



Paribus *Discovery*[™]

Version 1.5

Getting Started Guide

Document Version 1.3



QGate Software Limited
D2 Fareham Heights, Standard Way, Fareham
Hampshire, PO16 8XT
United Kingdom

Tel +44 (0)1329 222800

info@QGate.co.uk
www.QGate.co.uk

© QGate Software Limited – All rights reserved



Notices

Copyright Information

This document and all subject matter outlined within this document remain the copyright of QGate Software Limited. It may not be reproduced in part or whole or any manner, digitised, transcribed, translated or mass distributed without written permission from QGate Software Limited.

© 2014 Copyright QGate Software Limited

Trademarks

Paribus™ is a trademark of QGate Software Limited.

Paribus *Discovery*™ is a trademark of QGate Software Limited.

QMatch+™ is a trademark of QGate Software Limited.

Microsoft is a registered trademark of Microsoft Corporation.

All rights reserved.

Disclaimer

Although every effort has been made to ensure the processing performed by this software product will not damage or corrupt your data, we strongly recommend you perform the following actions to safeguard against such eventualities.

In the unlikely event that data becomes lost or corrupted, QGate Software Limited cannot be held responsible.

Recommended Safeguards

- Perform a complete backup of any host database(s) to which Paribus Discovery connects to obtain match data.

Note: Paribus Discovery does not alter any data in host databases as part of the matching and review process. Changes are only made by subsequent updates / imports via Paribus Discovery Plug-in Tools.

- If possible, perform any match processes or updates using live databases outside of normal operational hours.

- If applicable, all remote user databases should be synchronised with the main host database(s) before any updates are performed.

Table of Contents

Notices	3
Copyright Information	3
Trademarks.....	3
Disclaimer	3
Table of Contents	5
Overview	8
Other Related Information	8
About Paribus <i>Discovery</i>	9
Intelligent Matching with QMatch+	9
Phonetic Matching.....	9
Matching Synonyms	10
Sequence Variation Matching	10
Sanitization.....	10
Data Segmentation.....	10
Case and Spelling	10
International Language Support	11
System Requirements	12
System Requirements	12
Paribus <i>Discovery</i> Installation	13
Installing the Paribus <i>Discovery</i> Application.....	13
Running Paribus <i>Discovery</i>	19
Launching the Application	19
Logging into the Paribus Control Database.....	20
Paribus Assistant	21
Application Settings	22
Paribus <i>Discovery</i> Licence Control	23
Paribus <i>Discovery</i> Licensing	23
Server Licences.....	23
Plug-in Licences	23
Paribus <i>Discovery</i> Licence Manager	23
Adding a New Licence.....	24
Removing a Licence.....	24
Editing an Evaluation Licence	24
Paribus Control Database	25
Paribus Control Database	25
Supported Platforms.....	25
Paribus Control Database Creator	25

Paribus Discovery Definitions Overview	28
Match Sessions	28
Match Sets	28
Match Conditions	28
Data Sets	28
Data Providers	28
Importing Paribus Definitions.....	29
How to Import Definitions	29
Defining and Running a Paribus Match Process	31
Step 1: Launch the Paribus Assistant.....	31
Step 2: Select the Data Source Type	32
Step 3: Setting up Paribus Definitions	33
Data Providers.....	33
Data Sets	34
Match Sets	34
Match Conditions.....	34
Step 4: Create a Paribus Match Session.....	35
Match Session.....	35
Step 5: Running a Paribus Match Process	36
Step 6: Reviewing Your Paribus Match Results	36
Step 7: Processing Your Paribus Match Results.....	37
Achieving Effective and Efficient Matching.....	39
1. Use Multiple Match Process Passes for Best Results	39
2. Use Match Conditions	39
3. If no Match Conditions are used, increase Match Set Threshold	40
4. Use of 'Extended Matching'	40
5. Use 'Log Unmatched Records' with care	40
6. Use Native OLEDB Drivers in Data Providers	40
7. Select Data Set fields and joins with care	40
8. Index the underlying data tables.....	40
9. Set Paribus Match Dictionary rules carefully	40
Paribus Discovery – Troubleshooting	42
Unable to Connect to my Paribus Control Database.....	42
I don't get any Matches when I run my Match Session	42
Paribus Discovery – Technical Support	43
General Support.....	43
Troubleshooting.....	43
System Requirements	43
Product Information.....	43
Fault Reporting	43

Upgrades and Service Releases	43
Systems Integration	44

Overview

This document is a guide to the basic installation process for installing Paribus *Discovery*.

This guide outlines the following:

- About Paribus *Discovery* and QMatch+
- System Requirements
- Installation and configuration of Paribus *Discovery*
- Paribus *Discovery* Licensing
- The Paribus Control Database
- Overview of Paribus Definitions
- Basic guide to defining and running a Paribus Match Process
- Introduction to achieving effective matching
- Troubleshooting

Other Related Information

For further information on using Paribus *Discovery*, see the Paribus *Discovery* product online help.

For additional reference information on Paribus *Discovery*, see the QGate KnowledgeBase at <http://www.QGate.co.uk/knowledge/paribus-discovery>

For up to date information on Paribus *Discovery* and other QGate products, please visit the QGate Software website at www.QGate.co.uk.

About Paribus *Discovery*

Paribus *Discovery* is a sophisticated application for intelligently matching business information such as Company Names, Person Names and Addresses etc, for the purposes of identifying similarities in corporate customer information.

The name **Paribus** comes from the Latin, meaning "All Things Equal".

Using the sophisticated matching capabilities of QGate Software's **QMatch+** matching technology (see below), Paribus *Discovery* is able to identify matches in data regardless of phonetics, synonyms, sequence, abbreviations, segmentation, spelling and case.

Paribus *Discovery* is able to match on any combination of data, with the added ability to combine multiple elements into a single match process (i.e. the ability to match a person's name with matching address elements and company associations).

Furthermore, Paribus *Discovery* enables you total flexibility to control how you weight the significance of the elements you match on. For example, you can denote a match performed on person names to be a 'close match', and the match performed on related address to be 'relaxed'.

Once Paribus *Discovery* has identified matches within your data, Paribus *Discovery* will present a summary report of the matches found. This details the number of records processed and the number of matches established.

In addition, Paribus *Discovery* provides the ability to review in detail the matches found, together with information about each matched item and the match score it obtained (degree of likeness). The review process enables the user to view each match found, altering the results if required, before processing / exporting the results from Paribus *Discovery*.

Finally, once matches have been established and reviewed, Paribus *Discovery* provides a variety of tools and plug-ins with which to processes the match results against your data. These include the ability to export the match results for processes against legacy systems and proprietary applications.

Intelligent Matching with QMatch+

Paribus *Discovery's* matching capabilities are powered by QGate Software's intelligent matching technology – **QMatch+**. QMatch+ is a sophisticated matching algorithm that is able to intelligently identify matches within data, regardless of phonetics, synonyms, sequence, abbreviations, segmentation, spelling and case.

Phonetic Matching

QMatch+ is able to match data based upon phonetics, such as:

- **Photo Centre**, Foto Center
- **Robert Stephenson**, Bob Stevenson, Bobbie Stevensen
- **Sherly Wyiatt**, Cheryl Wiatt, Sheral Whiat

Matching Synonyms

QMatch+ provides support for matching synonyms, such as:

- **Robert**, Bob, Bobbie, Rob, Robbie, Roberto
- **William**, Will, Willy, Bill, Billy
- **Richard**, Rich, Ric, Dick
- **Geoff**, Jeff
- **Lewis**, Alois, Alos, Aloysius (*international names*)
- **Eugene**, Eugenijus, Eugenio, Eugenius, Eugenios (*international names*)
- **Frank**, Franck, Franko, Francois, Frankiskos (*international names*)
- **International Business Machines**, IBM, I.B.M

Sequence Variation Matching

QMatch+ provides a unique feature of identifying matches regardless of the sequence and extent of the words and/or names contained – a technique we call Sequence Variation, such as:

- **University of Florida**, Florida University
- **1st National Bank of Arizona**, Arizona First National Bank
- **West Midlands Police**, West Midlands and District Metropolitan Police

Sanitization

As part of the match process performed by **QMatch+**, is a technique called **Sanitization**. This is where certain words and abbreviations are managed (some excluded, some elaborated) to ensure that all possible permutations are handled in the match process. Examples such as:

- **Corporation**, Corp
- **Limited**, Ltd, Ltd.
- **University**, Uni

Data Segmentation

QMatch+ provides a unique feature of matching names regardless of segmentation (spaced/punctuated words), such as:

- **QGate Software**, Q Gate Software, Q-Gate Software
- **3Com**, 3 Com, 3-Com
- **GuideMark**, Guide Mark, Guide-Mark

Case and Spelling

QMatch+ is able to match data regardless of certain spelling variations, and variations in case, such as:

- **University of London**, Univercity of Londen
- **NewMarket Partners**, NuMarket Partners
- **Geoff Cooper**, Jeff Couper
- **MCManiman**, McManiman

International Language Support

QMatch+ provides support for international language as standard. With its open match rules engine, QMatch+ combines international language support seamlessly into each of its match processes.

