may not apply to your version of Data Quality XI for Informatica, depending on which product options you purchased.

Assumed audience and the content of this guide

This guide assumes that the people responsible for integrating Data Quality XI for Informatica are experienced users of the various Informatica applications including the PowerCenter Repository Manager, Designer, and Workflow Manager. It does not explain basic operation of the Informatica applications. For questions about the Informatica applications, consult the appropriate Informatica documentation or Informatica technical support.

This guide also does not explain the operation of the Data Quality server or Project Architect. For specific questions about Data Quality, consult the Data Quality documentation that is included in this package.

Complete documentation set

Your complete documentation set for Data Quality XI for Informatica includes the following:

- *Data Quality Administrator's Guide*
- *Data Quality International Address Processing*
- *Data Quality Integrator's Guide*
- *Data Quality Modifier's Guide*
- *Data Quality User's Guide*
- *Data Quality Release Notes*
- *Data Quality XI for Informatica Release Notes*
- This User's Guide

Access the latest documentation

You can access product documentation in several places:

- **On your computer.** Release notes, manuals, and other documents for each product that you have installed are available in the Documentation folder. Choose **Start** > **Programs** > **[product name or product family]** > **Documentation**.
- **On the Customer Portal.** Go to www.firstlogic.com/customer, and then click the Documentation link to access all the latest product documentation. You can view the PDFs online or save them to your computer for viewing or printing.

Conventions

This document follows these conventions:

Convention	Description
Bold	Highlights file names, paths, and text that you should type exactly as shown. For example, “Type **cd\dirs**.”
Italics	Both emphasizes text and indicates text for which you should substitute your own data or values. For example, “Type a name for your file, and the **.txt** extension (*testfile*.**txt**).”
> in menu commands	Indicates commands that you choose from menus in the following format: Menu Name > Command Name. For example, “Choose File > New.”
	Alerts you to important information and potential problems.
	Points out special cases that you should know about.
	Draws your attention to tips that may be useful to you.

About versions

Informatica versions

This version of Data Quality XI for Informatica is engineered for PowerCenter 7.1.2 and higher, or 8.1 and higher. It is *not* compatible with previous versions of Informatica.

Business Objects technology versions

Data Quality XI for Informatica uses BusinessObjects Data Quality XI version 11.5.1.0 to power its data quality processes.

Datatype support

Data Quality XI for Informatica supports only the string datatype.

Chapter 1: Overview

Before you use BusinessObjects Data Quality XI for Informatica, you need to understand some of the concepts behind data quality and address cleansing. A basic familiarity with these concepts will help you to understand what Data Quality XI for Informatica does and why your organization benefits from data quality.

What is Data Quality XI for Informatica?

Data Quality XI for Informatica is designed to ensure information quality within Informatica PowerCenter installations. With the capabilities provided by Data Quality, it can provide batch verification and correction of name and address data, and can include sophisticated processes to identify matching records.

Data quality functions

Data Quality XI for Informatica brings name and address integrity to your Informatica environment through preset transformations that your users can easily integrate into standard Informatica extract, transform, and load (ETL) mapping.

Name cleansing

The name cleansing capability, powered by Data Quality's Data Cleanse transform, uses extensive name dictionaries to find name and firm (business name) information and, if necessary, add or fix that information from your source data. For example, based on the name dictionary information, you can add to your source data other name-related components, such as gender, titles, and so on.

Name cleansing capability is included with Data Quality XI for Informatica.

Address cleansing

Data Quality XI for Informatica performs standardization and correction of U.S. and international addresses. This capability is powered by Data Quality's Address Cleanse transform. You can purchase different international engines for Data Quality. The engines provide a different level of address-checking for certain countries. For example, if you own the Canadian engine, Canadian addresses are verified down to the street address level. With the international engine, some countries' addresses are only verified to the city level. International address cleansing may not be available in your version of Data Quality XI for Informatica, depending on which product options you purchased.

Matching function

The matching function of Data Quality XI for Informatica brings sophisticated record-matching technology to your Informatica environment. This transformation is powered by Data Quality's matching functions transform. Use this matching capability in concert with address cleansing and name cleansing to improve the quality of your data by identifying matching or duplicate records.

Note: If a Data Quality dataflow contains a match transform, you must use an active Data Quality XI for Informatica transform in your mapping.

Flexible configuration

You can integrate the data cleansing and matching functions with relatively minor configuration. In addition, your Informatica users can further fine tune their results of Data Quality XI for Informatica's operations by making changes to the Data Quality projects.

Introduction to Data Quality

As a Data Quality XI for Informatica user, you may be wondering what Data Quality is and what it means to you. After installation and configuration, you can use Data Quality XI for Informatica with little thought given to the Data Quality component, because it handles the data quality processes behind the scenes. The version of Data Quality shipped as part of Data Quality XI for Informatica is customized to run only as part of Data Quality XI for Informatica, but the technology of Data Quality offers other potential solutions to your enterprise's data quality needs. This section provides an overview of what Data Quality does, independent of its use within Data Quality XI for Informatica, and tells you where you can find more information about Data Quality.

Data quality hub

Data Quality is the next generation data quality platform that allows you to tie in all of your data quality needs to a single "hub." Many users of Data Quality XI for Informatica also have data quality needs in other applications or other areas of their enterprise. Instead of having many disparate solutions for each area of your enterprise, you can use Data Quality as the backbone for all of your data quality needs. With Data Quality, you can centralize and standardize this processing through the following Data Quality features.

Centralized business rule repository

All of the business rule definitions are stored in a single repository that can be shared by all applications that use Data Quality. You can organize business rules into reusable objects that can be shared to meet your enterprise requirements.

You can use the same objects in a project for Data Quality XI for Informatica, your own custom application that incorporates Data Quality, or direct batch processing through Data Quality itself. This means that you can create data quality standards that are truly shared across your entire enterprise.

Service-oriented architecture

Data Quality is built on service-oriented architecture principles. That means that Data Quality is a data quality service that serves all of your other applications.

You can set up any number of instances of the Data Quality server software, and your applications (including Data Quality XI for Informatica) can communicate with that server directly via a socket connection, or via a web server using the Data Quality web service. This means that all of your data quality processing can happen in one centralized area. Data Quality XI for Informatica communicates with Data Quality via a socket connection.

Data Quality scenario

The following diagram shows a possible scenario for how Data Quality can be used as the central hub for all of your enterprise data quality needs. This is only one example; the number of scenarios is practically limitless because of Data Quality's flexibility.

More information about Data Quality

For more information about Data Quality, see the Data Quality documentation included in this shipment. The Data Quality documentation is available in electronic format, both in a printable PDF file and an online help format.

The online version requires an internet browser. We recommend Internet Explorer or Netscape (most recent versions recommended), or Mozilla 1.4 or higher, though other browsers may work as well. The printable version requires Adobe® Reader® 4.0 or higher (6.x recommended).

Documentation is a separate install component. The documentation for Data Quality is an optional installation component in most cases. If you don't install the documentation while installing Data Quality, it won't be available on your system.

Data Quality documentation descriptions

The following table describes the types of information included with each of the documents installed with Data Quality.

Document	Description
Data Quality User's Guide	Includes conceptual information about Data Quality, details about using the Project Architect interface, and a detailed reference for objects such as transforms and Data Quality Blueprints™. This guide is intended for the user who creates projects (dataflows) and configures the business rules of your Data Quality processing.
Data Quality Integrator's Guide	Includes information about integrating Data Quality functionality into custom applications via the web service. For example, it includes a detailed web method reference. This guide is intended for programmers and integrators.

Document	Description
Data Quality System Administrator's Guide	Includes information about system requirements, installation planning, and actual installation steps for Data Quality.
Data Quality Release Notes	Includes information about what is new in the latest Data Quality release.
Data Quality International Address Processing	The full title of this document is *Data Quality International Address Processing: Installing the Runtime Environment on UNIX*. This document is available only on UNIX platforms, and applies only to users of the International engine within the Address cleanse transform. This document explains how to install the runtime environment necessary for this scenario.
Data Quality Data Cleanse Modifier's Guide	Includes information about how to customize cleansing with the Data Cleanse transform. This guide is intended for advanced users of the Data Cleanse transform.

Access Data Quality documentation on Windows

After installing Data Quality, Windows users can access the documentation from shortcuts on the Start menu. To access the documentation:

1. Choose **Start > Programs > BusinessObjects XI Release 2 > Data Quality 11.5 > Documentation**.
2. Click the appropriate shortcut for the documentation you want to view.

Access Data Quality documentation on UNIX

After installing Data Quality, UNIX users can access the online documentation from their web browser. Enter the appropriate location and file name for the document you want to view, as shown in this table.

Document	Location and file name
Data Quality User's Guide	*install_location*/**documentation/userguide_html/index.htm**
Data Quality Integrator's Guide	*install_location*/**documentation/intguide_html/index.htm**
Data Quality Data Cleanse Modifier's Guide	*install_location*/**documentation/modguide_html/index.htm**

UNIX users can access the printable (PDF) documentation from Adobe Reader. You can find these documents in *install_location*/**documents/pdf/**.

Chapter 2: Installation

Before you can integrate BusinessObjects Data Quality XI for Informatica into your application, you must install BusinessObjects Data Quality XI software, copy dataflows to your Data Quality installation, and install the Data Quality XI for Informatica plug-in to your Informatica server and client machines.

Be sure to follow the instructions appropriate for your operating system.

System requirements

This section explains the requirements for installing the Data Quality XI for Informatica plug-in to operate with your Informatica PowerCenter environment.

System requirements for installing the Data Quality portion of Data Quality XI for Informatica can be found in the *Data Quality System Administrator's Guide*, which is included in this software package.

Requirements for Data Quality XI for Informatica

Installation of the Data Quality XI for Informatica plug-in requires that you have the following software and hardware specifications before installing Data Quality XI for Informatica.

Informatica versions

This version of Data Quality XI for Informatica is engineered for PowerCenter 7.1.2 and higher, or 8.1 and higher. It is *not* compatible with previous versions of Informatica.

Operating systems

Here is a list of the operating systems that are supported in this release.

- Windows 2000, 2003, and XP (32-bit)
- AIX 5.2 and 5.3 (32-bit)
- HP-UX 11i (32-bit)
- Solaris 8.0, 9.0, and 10.0 (32-bit)
- Red Hat Linux Enterprise Server 3.0 (32-bit, and for PowerCenter 8.*x* only)

The installation routine you run is specifically designed for your operating system.

Processor requirements

Running Data Quality XI for Informatica requires that you have a processor compatible with your operating system and which has a minimum clock speed of 800 MHz.

Installation

Before you can integrate Data Quality XI for Informatica into your application, you must first install the Data Quality Server and Project Architect. After you've installed Data Quality, you can then install the Data Quality XI for Informatica plug-in to your Informatica server and client machines.

Follow the instructions appropriate for your operating system:

- "Install Data Quality XI for Informatica (Windows and UNIX)" on page 16
- "Install the Data Quality XI for Informatica on UNIX" on page 29

About installation directories

Use these two variables when you enter paths for the Informatica server and the client: *$PMRootDir* and *Client_Install_Dir*.

$PMRootDir

In these installation steps, procedures may refer to the Informatica server variable *$PMRootDir*. This User's Guide uses *$PMRootDir* as an abbreviation for the installation directory (home) of your PowerCenter Server program.

For example, assume that your PowerCenter Server is installed in **informatica_server**. When the instructions refer to "Enter *$PMRootDir***BusinessObjects**\," you enter **informatica_server\BusinessObjects** (Windows) or /**informatica_server/businessobjects** (UNIX).

If you are entering a path within Informatica or the Data Quality interface, you can simply enter the *$PMRootDir* variable rather than the whole path.

Client_Install_Dir

Client_Install_Dir is an abbreviation we use to refer to the installation directory (home) of your PowerCenter client. The client is only installed on Windows.

Install Data Quality XI for Informatica (Windows and UNIX)

These installation instructions are applicable to Windows and UNIX. However, there are additional steps to perform if you are a UNIX user. Therefore, UNIX users should read “Install the Data Quality XI for Informatica on UNIX” on page 29 before proceeding with the steps in this section.

Though this section applies to both Windows and UNIX installations, the instructions favor Windows terminology. For example, path names contain the backward slash (**dir\file**), rather than the forward slash (**dir/file**) and show the GUI used for a Windows installation. UNIX users should substitute appropriate file names and UNIX conventions when following these instructions.

The total number of CDs that you receive with your purchase of Data Quality XI for Informatica is based on the country-specific address cleansing transforms that you purchased. The CDs for Data Quality XI for Informatica include:

- “BusinessObjects Data Quality XI Release 2” CD and the Auxiliary Directories CDs contain the installation of the Data Quality server and the extensive lookup directories used for address cleansing. The exact CD set that you receive depends on your purchase. See the Data Quality documentation for installation instructions.
- “BusinessObjects Data Quality XI Packages” CD contains additional files to install with Data Quality. See the Data Quality documentation for installation instructions.
- “BusinessObjects Data Quality XI Release 2 - Windows Components for UNIX” CD (UNIX only) contains software for UNIX customers.
- “BusinessObjects Data Quality XI for Informatica – Dataflows” CD contains the dataflows that you must copy to your Data Quality installation.
- “BusinessObjects Data Quality XI for Informatica” CD contains the installation of the plug-in for your Informatica server.
- “BusinessObjects Data Quality XI for Informatica – Windows Client” CD contains the plug-in for your Informatica clients.

Important: You must install and set up the Data Quality server before you install the Data Quality XI for Informatica plug-in.

The steps

There are several steps involved in installing and setting up Data Quality XI for Informatica in your Informatica environment:

Windows 2003 installation

For Windows 2003, you must install both Informatica PowerCenter and the Data Quality XI for Informatica plug-in with either the same user account or with accounts that have the same permissions. In addition, the PowerCenter Server service must be started using an account that has permission to all files in both the PowerCenter installation and the Data Quality XI for Informatica plug-in installation. This ensures that all users can write to the files and folders necessary to run Data Quality XI for Informatica.