Uniquely, **QMatch+** does not require additional languages to be selected explicitly, as it is believed this is not always feasible for an application such as Paribus *Discovery* to determine which language set to use when performing matches.

System Requirements

In the interest of providing customers with the latest product support information, QGate provides the following online resources within our QGate KnowledgeBase website:

System Requirements

For information regarding the minimum system requirements for installing and running Paribus Discovery, please visit the product system requirements page:

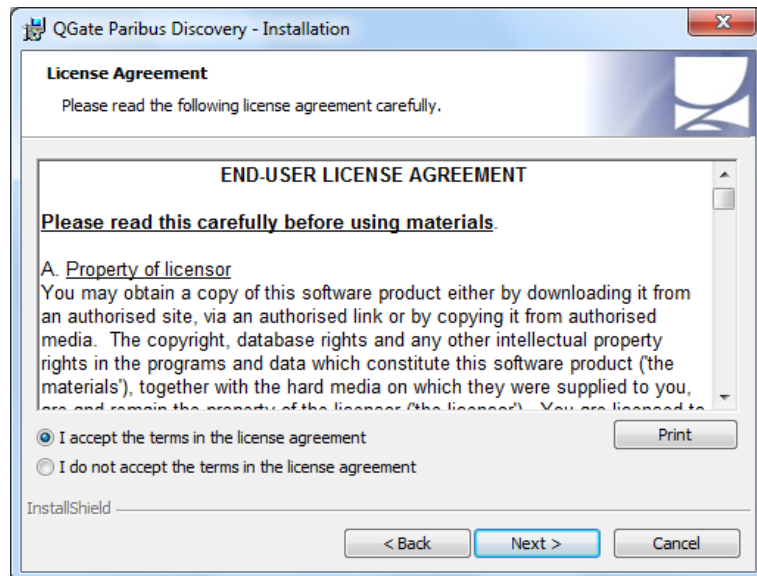
<http://www.QGate.co.uk/knowledge/paribus-discovery/sysreqs/>

Paribus *Discovery* Installation

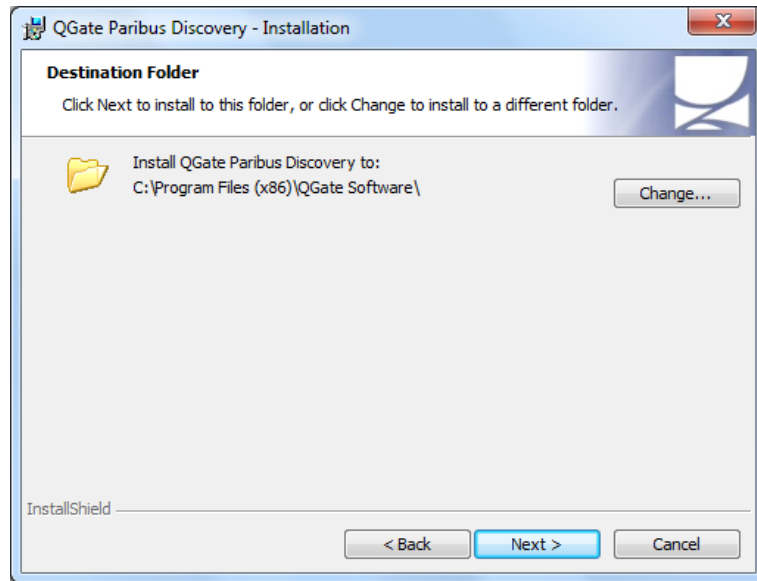
Installing the Paribus *Discovery* Application

Installation of the Paribus *Discovery* application is completed by use of an install wizard.

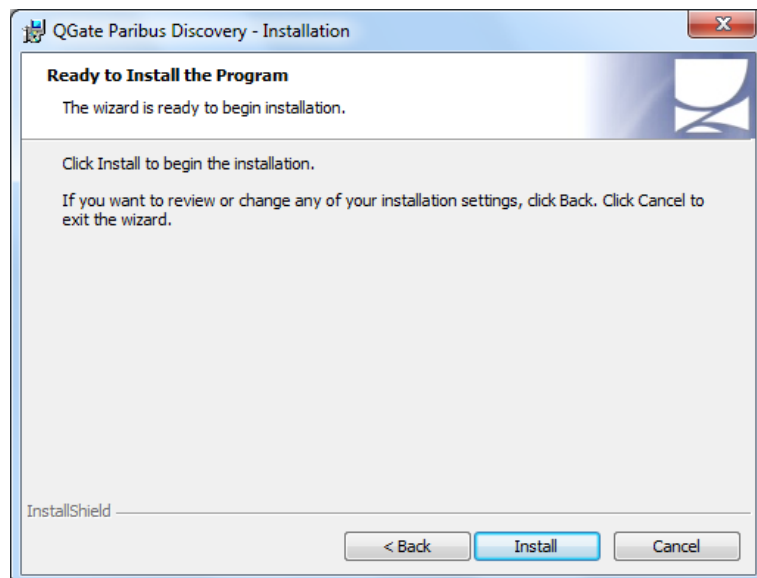
- 1) Run the **Paribus *Discovery* Setup.exe** file from within the **Setup** directory of your Paribus *Discovery* Installation Package.
- 2) From the *Welcome* dialog, click on the **Next >** button. Note: at any time during the installation procedure, clicking the Cancel button will end the installation and roll-back any files installed to that point.
- 3) From the **Welcome** dialog, click on the **Next >** button



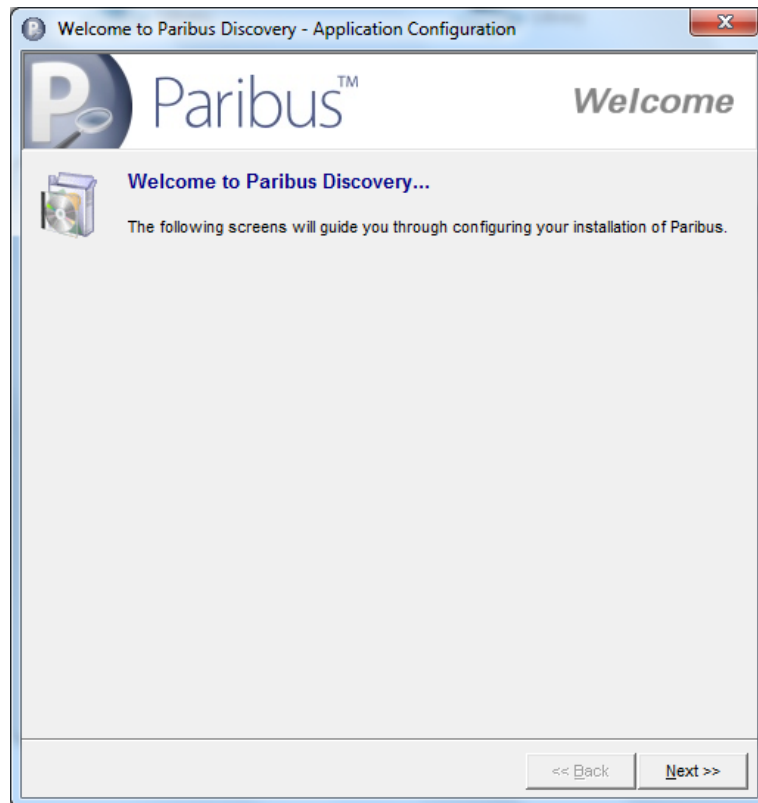
- 4) Read the license terms. If you do not accept the terms then cancel the installation now. If you do accept the terms then select the **I Accept...** radio button and click the **Next >** button.