Step 1: Install Data Quality and copy the dataflows

First, install the Data Quality Server and Project Architect following instructions in the *Data Quality System Administrator's Guide*. Then return to this User's Guide and perform the required manual steps described at the end of this section.

Data Quality installation notes

You can find the instructions for installing, configuring, and running the Data Quality portion of Data Quality XI for Informatica in the Data Quality documentation you received in this package. However, using Data Quality as a component of Data Quality XI for Informatica presents some special circumstances not covered in the Data Quality documentation.

This section describes some additional issues unique to installing Data Quality as part of Data Quality XI for Informatica.

Install Data Quality first

You must install Data Quality before installing the Data Quality XI for Informatica plug-in. After you have installed and configured Data Quality, you can proceed to "Step 2: Install the Data Quality XI for Informatica plug-in" on page 19.

Install Data Quality on a different machine than Informatica

We recommend you install Data Quality on a different machine than the one that contains your installation of Informatica PowerCenter. Information about the system requirements for the machine on which you install Data Quality are included with the Data Quality installation instructions.

Data Quality previously installed

If you are already an Data Quality customer (prior to the purchase and installation of Data Quality XI for Informatica) and therefore have already installed a version of Data Quality that you want to use with Data Quality XI for Informatica, you must proceed with different steps than you would if you were performing a fresh installation of Data Quality as part of the installation of this product. See the Data Quality documentation for information about upgrading Data Quality.

Copy dataflows and transforms to Data Quality installation

The "BusinessObjects Data Quality XI for Informatica – Dataflows" CD that ships with the Data Quality XI for Informatica contains dataflows and transforms in the form of **.xml** files that are specific to the functionality of the Data Quality XI for Informatica. You must copy the dataflows and transforms from this CD to your Data Quality installation.

To copy the dataflows and transforms to your Data Quality installation, perform this single step:

From the "BusinessObjects Data Quality XI for Informatica – Dataflows" CD, copy the folder **repository** (and all the folders and files within) to the folder ***DataQualityinstall***\, where ***DataQualityinstall*** is the location of your Data Quality installation (by default, **c:\dqxi\11_5**).

Step 2: Install the Data Quality XI for Informatica plug-in

The next installation step is to install the Data Quality XI for Informatica plug-in to your Informatica Server and client machine(s).

Install server plug-in to your Informatica Server

To install the server plug-in to your Informatica Server:

1. Insert the "BusinessObjects Data Quality XI for Informatica" CD into your CDROM drive. Choose **Start > Run** and enter *x*:**Server_setup.exe** (where *x* is the letter of your CDROM drive on Windows). The Welcome screen appears.
2. Click **Next**. The License Agreement window appears.
3. Accept the license agreement and click **Next**. The Ready to Install window appears.
4. Click **Install** to install the plug-in.
5. When the installation is complete, you are prompted to restart your computer. You don't have to restart it now, but you must restart your computer before you can actually run Data Quality XI for Informatica.

UNIX install: These instructions are for Windows installations of the server plug-in. To install the plug-in on UNIX, run the executable from the CD and follow the intructions provided by the command prompt. Also, if you are performing a UNIX install, proceed to "Add to environment variables" on page 29 and "Install C++ runtime upgrade (Linux and HP-UX)" on page 30 before continuing with "Step 3: Register the DQ_XI_plugin.xml file" on page 20.

Install client plug-in to your Informatica client(s)

To install the client plug-in to your Informatica client(s):

1. Insert the "BusinessObjects Data Quality XI for Informatica – Windows Client" CD into your CDROM drive. Choose **Start > Run** and enter *x*:**Client_setup.exe** (where *x* is the letter of your CDROM drive on Windows).

   The Welcome screen appears.
2. Click **Next**. The License Agreement window appears.
3. Accept the license agreement and click **Next**. The Choose Destination Location window appears.
4. Click **Change** to browse to a destination folder where you want the samples and documentation files to be installed, then click **Next**.

   The Ready to Install window appears.
5. Click **Install** to install the plug-in.
6. When the installation is complete, you are prompted to restart your computer. You don't have to restart it now, but you must restart your computer before you can actually run Data Quality XI for Informatica.

Step 3: Register the DQ_XI_plugin.xml file

Before you can use Data Quality XI for Informatica's custom transformation, you must register it as a plug-in in the repository, as you would with any plug-in for Informatica PowerCenter.

For more information about registering plug-ins, see your Informatica documentation.

The procedure for registering the plug-in is different for 7.*x* and 8.*x* versions of Informatica PowerCenter.

Register to the repository (Informatica 7.*x*)

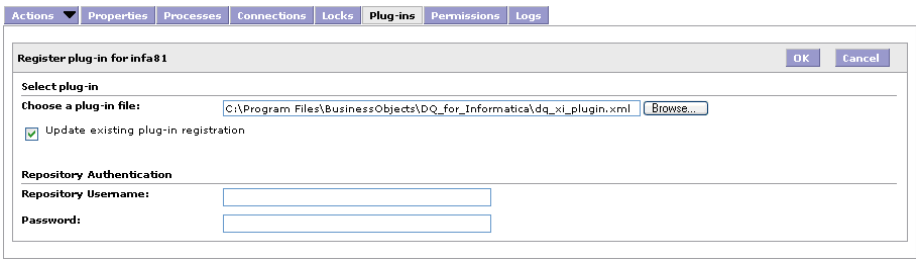
To register the plug-in to the repository on Informatica 7.*x*:

1. Manually copy the **DQ_XI_plugin.xml** file from **c:\Program Files\BusinessObjects\DQ_for_Informatica** to the **RepositoryServer\bin\Plugin** directory.
2. Make sure that no application is connected to the Repository Server, including PM Server, Designer, Workflow Manager, Workflow Monitor, and Repository Manager.
3. In your Repository Server Administration Console, connect to the repository server.
4. Click the **Available Packages** node in the left pane. The **DQ_XI_plugin.xml** plug-in is listed in the right pane.
5. Right-click on the **DQ_XI_plugin.xml** file and click **Register**. The Register Plug-in window opens.
6. Select the repository in which you want to register the plug-in, enter your user name and password, and click **OK**.

Register to the repository (Informatica 8.*x*)

To register the plug-in to the repository on Informatica 8.*x*:

1. In the Informatica PowerCenter Administration Console, place the repository in "exclusive" mode. See the Informatica documentation for details on this step.
2. In the plug-ins tab for this node, click **Register Repository Server plug-in**. The Choose file window appears.
3. Browse to the location of the **DQ_XI_plugin.xml** plug-in. Select the plug-in, and click **Open**. The Console displays a screen to register the plug-in.

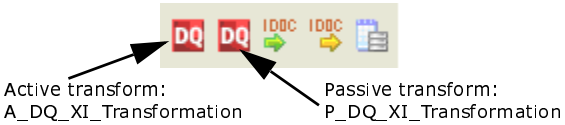


4. Enter the **Repository Username** and **Password**. Check the box for **Update existing plug-in registration** if you have previously registered this plug-in. Click **OK**.