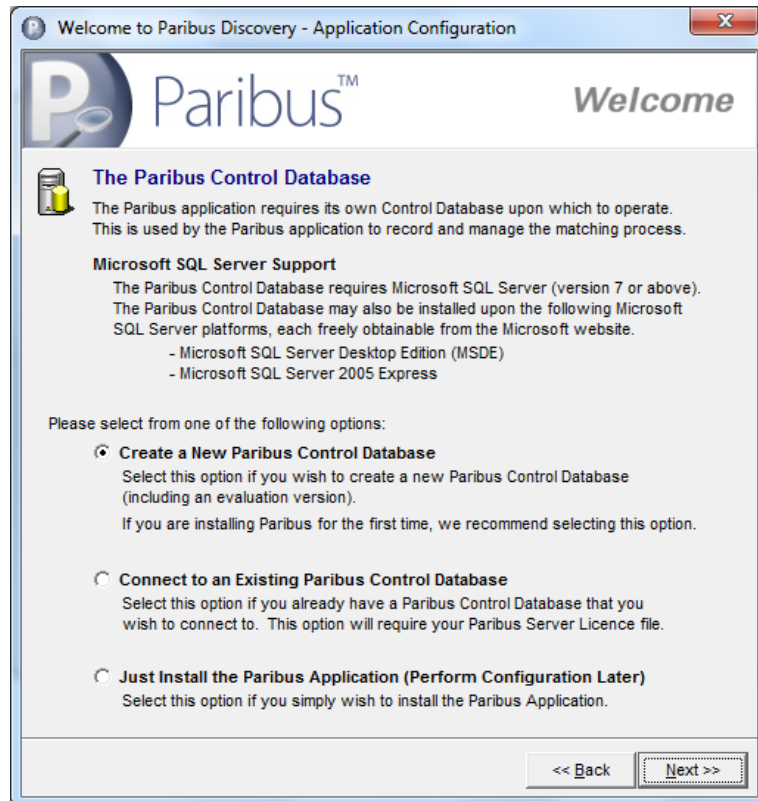
- 5) From the *Destination Folder* dialog, click on the **Next >** button to accept the default destination folder or click on the **Change...** button to locate an alternative path and folder.



- 6) When you are ready to proceed, click the **Install** button.
- 7) On the **Welcome to Paribus Discovery – Application Configuration** dialog, click **Next >>** to begin configuring your installation of Paribus Discovery.



- 8) On the **Paribus Control Database** dialog, select one of the following options:



Create a New Paribus Control Database

Select this option if you wish to create a new Paribus Control Database. If this is the first time you are installing Paribus, we recommend selecting this option to achieve the necessary Paribus application support.

Note: If you are installing an evaluation of Paribus, also select this option to create a new evaluation Paribus Control Database.

For more information on the Paribus Control Database, see the section on page 25.

On selecting this option and clicking the **Next >>** button will display the **Paribus Control Database Creator** (see page 25 for more information)

Connect to an Existing Paribus Control Database

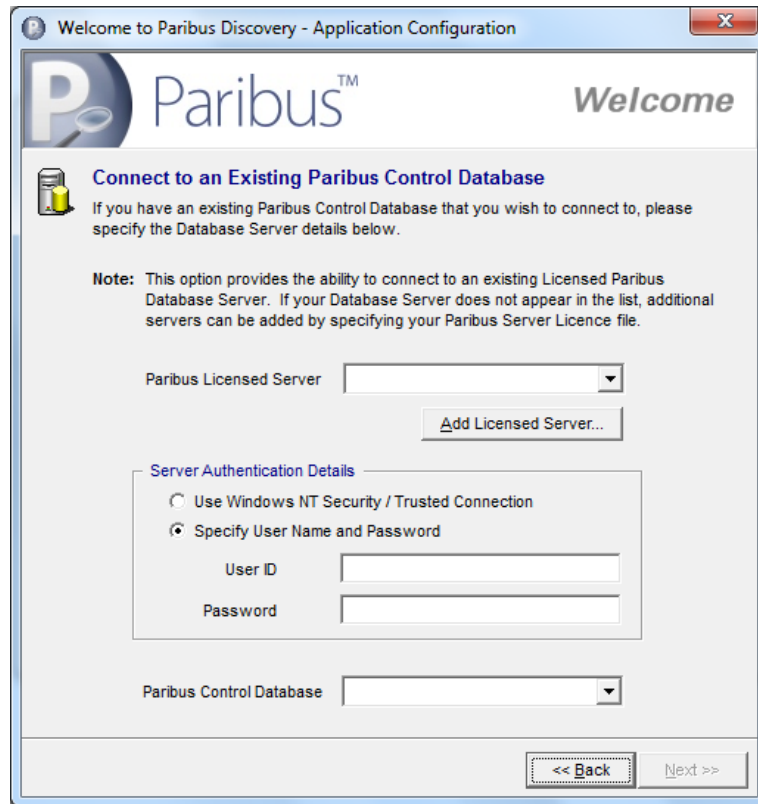
Select this option if you already have an existing Paribus Control Database that you wish to connect to.

On selecting this option and clicking the **Next >>** button, the **Connect to an Existing Paribus Control Database** page will be displayed (go to Step 9).

Just Install the Paribus Application

Select this option if you wish to simply install the Paribus application and configure your installation later.

- 9) If choosing to connect to an existing Paribus Control database, the following configuration page will be displayed.



The screenshot shows a window titled "Welcome to Paribus Discovery - Application Configuration". The window contains the Paribus logo and a "Welcome" message. The main heading is "Connect to an Existing Paribus Control Database". Below this, there is a note: "If you have an existing Paribus Control Database that you wish to connect to, please specify the Database Server details below." A "Note" section states: "This option provides the ability to connect to an existing Licensed Paribus Database Server. If your Database Server does not appear in the list, additional servers can be added by specifying your Paribus Server Licence file." The configuration fields include: "Paribus Licensed Server" (a dropdown menu), an "Add Licensed Server..." button, "Server Authentication Details" (with radio buttons for "Use Windows NT Security / Trusted Connection" and "Specify User Name and Password", and input fields for "User ID" and "Password"), and "Paribus Control Database" (a dropdown menu). At the bottom right, there are "Back" and "Next" navigation buttons.

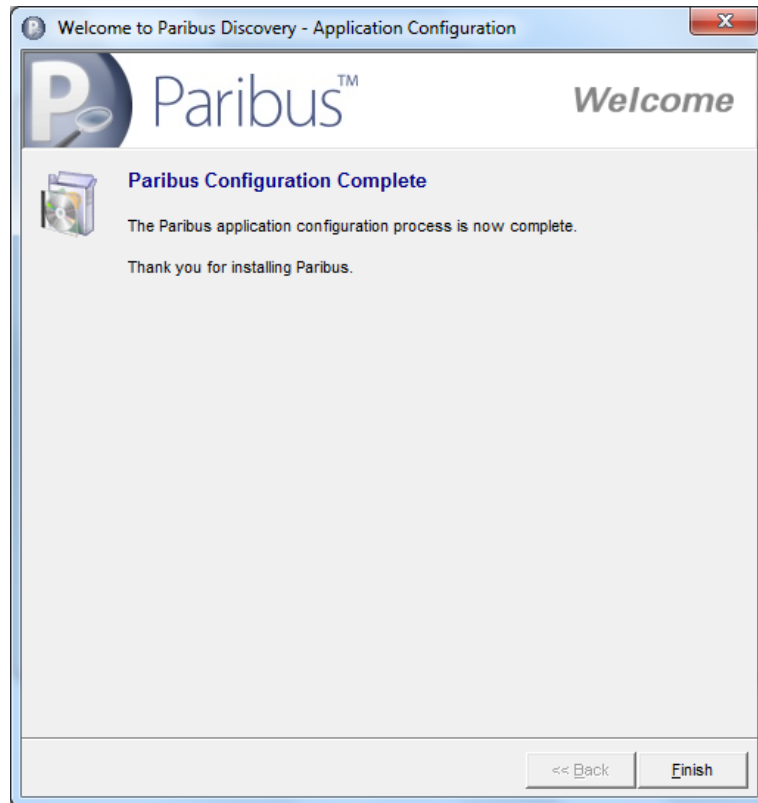
Select the **Paribus Licensed Server** you wish to connect to from the drop-down list of licensed servers. If the required server is not listed, additional servers may be added by clicking the **Add Licensed Server...** button. This button displays the **Paribus Licence Manager** allowing you to add a new Paribus Server Licence File (enabling access to the associated server). Please see page 23 for more information on the Paribus Licence Manager.

Choose your SQL Server Authentication details to enable connection to the desired Paribus Control Database on the specified licensed server.

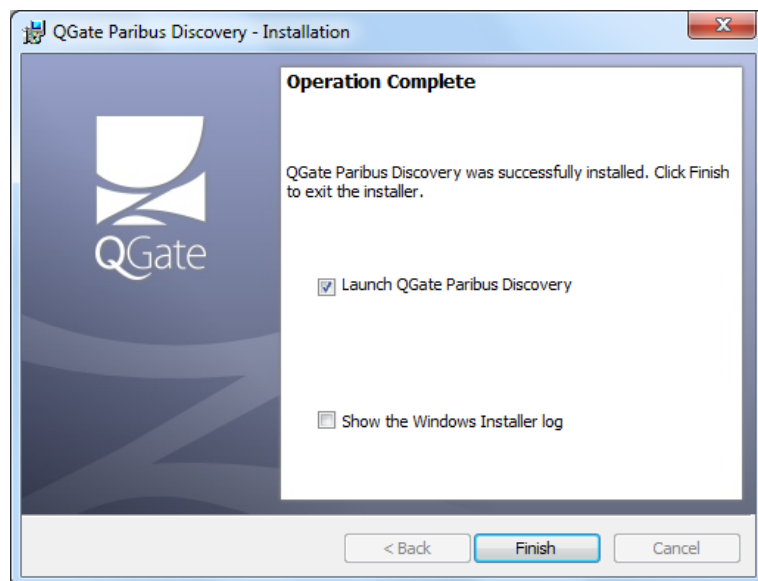
The **Paribus Control Database** dropdown-list provides a list of available databases on the specified server.

On completing the required details, clicking the **Next >>** button will complete the configuration process.

- 10) Once configuration is complete, click the **Finish** button to close the Paribus configuration window.



- 11) When the install has finished, the following window will be displayed.
- 12) Click the **Finish** button to complete the installation (optionally launching the Paribus *Discovery* application).



- 13) Click on the **Yes** button if prompted to restart your PC.

Running Paribus *Discovery*

Launching the Application

Paribus *Discovery* can be launched from the Paribus *Discovery* Application Icon.



Windows Start Menu: **Programs** → **QGate Paribus** → **Paribus *Discovery***

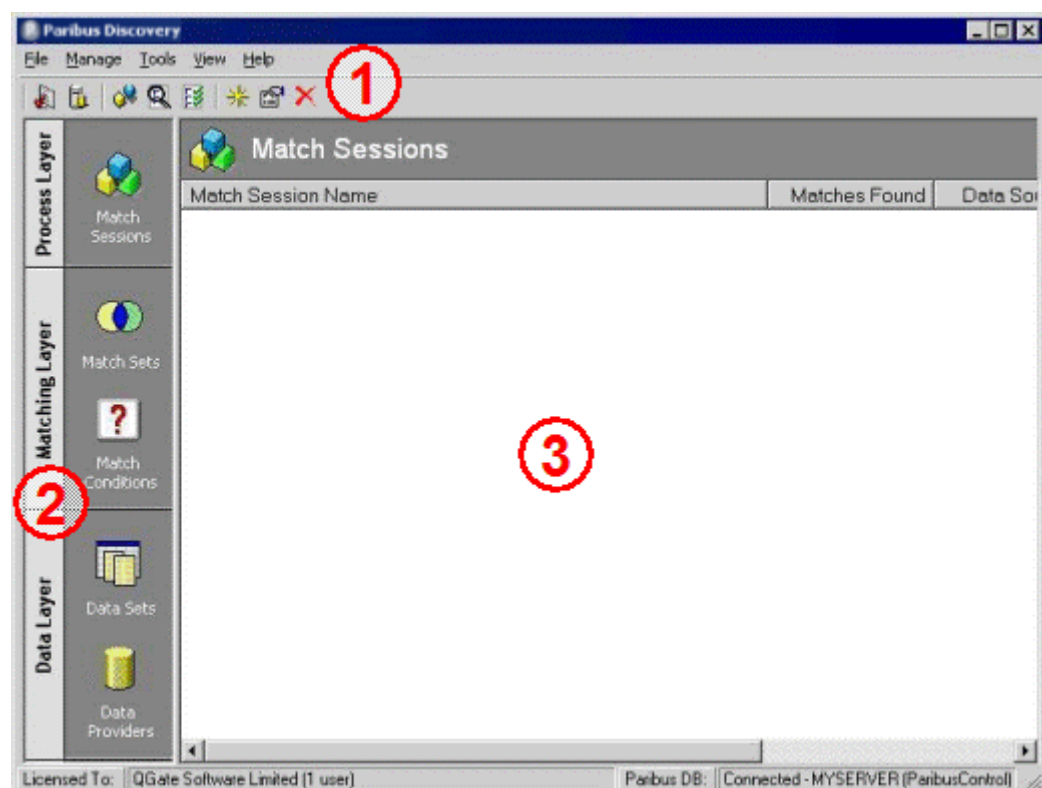


Figure 1 - Main Paribus *Discovery* Application Window

The main Paribus *Discovery* Application window comprises of three main areas:

1. The application menu/toolbar at the top of the window
2. The Paribus *Discovery* Navigation Bar on the left of the main window. Clicking on each of the icons in this navigation bar displays a list of the associated items in the definition list area.
3. The Paribus *Discovery* definition list area comprises the majority of the window. Here, a context-sensitive list of items is displayed. Double-clicking on an entry loads the item's definition for edit. Alternatively, a right mouse button (RMB) click on an item brings up a menu of choices (e.g. Edit, Delete, Create New, Run Session, Review Session, etc.).

Logging into the Paribus Control Database

Upon launching *Paribus Discovery*, you will typically be presented with the **Login To Paribus Control Database** window, as shown below.



Figure 2 - Paribus Control Database Login

This dialog provides the means of connecting to your Paribus Control Database.

If, during the installation process, you opted to Create a New Paribus Control Database or connect to an existing Control Database, the Server Name and Database Name details will already be defined. Select the Authentication method (and enter a valid User Name and Password if necessary) and click **OK**.



If you wish to change the Paribus Server/Database settings, click the **Configure Connection** button. This will display the **Application Settings** dialog (see page 21).

Paribus Assistant

When you log in to a new Paribus Control Database (or one containing no Match Sessions), the Paribus Assistant is displayed automatically, as shown below:



Figure 3 - Paribus Assistant

Select the appropriate option to help you get started using Paribus:

- **Define Paribus Definitions** – use this option if you wish to define your Paribus definitions manually. Please see page 31 for a basic guide to defining your Paribus definitions. Note: Paribus definitions can be imported later by selecting **Import Paribus Definitions...** from the main **File** menu.
- **Import Paribus Definitions** – use this option if you have a set of pre-defined Paribus definitions that you wish to import. Click the **Import Definitions...** button to display the **Import Paribus Definitions** dialog. Please see page 29 for more information on importing Paribus definitions.

For an overview of the available Paribus definitions, please see page 28.

Application Settings

The Application Settings dialog allows you to define your Paribus Control Database settings:

- Database Server (SQL Server)
Note: Available servers are based upon your installed licences.
- Database Authentication
- Paribus Control Database

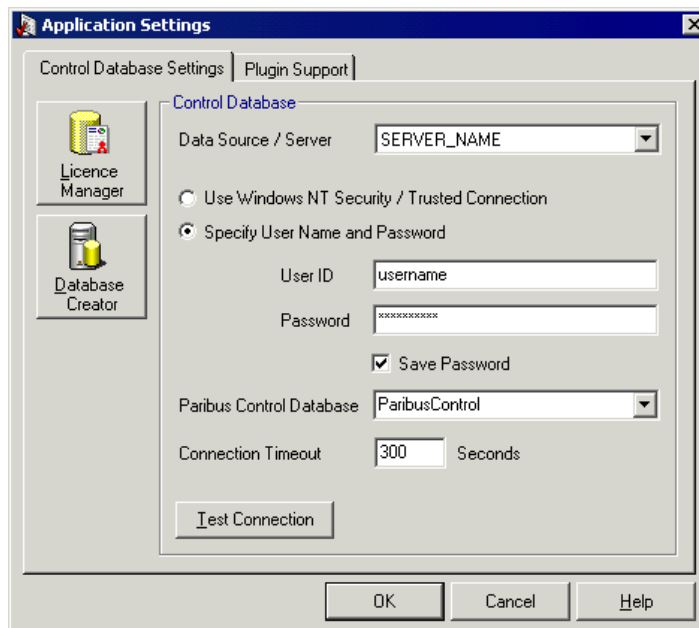


Figure 4 - Paribus Application Settings

New Paribus licences can be added from here by clicking the **Licence Manager** button. For more information on Paribus Licensing please see page 23.

If necessary, a new Control Database can be created from here by clicking the **Database Creator** button. For more information on using the Paribus Control Database Creator, please see page 25.

Paribus *Discovery* Licence Control

Paribus *Discovery* Licensing

There are two licence types used within Paribus *Discovery*:

Server Licences

Server Licences provide access to Paribus Control Database servers (see page 25 for more information on the Paribus Control Database). A server licence is required for each Paribus Control Database server. The number of concurrent connections to any Paribus Control database is limited to the number of users allowed by the licence.