5. The Console indicates that it is importing the plug-in. Upon completion, place the repository back into normal mode.

Step 4: Verify the Data Quality XI for Informatica toolbar

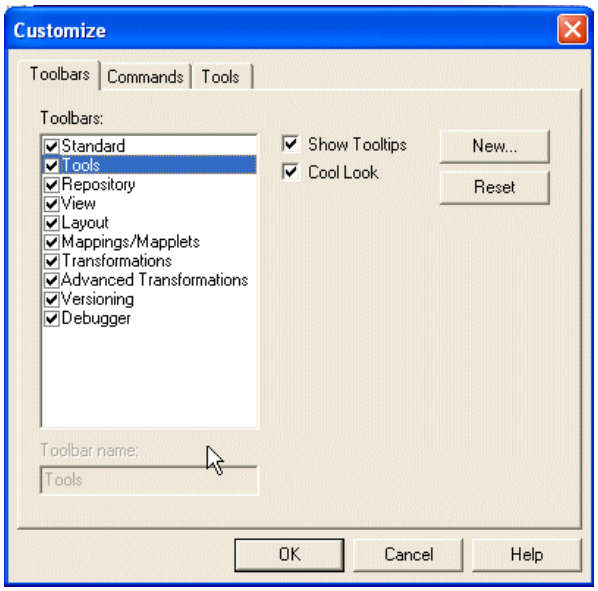
After you install the client software, the Data Quality XI for Informatica toolbar should appear in your PowerCenter Designer application.



There are two toolbar items for Data Quality XI for Informatica: one for an active transform and one for a passive transform.

If the toolbar does not appear, perform the following steps:

1. In the PowerCenter Designer, open the appropriate repository.
2. Select **Tools > Customize**. The Customize window opens.



3. Select the **Transformations** and the **Advanced Transformations** options.
4. Click **Ok**.

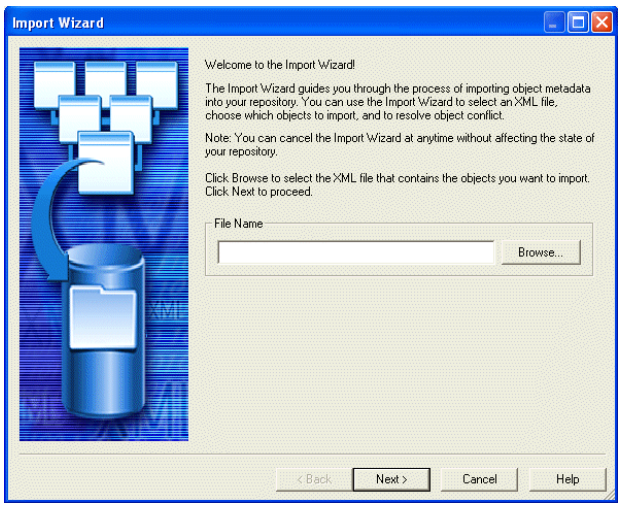
Step 5: Import and set up the sample mapping

The sample mapping is installed as an XML object. It includes sources, targets, and transformations. You must import and set up the sample mapping for use with Data Quality XI for Informatica.

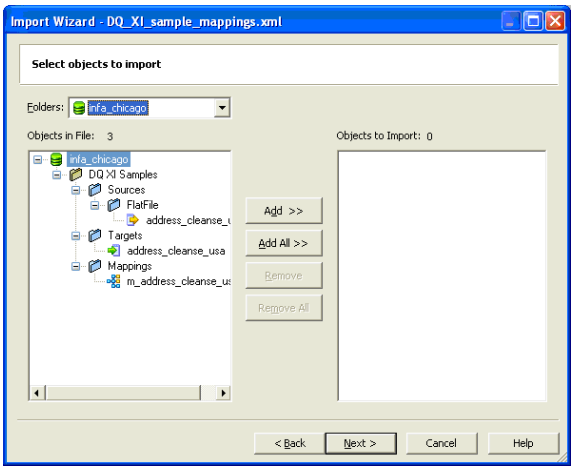
Import the sample mapping

To import the sample mapping, follow the normal steps (listed below) for importing an object.

1. In PowerCenter Repository Manager, choose **Folder > Create**. The Create Folder window opens.
2. In the Name box, enter **DQ XI Samples**.
3. Set the appropriate owners, groups, and permissions for the folder and click **OK**.
4. Choose **Repository > Exit**.
5. In PowerCenter Designer, right-click the DQ XI Samples folder and choose **Open**.
6. Choose **Repository > Import Objects**. The Import Wizard window opens.



7. Click Browse to locate and select the **DQ_XI_sample_mappings.xml** file and click **OK**. By default, the sample mappings are installed in **c:\Program Files\BusinessObjects\DQ_for_Informatica**.
8. Click **Next**. The wizard's Select Objects to Import window opens.



9. Click **Add All** to import all sample sources, targets, transformations, and mappings, and click **Next**. The wizard's Resolve Conflicts in Folder window opens.
10. Resolve any conflicts, as necessary. When finished, click **Close**.
11. In PowerCenter Designer, choose **Repository > Save**.

Set up the sample mapping

You set up the sample mapping by configuring the DQ XI Properties in the Edit Transformations window.

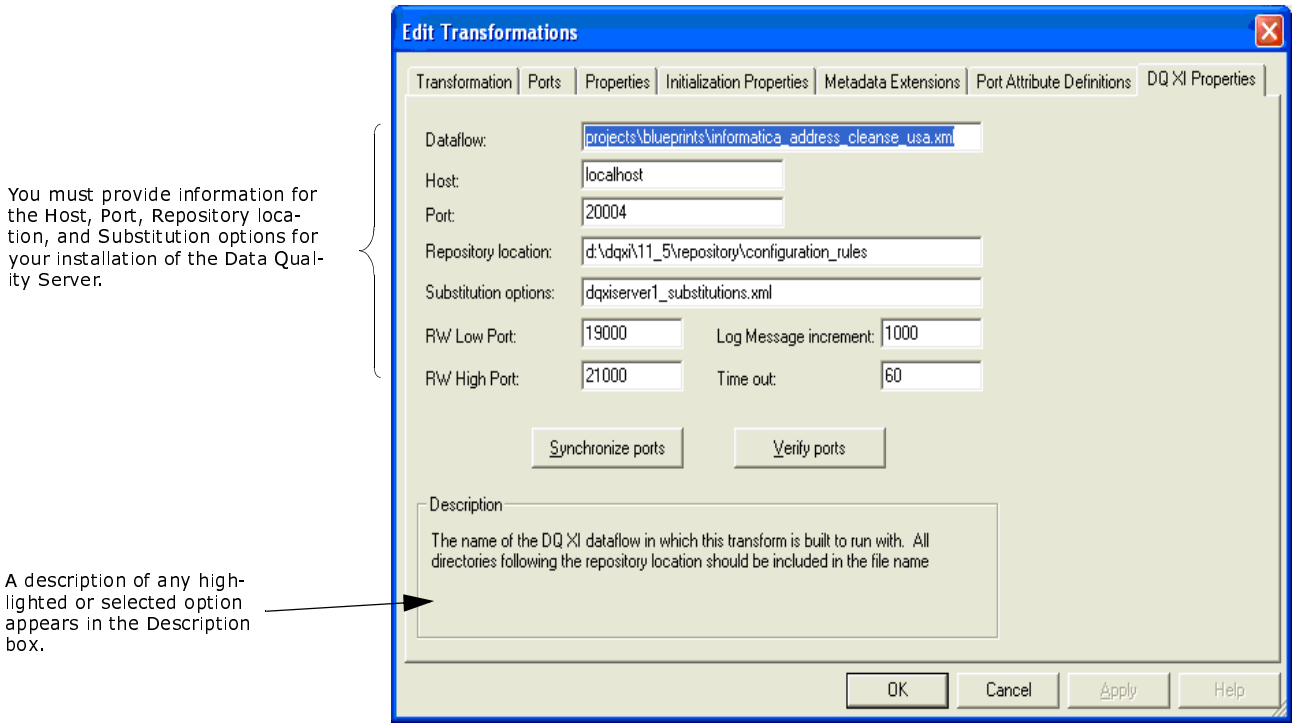
To configure the set up the sample mapping:

1. In the **Mapping Designer**, open the mapping m_address_cleanse_usa.
2. Right-click Edit.

   The **Edit Transformations** window opens.

Tip: The first time that you open the Edit Transformation window, select the corner of the window and drag it so that you can view all of the tabs.

3. Click the **DQ XI Properties** tab.



5. Set the **properties** as necessary. See the table in the following section to understand which properties are required or optional.
6. Click **Verify ports** to check that the fields in the specified Data Quality dataflow correspond to the input and output ports on the Informatica transform. The Data Quality Server must be running for the verification to be successful.
7. Click **Apply** and then **Ok**.

Option descriptions

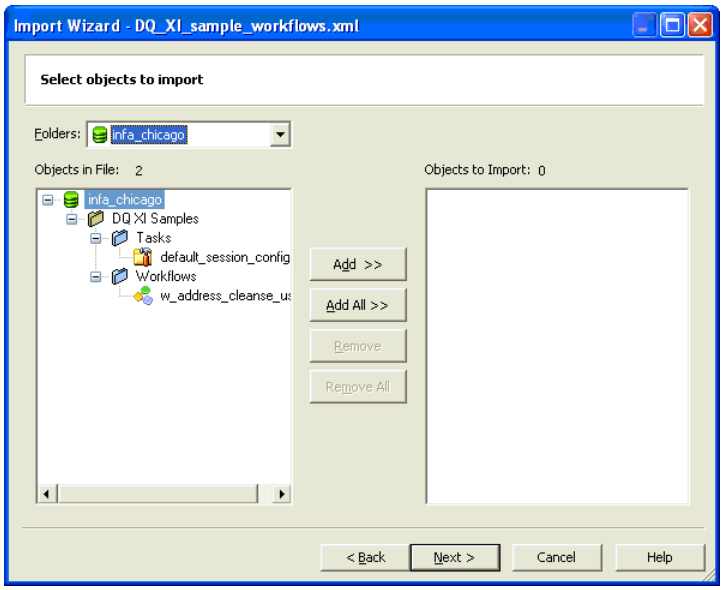
This table describes each of the properties available for the Data Quality XI for Informatica custom transformations. Descriptions also appear in the Description box in the Edit Transformation window.

Option	Description
Dataflow	The name of the Data Quality dataflow in which this transform is built to run. All directories following the repository location should be included in the file name
Host	The host name of the machine that the Data Quality server is installed on.
Port	The port number that the Data Quality server is listening on.
Repository location	The location of the repository in which your dataflow and all supporting business rules are located.
Substitution options	The file name of the optional substitution options file for your Data Quality server. This should be specified as a file name only.

Option	Description
RW Low Port RW High Port	The lowest and highest port number that a socket listener should attempt to listen on. This is on your Informatica server machine.
Log Message increment	The number of records between each request to the Data Quality server for updating log or error messages.
Time out	The number of seconds to wait for a connection to be established, or for data to be sent/received.
Synchronize ports	Click to synchronize the input and output ports. Data Quality XI for Informatica prepends an I_ before the input port names and an O_ before the output port names.
Verify ports	Click to check that the fields in the specified Data Quality dataflow correspond to the input and output ports on the Informatica transform.

Step 6: Import the sample workflow

1. Open PowerCenter Workflow Manager and right-click the DQ XI Samples folder. Click **Open**.
2. Choose **Repository > Import Objects**. The Import Wizard window opens.
3. Click **Browse** to locate and select the **DQ_XI_sample_workflows.xml** file (located by default in *Client_Install_Dir***samples**), which contains the sample sessions). Click **OK**.
4. Click **Next**. The wizard's Select Objects to Import window opens.



5. Click **Add All** to import all sample session and workflow. Click **Next**.

   The wizard's Conflict Resolution Wizard window opens. You must resolve each mapping with the mappings that you imported in PowerCenter Designer.
6. Resolve any conflicts, as necessary. When finished, click **Close**.
7. In PowerCenter Designer, choose **Repository > Save**.

Step 7: Verify the installation

After installing the software, we recommend that you set up and run the sample mapping, wf_address_cleanse_usa, to verify that the installation process was successfully completed. First you must shutdown and restart your computer if you haven't already done so during the installation procedure.

1. In PowerCenter Workflow Manager, connect your server to your repository.
2. Run the mapping wf_address_cleanse_usa using your PowerCenter Workflow Manager. If the session completes successfully, Data Quality XI for Informatica was properly installed for that module.

   If the session does not properly complete, verify that you've followed all of the installation steps.

Install the Data Quality XI for Informatica on UNIX

The installation procedures for UNIX are similar to the procedures for Windows. For installation steps that apply to both Windows and UNIX, see "Install Data Quality XI for Informatica (Windows and UNIX)" on page 16. Read this section for any additional installation steps, or for information that relates directly to your UNIX installation.

Installation directories

See "About installation directories" on page 15 to learn about directory variables that you can use when entering the default location for Informatica Server and PowerCenter.

Permissions

The owner of your Informatica Server process should perform the installation if possible. Otherwise, must go back and set appropriate permissions to the Data Quality XI for Informatica files and directories after the installation.

Important: Before running the Informatica Server process (PMServer), you should change your default permissions for your Informatica user to **rw-rw-r--**. When you run the server process, certain files are overwritten and the server process owner becomes the owner of those files. Using **rw-rw-r--** permissions ensures that people in the group will still have write access to these files.

After you perform "Step 4: Verify the Data Quality XI for Informatica toolbar" on page 22, you must add to your environment variables and install the Linux C++ runtime upgrade.

Add to environment variables

Add or edit the following environment variables in your **.login** or **.profile** file (in the Bourne shell), or in the **.login** or **.cshrc** file (in the C shell).

Note: If you also use a stand-alone (batch) Business Objects or Firstlogic application on the same machine, you may encounter conflicts in some situations. To avoid this conflict, we recommend that you set up different user profiles: one for Data Quality XI for Informatica and one for the stand-alone application.

For the user who launches the PowerCenter Server

Each user who launches the PowerCenter Server must alter some environment variables in their individual **.login** or **.profile** file (in the Bourne shell), or in the **.login** or **.cshrc** file (in the C shell).

1. Add *$PMRootDir*/**businessobjects** to the LD_LIBRARY_PATH (for Solaris and Linux) or SHLIB_PATH (for HP-UX) or LIBPATH (for AIX) environment variable.
2. Restart the Informatica Server for the variables to take effect.

Install C++ runtime upgrade (Linux and HP-UX)

If you use Red Hat Linux, you must install the GNU Standard C++ Library, version 3.2.3 (for **libstdc++.so.5** support). We have found a technical issue requiring this new C++ runtime support. To learn more, go to http://www.redhat.com/apps/download and search for the keyword "libstdc++." Be sure to review Red Hat's specific instructions in the System Environment Libraries section related to dependencies. If you need further assistance, contact Red Hat.

If you use HP-UX, you must install the HP-UX aC++ runtime environment version A.03.26 or later. To learn more and download this runtime environment, go to http://docs.hp.com/en/5187-2783/ch15s03.html.

Continue with "Step 5: Import and set up the sample mapping" on page 23 to complete your installation.

If you also use a stand-alone Business Objects (Firstlogic) product

Some users of Data Quality XI for Informatica also use a stand-alone (batch) Business Objects (Firstlogic) application. If Data Quality XI for Informatica and the stand-alone application are on the same machine, you may encounter conflicts in some situations.

What causes the conflict?

Your PATH and PW_PATH environment variables point to locations of your Business Objects products. If you try to run Data Quality XI for Informatica but your environment variable points to the stand-alone application location first, Data Quality XI for Informatica will try to use the software installed for the stand-alone application, and vice versa.

Different user profiles

To avoid this conflict, Business Objects recommends that you set up different user profiles: one for Data Quality XI for Informatica and one for the stand-alone application. The environment variables in Data Quality XI for Informatica user profile should point to the location of the Data Quality XI for Informatica software. For UNIX, this location is specified in "Add to environment variables" on page 29. For Windows, the environment variables are automatically set up during installation.

The environment variables in the stand-alone application user profile should point to the location of the stand-alone application.

Scripts on UNIX

If you run Data Quality XI for Informatica on a UNIX system, you can run Data Quality XI for Informatica and the stand-alone application with the same user profile. To do this, your login script must initially set the environment variables as appropriate for Data Quality XI for Informatica. You must also create a script (or enter the appropriate command) that resets your environment variables to the settings required by the stand-alone application.

1. Login with the profile set up for Data Quality XI for Informatica.
2. Start the Informatica Server process.
3. Run the script (or enter the appropriate commands) to set your environment variables for your stand-alone application. For example:

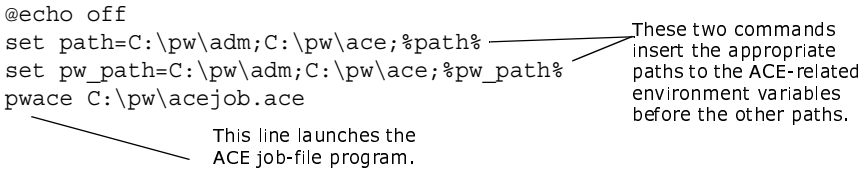
   ```
   setenv PATH /postware/ace
   setenv PW_PATH $PATH
   ```

4. Run your stand-alone application, as needed.
5. Repeat these steps each time you log on or when you want to restart the Informatica Server process.

Batch files on Windows

On Windows, you can use batch files to run the stand-alone application and Data Quality XI for Informatica from the same system. Follow these steps:

1. Create a batch file that sets the proper environment variables for your stand-alone application and launches the stand-alone application. The following example shows a batch file for launching an ACE job file.



2. Create additional batch files for each stand-alone application you want to run.

When you use this procedure, the application you run from the batch file uses the environment you set up in the batch file. All other applications you run use the normal environment variables set up for the user who is logged in.

A command-prompt window remains open until you close the application that you launched from the batch file.

You may want to create a shortcut to the batch files and place the shortcut on your desktop or in your Start menu. You can then double-click the shortcut to run the batch file and launch the stand-alone application.

Chapter 3:
Add and use Data Quality XI for Informatica transformations

After you've installed the BusinessObjects Data Quality XI for Informatica software, there are still other steps necessary before you can use and run Informatica processes that use the Data Quality XI for Informatica.

Read this chapter to get an idea of which type of transformation, active or passive, is right for your task. Then refer to the appropriate sections of this chapter for information about setting up and using a transformation.

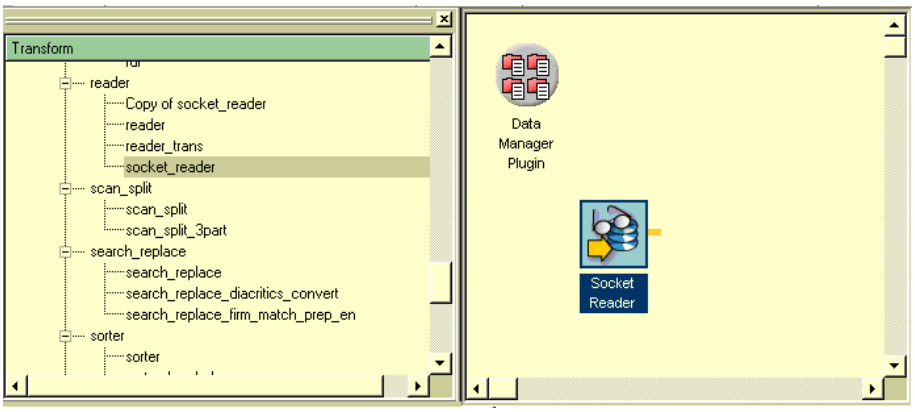
Create Data Quality dataflows

Before you add and use new transformations in Informatica, you must create Data Quality dataflows to work with the Informatica transformations you use. General information about creating a Data Quality dataflow can be found in the Data Quality documentation.

This section describes issues unique to creating a dataflow to work with Data Quality XI for Informatica.

Socket reader/writer

A typical Data Quality dataflow uses a Reader transform and a Writer transform to input data from a data source and output data back to that data source. However, Data Quality XI for Informatica requires that the dataflow instead use the Socket Reader and Socket Writer transforms, which are provided with Data Quality XI for Informatica.



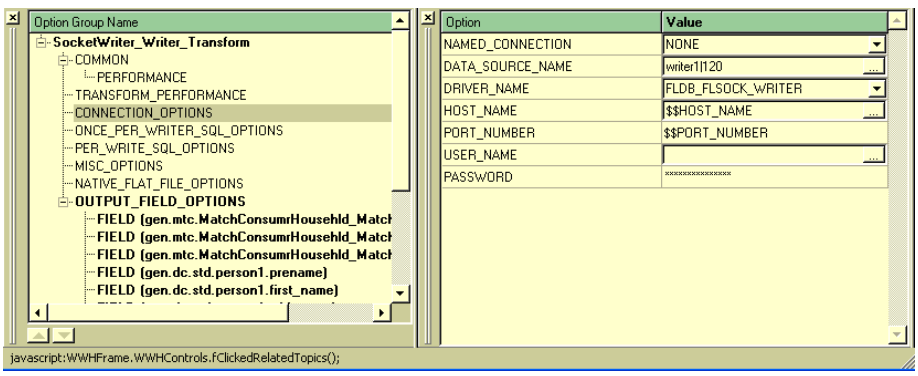
Unlike typical Data Quality Reader and Writer transforms, the Socket Reader and Socket Writer require that input and output data be of the string datatype. Use of any other data type causes an error.