Evaluation Licence

Under an Evaluation Server Licence a connection can be made to any SQL Server installation and connect to any Paribus Control Database. Concurrent connections under an evaluation licence are restricted to a single user. Match results from an evaluation Paribus Session cannot be exported or processed using Paribus *Discovery* plug-in tools

Plug-in Licences

Plug-in Licences provide access to Paribus *Discovery* Plug-in tools.

- Some Plug-ins provide limited functionality without the presence of a licence.
- Plug-in Licences are not covered in this Getting Started guide.

Paribus *Discovery* Licence Manager

The Paribus *Discovery* Licence Manager is used to add, edit, and delete Paribus *Discovery* Licences and is accessed via the 'Licence Manager' button on the **Application Settings** dialog, or from the Paribus *Discovery* main menu: **File > Licence Manager...**

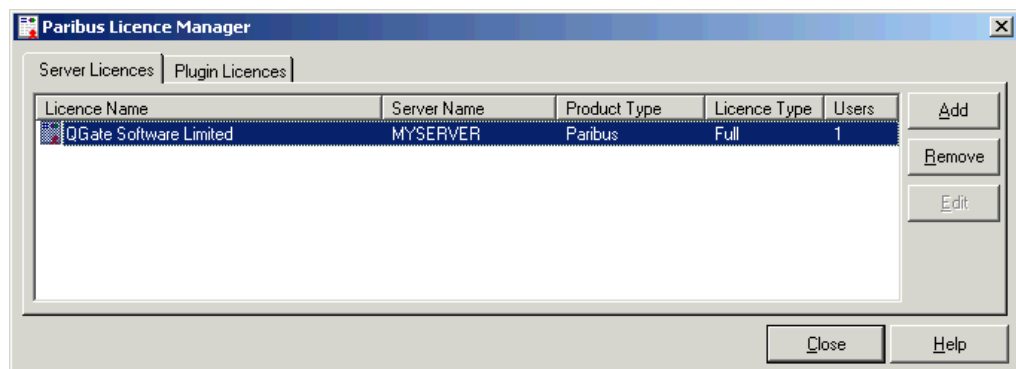


Figure 5 - Paribus *Discovery* Licence Manager

Adding a New Licence

New licences are added by clicking the **A**dd button on the Licence Manager. A standard file selection dialog is displayed from where the required Paribus *Discovery* Server Licence (*.lic) can be selected.

Assuming a valid Server Licence has been selected, the new licence will be displayed in the **Server Licences** list.

Removing a Licence

A Server Licence can be removed by selecting it in the Licence Manager list and clicking the **R**emove button.

Note: Removing a licence does not remove any Paribus Control database on the related SQL Server.

Editing an Evaluation Licence

Editing an evaluation licence allows the SQL Server name to be changed (*only evaluation licences can be edited*). An evaluation licence can be edited by selecting it in the Licence Manager list and clicking the **E**dit button. The Evaluation Licence Server Name dialog (shown below) is displayed. Enter the new SQL Server name and click **OK**.

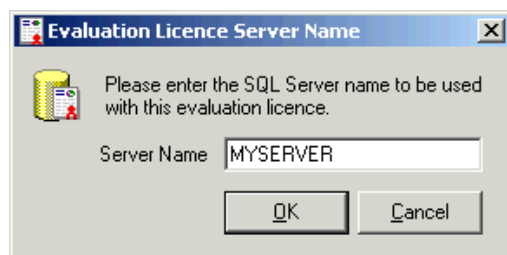


Figure 6 - Edit Evaluation Licence Server Name

Paribus Control Database

Paribus Control Database

The Paribus Control Database is the database upon which the Paribus *Discovery* application operates. The Paribus Control Database is used to manage the match process, and store the definitions and configurations you construct whilst using Paribus *Discovery*.

Supported Platforms

Currently the Paribus Control Database is only supported on Microsoft SQL Server and Microsoft SQL Server Express Edition (versions 2005 or 2008).

Paribus Control Database Creator

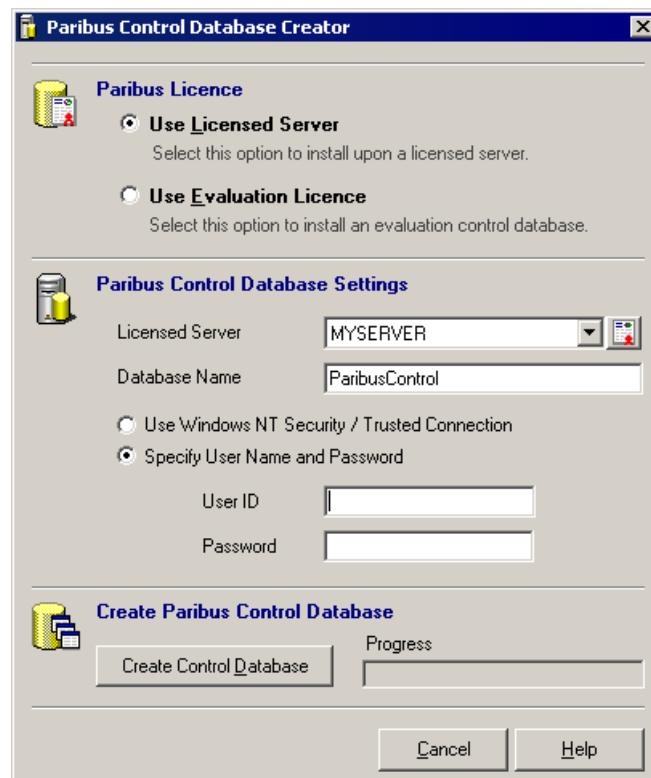


Figure 7 - Paribus Control Database Creator

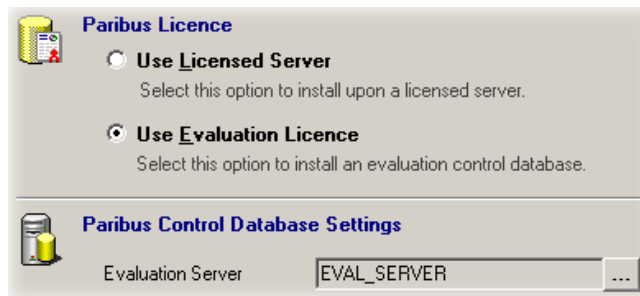


Figure 8 - Paribus Control Database Creator (Use Evaluation Licence)

The **Paribus Control Database Creator** provides the ability to generate new Paribus Control Databases within a specified Microsoft SQL Server database (including MSDE).

Paribus Discovery Licence

Two licensing options are available when creating a new control database:

- **Use Licensed Server** – Select this option if you have an existing Paribus *Discovery* Server License installed or available to install and wish to create a new control database on the associated server.
- **Use Evaluation Licence** – Select this option if you do not have any Paribus *Discovery* Server Licenses available and wish to trial Paribus *Discovery* as an Evaluation.

Note: Running Paribus under evaluation will provide full access to the match processes, but does not permit any ability to process or export the match results.

Licensed / Evaluation Server

The Licensed / Evaluation Server value denotes the server name of your database server.

If **Use Licensed Server** is selected in the **Paribus Discovery Licence** section, then this information is derived from the Licence. Click on the **Add Licence...** button beside the Licensed Server drop-down list to add new Paribus Server Licences. This displays the Paribus *Discovery* Licence Manager as described on page 23.

If **Use Evaluation Licence** has been selected in the **Paribus Discovery Licence** section, the required server name can be set by clicking the ellipsis button to the right of the **Evaluation Server** text field and entering the server name in the displayed dialog window.



Figure 9 - Evaluation Licence Server Name

Database Name

The Database Name is the name you wish to call your new Paribus Control database. By default, this will be named *ParibusControl*. Note: subsequent control databases generated on the *same* server must have unique names.

Use Windows NT Security / Trusted Connection vs. Specify User Name and Password

A connection can be made to a SQL Server database either by using the login credentials of the client machine if that Windows logon has access rights to the SQL Server database or by explicitly specifying a SQL Server User ID and Password. Select the appropriate option to allow connection to the required SQL Server (see your Systems Administrator for assistance, if necessary).

Note: Whichever login method is used, the login user must have the appropriate privileges to create databases on the specified database server. If appropriate, this can be the database administrator User ID

User ID

If **Specify User Name and Password** has been selected, enter the User ID required by your database server to log into and create the Paribus Control Database.

Note: The SQL Server default system administrator ID is **sa**.

Password






If **Specify User Name and Password** has been selected, enter the Password required by your database server to log into and create the Paribus Control Database.