Socket reader/writer reference

As you create Data Quality dataflows with the Socket Reader and Socket Writer transforms, you must understand the values of the Connection_Options settings and Data_Source_ID field.

Connection_Options

The Connection_Options group includes information about accessing the input or output source.



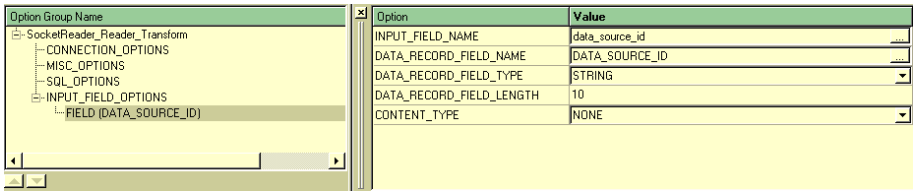
The following table defines the options of the Connection_Options group, and the value required by the Socket Reader and Socket Writer.

Option	Description
Data_Source_ Name	Specifies the name and path to the data source. For the Socket Reader and Socket Writer, the value must be in the form of *writer_name*|*timout_in_seconds* (for example, **writer1|120**).
Driver_Name	Specifies the database driver to use when connecting from the Reader or Writer to the data source. For the Socket Reader, the value must be **FLDB_FLSOCK_READER**. For the Socket Writer, the value must be **FLDB_FLSOCK_WRITER**.
Host_Name	Specifies the server name where the data source resides. For the Socket Reader and Socket writer, use the substitution variable **$$HOST_NAME**.
Named_Connection	Specifies the name of a pre-configured data source connection to either a relational database or a flat file. For the Socket Reader and Socket Writer, no value is required.
Password	This option is not used with the Socket Reader and Socket Writer.
Port_Number	Specifies the port number used for connections. For the Socket Reader and Socket Writer, use the substitution variable **$$PORT_NUMBER**. **NOTE:** Because of a known issue with the Data Quality Project Architect, the use of this substitution variable causes the Project Architect to fail to verify the dataflow. You should ignore this error.
User_Name	This option is not used with the Socket Reader and Socket Writer.

Data_Source_ID field

In a typical Data Quality Reader or Writer transform, the Data_Source_ID field adds a new column and value that you specify before the data passes to the next transform. You would use this field if you are processing more than one Reader or Writer transform in a dataflow and you want to identify which Reader or Writer the record came from at the end of processing.

With the Socket Reader and Socket Writer, this field exists, but may have no use if you always use a single data source, Informatica. In such a scenario, this field does not map to any Informatica field, and when processed, Informatica flags it as an unmapped field. To avoid this error, you should default the Data_Source_ID field to some value.



Different types of Data Quality XI for Informatica transformations

Before incorporating Data Quality XI for Informatica transformations into your Informatica mappings, you must decide which type of transformation to use.


If you use a Data Quality dataflow that includes a match transform, you must use active transforms.

This section will be helpful when making your decision between the two transformation types.

Passive custom transformation

For flexibility in mappings and ease of setup, Data Quality XI for Informatica's passive custom transformation is the best choice. We recommend that you use this type of transformation if you are not using Data Quality's match capability and if you are new to Data Quality XI for Informatica.

A passive transformation outputs one row for every input row.

See "Create a Data Quality XI for Informatica transformation" on page 37 for more information about adding this transformation.

Active custom transformation

For performance and to use match capabilities, the active Data Quality XI for Informatica custom transformation is the better choice.

See "Create a Data Quality XI for Informatica transformation" on page 37 for more information about adding this transformation.

Reusable transformations

You can reuse a transformation if you create it in the Transformation Developer by using an *instance* of that transformation in mappings.

If you make changes in the main transformation, the changes are inherited by each instance of that transformation. You cannot make any changes within an instance of the transformation.

Nonreusable transformations

When you create a transformation within a mapping using the Mapping Developer, it is not reusable in other mappings, but it is modifiable.

You can make nonreusable transformations reusable by selecting the Reusable check box in the MetaData Extensions tab of the Transformation Developer. When you do this, Data Quality XI for Informatica places a reusable version of the transformation in the Transformations folder and makes your original transformation a non-modifiable instance.

If you want to fine-tune an instance of Data Quality XI for Informatica's transformations for the mapping process into which it is incorporated, you should create the transformation within the mapping.

Create a Data Quality XI for Informatica transformation

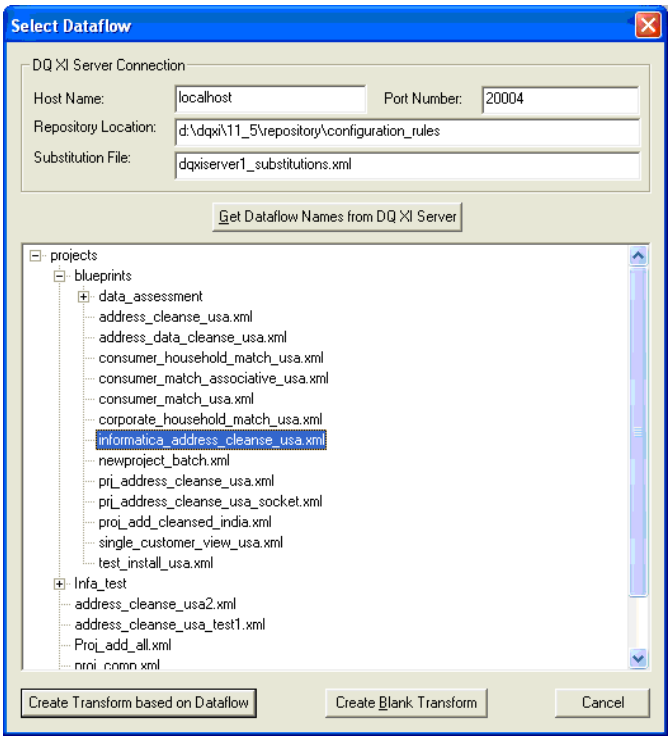
This section explains the steps involved in creating a custom transformation. For specific information about Informatica controls and program operation, see your Informatica documentation.

Step 1: Create a new transformation

1. Open **Informatica PowerCenter Designer** and the appropriate repository.
2. Open the **folder** that you want to work in.
3. Choose **Tools > Transformation Developer**.
4. On the Data Quality XI for Informatica **toolbar**, click the custom transformation that you want to create.

   If the toolbar is not displayed, see "Step 4: Verify the Data Quality XI for Informatica toolbar" on page 22 (for Windows) or "Step 4: Verify the Data Quality XI for Informatica toolbar" on page 22 (for UNIX).

5. Click the area of the **Transformation Developer workspace** where you want to place the transformation. The **Select Dataflow** window opens.



6. Make any necessary changes to the **Host name**, **Port Number**, **Repository Location**, and **Substitution File** fields.