Create Control Database

Clicking the **Create Control Database** button will automatically create a Paribus Control Database, based upon the database details and login credentials specified.

Paribus *Discovery* Definitions Overview

This section provides a brief overview to the Paribus Definitions that are used within the Paribus *Discovery* application to support the match process. For more detailed information on Paribus Definitions, see Paribus *Discovery* online help.

Process Layer	 <h3>Match Sessions</h3> <p>A Match Session is the highest-level object in Paribus and defines the rules that will be applied in a match process. It also contains the match results for that match process and allows those results to be viewed, reviewed, and exported/processed via the various tools and plug-ins.</p>
Matching Layer	 <h3>Match Sets</h3> <p>A Match Set is a template defining the initial two items of data (Data Sets) that will be compared within the match process.</p> <p>A Match Set also defines related Match Conditions that can be optionally applied to the Match Set when used within a Match Session.</p> <p>When used within a Match Session, the definitions of the Match Set form the default (template) settings, however these can be overridden within the Match Session.</p>  <h3>Match Conditions</h3> <p>A Match Condition is a template defining additional match criteria that can be optionally applied to the matches established from a Match Set.</p> <p>Once a Match Set has established a collection of initial matches (e.g. matching Company Names), a Match Condition can then be applied to those matches (e.g. on Postal code).</p> <p>One or more Match Conditions can be applied during the match process, and only those results meeting the criteria of <i>all</i> or any <i>one</i> of the conditions (AND / OR) will remain in the match results.</p> <p>Match Conditions are linked to Match Sets based upon their related data and are available for selection in the match process via the Match Session.</p>
Data Layer	 <h3>Data Sets</h3> <p>A Data Set defines how the match data and supporting information are to be retrieved from a specified data source (accessed by a Data Provider) during a match process. It also allows the definition of filters that can be optionally applied during a match process to limit the type or extent of data being retrieved.</p> <p>Data Sets are referenced by both Match Sets and Match Conditions and a given Data Set can be used in any number of Match Sets and Match Conditions.</p>  <h3>Data Providers</h3> <p>A Data Provider is the lowest-level object in Paribus and defines a connection to a source of data to be used in a match process.</p> <p>Data Providers are used by Data Sets to access the underlying data. A given Data Provider can be used in any number of Data Sets.</p>

Importing Paribus Definitions

Paribus *Discovery* provides the ability to import a set of pre-defined Paribus definitions, which have been pre-configured for a given database model or application (i.e. CRM application database).

This facility enables Paribus *Discovery* to be rapidly implemented upon a given database application/model, without the need for manually defining each of the data entities.

Contact your Paribus *Discovery* software supplier for a list of pre-defined Paribus Definitions.

How to Import Definitions

To import a set of Paribus definitions, select **Import Paribus Definitions...** from the **File** menu in the main Paribus *Discovery* application. The **Import Paribus Definitions** dialog will be shown, as follows:

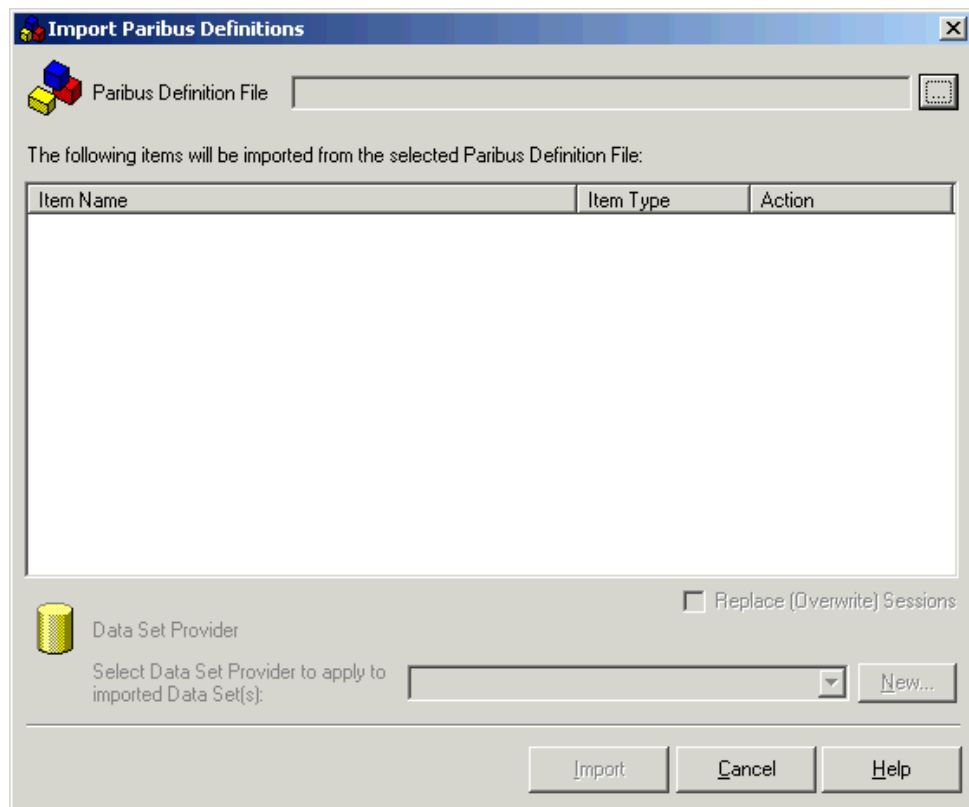


Figure 10 - Import Paribus Definitions Dialog

- Select the Paribus Definition file you wish to import by clicking the ellipsis (...) button.

Note: This action will happen automatically when first entering this dialog.

- From the **Select Paribus Definition File** dialog, select the Paribus Definition file you wish to import. Paribus Definition files have a file extension of **.pds**.
- Once selected, on returning to the **Import Paribus Definitions** dialog the main list should contain the definition items you wish to import.

- The **Action** column denotes which action will be taken during the import. If a definition item already exists, this will be overwritten (if required). If a definition item does not exist, the action will be to add as new.
- The **Replace (Overwrite) Sessions** checkbox denotes if an existing Match Session will be overwritten (if applicable).
- **Data Set Provider:** If you are importing Data Sets, these will require a Data Set Provider to provide them a connection to a database.

Select from the drop-down list of Data Set Providers an appropriate provider. If an appropriate provider is not available or no providers are listed, clicking the **New...** button will enable you to create a new Data Provider.

- Finally, click the **Import** button to import the selected Paribus Definitions.

Defining and Running a Paribus Match Process

This section outlines the steps required to define and perform a basic Paribus match process, based upon a single source of data (assuming a data de-duplication scenario).

It uses the **Paribus Assistant** to define the match process elements and then run the resulting Match Session.

Furthermore, the section outlines how to review the match results and how to process/export them using the various Paribus *Discovery* tools and plug-in components.

For more detail information on how to use and configure Paribus *Discovery*, see the Paribus *Discovery* product online help.

Step 1: Launch the Paribus Assistant

The first step is to launch the Paribus Assistant. If your Paribus Control Database is empty (or contains no Match Sessions), the paribus Assistant will be displayed automatically when you log in to the database.


Thereafter, the Paribus Assistant can be displayed by selecting **File > Paribus Assistant** from the main Paribus *Discovery* menu or by clicking the Paribus Assistant toolbar  icon:



Figure 11 - Paribus Assistant

Click the **Define...** button to begin creating your Paribus definitions.

Step 2: Select the Data Source Type

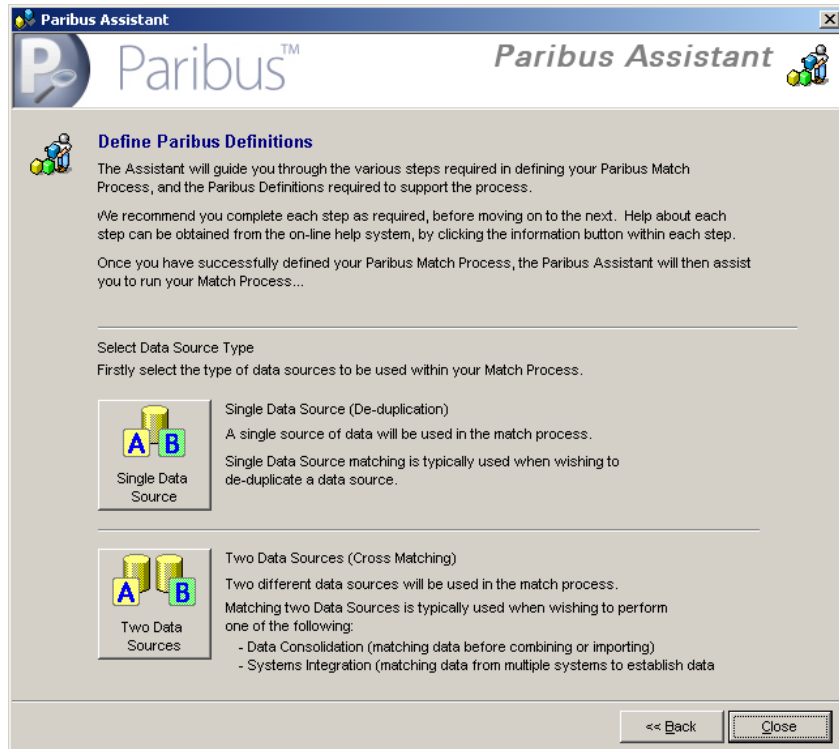


Figure 12 - Paribus Assistant Data Source Type

To begin creating your Paribus definitions, select the **Data Source Type** you wish to work with. This affects how the subsequent Paribus Definitions are created.

In this example, we will be using a *Single Data Source* (i.e. a data de-duplication scenario), so click the **Single Data Source** button.

Step 3: Setting up Paribus Definitions

The next step is to set-up the Paribus definitions to support your match process.



Figure 13 - Paribus Assistant Create Definitions

Data Providers

The first definition required is a Data Provider, which is responsible for providing the data to the match process.

To create a Data Provider, click the **Providers...** button. The **Data Providers List Window** will be displayed. Click the **New** button to create your Data Provider.

This action will display the **Configure Data Provider** dialog, enabling you to define the connection details to your data source. This can be any data accessible via an OLE DB Provider or ODBC connection (we recommend using OLE DB Providers wherever possible).

Note: If your data is not in a database table format (e.g. your file is a CSV file or other text file), we recommend importing your data into a database such as a Microsoft Access or SQL Server to enable access via OLE DB or ODBC.

For more information on how to configure a Data Provider, click the **Help** button on the **Configure Data Provider Window**.

Once you have successfully established a connection via a Data Provider, you can proceed to define Data Sets.



Data Sets

Once you have defined a Data Provider to source the data from your database, you can then define one or more Data Sets.

Data Sets define the mapping of data from database tables and columns, providing information and values from the database to be used in the match process (e.g. company names, person names, addresses etc).

To create a Data Set, click the **Data Sets...** button. The **Data Sets List Window** will be displayed. Click the **New** button to create your Data Set.

This action will display the **Data Set Definition Window**, enabling you to define the data mapping of the **Data Set Definition Wizard**.

For more information on how to configure a Data Set, click the **Help** button on the **Data Set Definition Wizard**.

You will need to define a Data Set for each element you require in the match process. For example, if you wish to match on company names, with an additional match on postal code, you will need to define two Data Sets – one providing company names, another providing postal codes.



Match Sets

A Match Set is a combination of two Data Sets that are compared against each other in the match process.

The Match Set is always the driving element in the match process, and is what forms the initial matching criteria.

Additional matching criteria can be applied to the match process, in the form of Match Conditions (see below). Match Conditions are associated to the match process via the Match Set. Associated Match Conditions are not enforced by association, but can be enabled/disabled when used within a Match Session.

To create a Match Set, click the **Match Sets...** button. The **Match Sets List Window** will be displayed. Click the **New** button to create your Match Set.

This action will display the **Match Set Definition Window** enabling you to define your Match Set.

Settings defined in a Match Set definition form the default template settings whenever this Match Set is used within a Match Session. Once included within a Match Session, these settings can be overridden.

For more information on how to configure a Match Set, click the **Help** button on the **Match Set Definition** dialog.



Match Conditions

If your match process requires the matching of at least two or more criteria (e.g. company names, with corresponding postal codes), you will need to create Match Conditions for each of the additional criteria.

Note: You will require additional Data Sets to support each Match Condition you wish to apply.

A Match Condition is a combination of two Data Sets that are compared against each other in the match process, to apply a conditional match check against any matches established from a Match Set (see above).

To create a Match Condition, click the **Match Conditions...** button. The **Match Conditions List Window** will be displayed. Click the **New** button to create your Match Condition.

This action will display the **Match Condition Edit Window** enabling you to define your Match Condition.

Settings defined in a Match Condition form the default template settings for whenever this Match Condition is used within a Match Session. Once included within a Match Session, these settings can be overridden.

For more information on how to configure a Match Condition, click the **Help** button on the **Match Condition** dialog.

Linking Match Conditions to Match Sets

Having created a Match Condition, it must be linked to your Match Set(s) before it can be referenced in the match process.

The **Link Match Condition Window** is displayed automatically after a new match Condition has been created. Select the Match Set from the list to which the Match Condition should be linked, then click the **Link** button.

Step 4: Create a Paribus Match Session

The next step is to create a Paribus Match Session to define and contain your match process.



Match Session

To create a Match Session, click the **Create...** button in the Match Sessions section of the window.

This action will display the **Match Session Settings Window**, together with the **Add Session Match Set Window**. Select the Match Set from the list of available Match Sets and click **OK**. This will populate the **Match Session Settings Window** with details of the selected Match Set.

For more information on how to configure a Match Session, click the **Help** button on the **Match Session Settings** dialog.

Based upon a common match process, we recommend starting with the following settings for your Match Session. Naturally, variations to these settings are encouraged once an understanding of their significance is established.

Match Method

On the Match Method Tab, set the following settings:

- Set the match method to use **QMatch+ Intelligent Matching**
- Set the Match Score Threshold to **80%**

Match Conditions

On the Match Conditions Tab, set the following settings:

- Enable your Match Condition by clicking its checkbox.
- Double clicking the Match Condition enables editing of the condition.

Match Process

On the Match Process Tab, set the following settings:

- Set the match process to use **Standard Matching**

Processing Options

On the Process Options Tab, set the following settings:

- Check the **Automatically Identify a Primary Group Member** checkbox

Click **OK** to save your Match Session.

Step 5: Running a Paribus Match Process

Once you have created your Paribus Match Session, you are now in a position to run the match process.

To run a Paribus Match process, click the **Run Session...** button in the Match Sessions section of the **Paribus Assistant**. (Subsequently, to run a Paribus Match process, right mouse click the Match Session you wish to run on the main Paribus *Discovery* application window and select **Run Session...** from the popup menu).

This action will display the **Identify Matches** dialog detailing the match process outlined in your Match Session.

Clicking the **Run** button will start the match process.

During the process, the dialog will provide information on the progress and number of matches found.

Once the match process is complete, Paribus *Discovery* will display a Session Summary Report providing information about the match process, match operations performed, and matches found.

On closing the report window and returning to the **Identify Matches** dialog, you can review the matches found by clicking the **View Results...** button (only if any matches were found).

Click **Close** to close the **Identify Matches** dialog.

Note: If no matches were found, refine your match criteria within your Match Session and re-run the match process. Alternatively, see the Trouble-Shooting section on page 42.

Step 6: Reviewing Your Paribus Match Results

If your match process has established a collection of matches, the next step is to review and approve the match results. This is achieved using the Paribus Session Match Review window.

To review Session Match results, select the Match Session from the Match Session list, and right mouse click the Match Session you wish to review and select **Review Session...** from the popup menu.





This action will display the **Session Match Results** window.

The main aim of the **Session Match Results** window is to provide the ability to review the match results found and, importantly, set the status of each Match Group and Group Member accordingly.

Important Note: Only Match Groups and Group Members that have been reviewed can be processed by the various tools and Paribus *Discovery* plug-in components.




Match Groups

By default, each Match Group has a group status of “Not Reviewed”. Using the **Session Match Results** window, a Match Group’s status can be set to one of the following:

	Not Reviewed (Default)	Default Group Status before reviewed.
	Reviewed	Group has been reviewed and approved. (will be processed by tools/plug-ins).
	Queried	Group has been marked as queried (will not be processed by tools/plug-ins).
	Void	Group has been marked as void (will not be processed by tools/plug-ins).

Match Group Members

By default, each Group Member has a member status of either “Member” or “Primary”. Using the **Session Match Results** window, a Group Member’s status can be set to one of the following:

	Group Member (Default)	Group Member is a member of the group (in a duplicate cleansing exercise, this member would be denoted as a duplicate). Member will be processed by tools/plug-ins.
	Primary Group Member	Group Member is the Primary Group Member (in a duplicate cleansing exercise, this member would be denoted as the master record). Member will be processed by tools/plug-ins.
	Excluded Member	Group Member has been excluded from the group. Member will not be processed by tools/plug-ins.

Note: The Group Member statuses listed here are for Single Data Source matches only.