7. To create a blank transformation, click **Create Blank Transform**.

   or

   To create a transformation based on an existing dataflow, click **Get Dataflow Names from DQ XI Server** to populate the list of dataflows, then select the Informatica dataflow, and click **Create Transform based on Dataflow**.

The transformation is created and placed on your Transformation Developer workspace.

Option descriptions

This table provides brief descriptions for each of the options on the **Select Dataflow** window.

Option	Description
Host Name	The name of your Data Quality server's host machine. Data Quality XI for Informatica automatically completes this field based on your Data Quality server setup.
Port Number	The number of your Data Quality server's port. Data Quality XI for Informatica automatically completes this field based on your Data Quality server setup.
Repository Location	The location of your Data Quality repository. Data Quality XI for Informatica automatically completes this field based on your Data Quality server setup.
Substitution File	The name of any substitution file set up for Data Quality. Data Quality XI for Informatica automatically completes this field based on your Data Quality server setup.
Get Dataflow Names from DQ XI Server	Click to populate a list of dataflows in the Select a Dataflow for the Transform area.
Create Transform based on Dataflow	Click after you've highlighted the dataflow that you want to use. The newly created transformation uses the same ports and setup as the dataflow that you highlighted.
Create Blank Transform	Click to create a new transformation (active or passive based on which toolbar icon you chose) that does not have ports or settings.

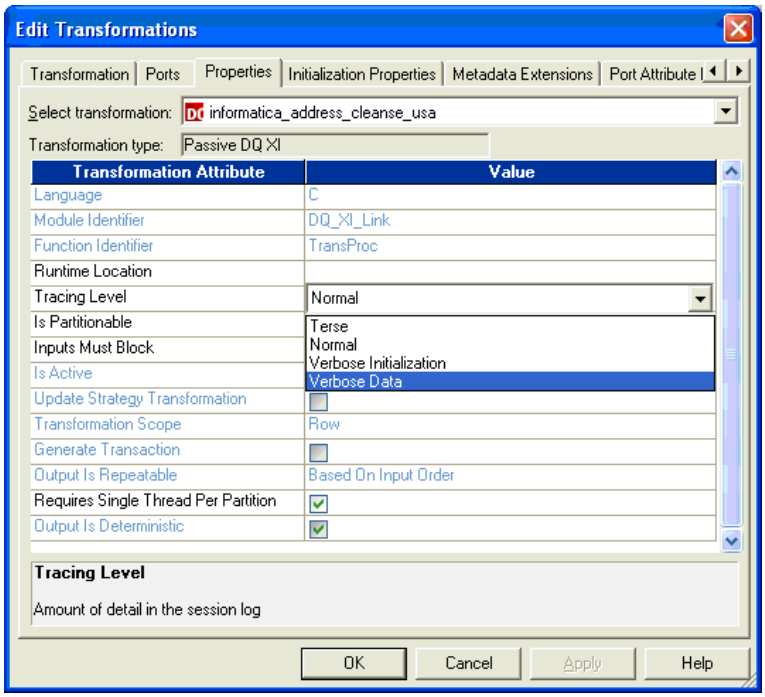
Chapter 4: Troubleshoot Data Quality XI for Informatica

If you can’t get something to work properly within BusinessObjects Data Quality XI for Informatica, this section may help you correct the problem. This section deals mostly with issues within Data Quality XI for Informatica itself. At times, issues may actually be caused by your Informatica applications. Be sure to consult the Informatica PowerCenter documentation as well.

Set the tracing level

In addition to setting the DQ XI properties, you may also want to use the **Edit Transformations** window to control the level of information recorded in the Informatica log files when you run a mapping. You control this through the **Tracing Level** setting on the **Properties** tab.

1. From the **Transformation Developer workspace**, double-click the transformation and click the **Properties** tab of the **Edit Transformation** window.



2. In the **Tracing Level** property, set the appropriate level.

   In addition to the normal logging produced by any of the settings, the Tracing Level **Verbose Data** creates two files, **SocketReader.log** and **SocketWriter.log**, in the Informatica Server's working directory. For more information about level of information provided with each tracing level, see your Informatica documentation.

Data Quality XI for Informatica transformation/mapping issues

The sample mappings are failing in the Workflow Manager

If you try to run the sample mappings through the Workflow Manager but they fail to complete, there are a few things to check.

Did you install in the default location?

If you installed the Data Quality XI for Informatica plug-in in a directory other than the default, you will have to edit some of the session properties and transformation properties. Edit the following items so that they point to the location where you installed the Data Quality XI for Informatica plug-in:

- Location of the source in the Source Options.
- Location of the target in the Target Options.
- The path for the Runtime Location property of the transformation.

Follow these steps to get to these properties:

1. In the **Workflow Manager**, Double-click on the workflow in the **Workflow Designer** tool. The Edit Tasks window opens.
2. Select the **Mapping** tab.
3. Open the **Source** or **Target** folders to edit the locations.
4. Open the **Transformation** folder to edit the Runtime Location.

Do the Data Quality XI for Informatica plug-in's files and directories have appropriate permissions?

Double-check to be sure that you have set the Data Quality XI for Informatica plug-in files' permissions properly for the person who is running the session. The directory where the sample job files are located needs write access for the session owner, for example.

Data isn't being processed when sent through a Data Quality XI for Informatica transformation

If your mappings are running, but data isn't being cleansed or consolidated, Data Quality XI for Informatica may not be properly installed.

If you have other transformations that are working properly, you know that Data Quality XI for Informatica is installed properly. Try running one of the mappings in the sample mapping folder if you're unsure. See Appendix A: "Sample mappings" on page 44 for more information about running a mapping. If the samples don't work, verify that you've performed all of the installation procedures correctly as documented in Chapter 2: "Installation" on page 13.

Process terminating on signal/exception

You might receive this error if you improperly edit some of the properties in a custom transformation. To fix this error, delete the custom transformation from your mapping and add a new custom transformation. Do not modify the Procedure Name property or the ProcIndex properties from their initial values.

Glossary

active transformation

A transformation in which Informatica PowerCenter can export records in succession without waiting for a return record.

custom transformation

An Informatica PowerCenter mapping element, which receives every row of data through input ports, performs some operation on that data set, and returns a set of output rows through output ports. These are implemented with two types of transformations, active and passive.

Data Quality

Data Quality is the Business Objects technology that performs data quality operations on the back end of Data Quality XI for Informatica.

dictionaries

Files that contain a lexicon of words and phrases that BusinessObjects Data Quality XI for Informatica uses to identify, parse, and standardize address data. These files have a **.dct** extension.

directories

Files of national address data compiled by the USPS and used by Data Quality to match, assign, standardize, and verify addresses. These files have a **.dir** extension.

discrete fields

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

metadata extensions

Key/value pairs defined within a transformation. These provide flexibility, allowing you to store different data types of information including a blob of information.

multiline fields

Input or output data that has certain address and name data combined in one field.

passive transformation

A transformation that outputs one row for every input row.

transformation

A repository object that generates, modifies, or passes data. BusinessObjects Data Quality XI for Informatica provides both active and passive custom transformations for Informatica PowerCenter.

Index