For more information on how to use the Session Match Results window, click the **Help** button on the **Session Match Results** window.

Step 7: Processing Your Paribus Match Results

The final step having achieved an effective set of match results, reviewed them and set each Match Groups status accordingly, is to process them using one of the Paribus *Discovery* tools or plug-in components.

If the match process did not provide suitable results, refine the Paribus definitions and Match Session details, and re-run the Paribus Match process (see previous steps).

All Paribus *Discovery* tools and plug-in components can be accessed from the Paribus *Discovery Tools* menu from the main application window.

Note: Paribus *Discovery* tools are not accessible under the evaluation version of the product.

For more information on the various Paribus *Discovery* tools and plug-in components, see the online help or additional plug-in documentation.

Achieving Effective and Efficient Matching

A number of steps can be taken to achieve effective matching using Paribus *Discovery*. These should be treated as guides, and not as strict rules.

1. Use Multiple Match Process Passes for Best Results

Depending on the structure of the data you wish to match on, it may be strongly advisable to run more than one match process on the data to achieve the most effective results.

For example, if performing a Contact record de-duplication process, duplicate Contacts may relate to the same Company record, or to duplicate Company records. It is, therefore, advisable to run a de-duplication process on the Company records first, with subsequent consolidation of the Company records (i.e. re-assignment of all records related to the duplicate Companies to a single 'master' company followed by deletion of the duplicate Company records). Then, run a Contact de-duplication process to identify duplicate Contacts within the remaining Company records.

Typical Scenarios

- **Company Name De-duplication** – only a single pass is necessary. Perform a match process on the Company records alone, with suitable Match Conditions applied (see Point 2, below).
- **Company Contact De-duplication** – two passes are advised. The first pass identifies duplicate Companies (as in Company Name Match, above), followed by consolidation of the duplicate Companies. The second pass identifies duplicate Contacts, with a Match Condition based on each contact's Company ID applied (see Point 2, below).
- **Private Contact De-duplication (no Company relationship)** – only a single pass is necessary. Perform a match process on the Contact records alone, with suitable Match Conditions applied (see Point 2, below).

2. Use Match Conditions

Applying Match Conditions to a Match Session can increase the accuracy of match results considerably. Depending on the match threshold used against the Match Set, relying on the matches from the Match Set alone can yield some over-optimistic matches. Applying one or more Match Conditions can significantly reduce this problem.

Note: if too many Match Conditions are applied (especially if the Match Condition Operator for the Session is set to 'AND') the match process may yield very few or no results. Likewise, the underlying data supporting the Match Condition(s) should also be well populated to ensure additional matches occur successfully.

As a rule, the Match Set threshold should be set to about 80%, while match thresholds for Match Condition(s) should be a little higher at about 90% (dependent upon the data).

Typical Scenarios

- **Company Name Match** – if a Match Set is based on Company Names, applying Match Conditions based on Postal Code, Town/City and/or Phone Number may provide more accurate results.

- **Company Contact Match** – if a Match Set is based on Company Contact Names, applying a Match Condition based on the Company ID (assuming the Company data has been thoroughly de-duplicated first - see Point 1, above), will ensure only duplicate Contacts at the same Company are matched.
- **Private Contact Match (no Company relationship)** – if a Match Set is based on Private Contact Names, applying Match Conditions based on Postal Code, Town/City and/or Phone Number may provide more accurate results.

3. If no Match Conditions are used, increase Match Set Threshold

If no Match Conditions are available (or none are applied to the Match Session) increase the Match threshold of the Match Set to 90%+. This will effectively reduce the number of over-optimistic matches.

4. Use of 'Extended Matching'

Extended Matching (under a Session's Match process options) uses more rigorous match rules and can often result in matches being found that would not otherwise occur. This, however, is at the expense of (significantly) increased processing time.

5. Use 'Log Unmatched Records' with care

'Log Unmatched Records' (under a Session's Process Options) is a useful option that stores within a Session's match results any records that did not match any other in the underlying data. However, storing the extra records incurs additional overhead that can be avoided if the logging of unmatched records is not required.

6. Use Native OLEDB Drivers in Data Providers

When configuring the connection to the underlying data in a Data Provider, always use the native OLEDB Provider where possible. Native OLEDB providers are usually more efficient than their ODBC counterparts. Only use an ODBC connection if a native OLEDB provider is unavailable.

7. Select Data Set fields and joins with care

Including appropriate information columns in a Data Set's definition can be very beneficial when reviewing the match results. However, more information columns (and associated database table joins) can slow the match process. Therefore, only include those columns that are useful and likely to be well-populated with data.

8. Index the underlying data tables

Look-ups on the underlying data will be quicker if the relevant columns in the database table(s) have appropriate indexes in place. Typically, the Primary Key, Foreign Key and any join columns defined in the Data Set should be indexed. Likewise, if any filters are being used, the look-up columns of those filters should also be indexed.

9. Set Paribus Match Dictionary rules carefully

Each Data Set has one or more associated Paribus Match Dictionaries (depending on the use of filters). A Match Dictionary contains intelligent references to the data

being matched upon, and is used by Paribus to achieve intelligent matching using QGate's QMatch+ matching component.

A Match Dictionary is always built the first time a match process is run against a Data Set. By default, the Match Dictionary is retained for a number of days (set by the options on the Data Set Options tab) to prevent the overhead of rebuilding the Match Dictionary repeatedly with subsequent Session runs.

If the underlying data changes, however, it is important to force a rebuild of the Match Dictionary by using the 'Delete Dictionary' function under the Data Set Options tab. If the underlying data is constantly changing, it may be worth enabling the 'Generate Match Dictionary every time this Data Set is used' option.

Paribus *Discovery* – Troubleshooting

This section provides some assistance in identifying common issues and providing potential resolutions.

Unable to Connect to my Paribus Control Database

If you are unable to connect to your Paribus Control Database, this may be due to one of the following:

- The SQL Server service on which the Control Database is located is not running. Start the SQL Server service.
- The Control Database does not exist on the selected SQL Server. Create a Control Database on the designated server, or connect to the server that is hosting the required database (a Paribus *Discovery* Server Licence will be required for this server)
- Maximum number of Paribus Control Database connections exceeded. Each Paribus *Discovery* Server Licence defines the maximum number of connections that can be made by any Paribus installation to a given SQL Server instance. Once this has been reached, no further connection can be made by Paribus *Discovery* to the given server. Either install a new Paribus *Discovery* Server Licence with an increased number of connections or disconnect from the SQL Server from one or more Paribus *Discovery* installations.
- Paribus *Discovery* requires Microsoft Data Access Components 2.6 to run correctly. Make sure this is installed on the client machine.

I don't get any Matches when I run my Match Session

If you obtain no match results on running a Match Session, one or more of the following may be the cause:

- Your Session match criteria are too stringent. Try setting the Match Threshold of the Session to a lower threshold value or disabling one or more Match Conditions in the Session.
- The information contained within the Match Dictionary is no longer valid, so nothing matches against it. A solution to this issue is to edit the Data Set used by the Match Set, and delete the Match Dictionary (Options tab).
- Of course, this may be due to there being no matches or duplicates in the underlying data. If filters have been applied, is the criteria too narrow?

Paribus Discovery – Technical Support and Troubleshooting

In the interest of providing customers with the latest product support information and troubleshooting assistance, QGate provides the following online resources within our QGate KnowledgeBase website:

General Support

For general technical support information, please visit the main product support page:

www.QGate.co.uk/knowledge/paribus-discovery

Troubleshooting

For troubleshooting common problems and scenarios relating to Paribus Discovery, please visit the troubleshooting page:

<http://www.QGate.co.uk/knowledge/paribus-discovery/troubleshooting>

System Requirements

For information regarding the minimum system requirements for installing and running Paribus Discovery, please visit the product system requirements page:

<http://www.QGate.co.uk/knowledge/paribus-discovery/sysregs>

Product Information

For more information about Paribus Discovery or other products from QGate Software, see one of the QGate web sites:

EMEA: www.QGate.co.uk

Americas: www.QGateSoftware.com

Fault Reporting

If you wish to report a fault or an issue with Paribus Discovery, please contact your Paribus Discovery software supplier.

Upgrades and Service Releases

Revision upgrades and service releases of Paribus Discovery are available on request from your Paribus Discovery software supplier.

Warning: We strongly recommend you review the release information provided with each release, to ensure each upgrade is suitable to apply. If you are in any doubt, we recommend you contact your Paribus Discovery software supplier.

All major version release upgrades will require the purchase of a new licence.

Systems Integration

For advice on integrating Paribus Discovery into other products and applications, please contact your Paribus Discovery software supplier.