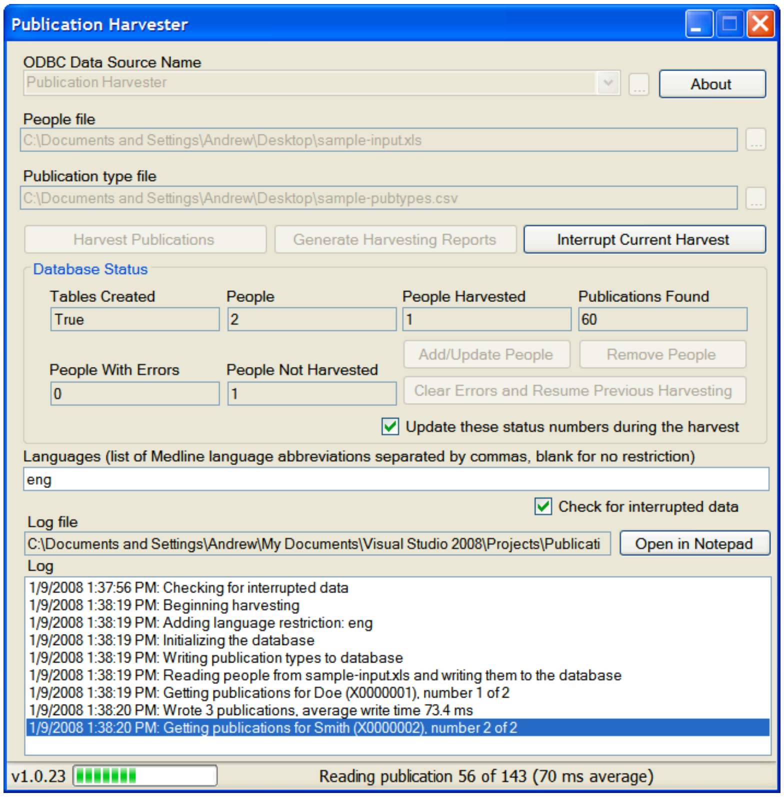
Publication Harvester

User Manual, v1.0.23



ODBC Data Source Name: Publication Harvester

People file: C:\Documents and Settings\Andrew\Desktop\sample-input.xls

Publication type file: C:\Documents and Settings\Andrew\Desktop\sample-pubtypes.csv

Tables Created	People	People Harvested	Publications Found
True	2	1	60

People With Errors	People Not Harvested
0	1

☑ Update these status numbers during the harvest

Languages (list of Medline language abbreviations separated by commas, blank for no restriction): eng

☑ Check for interrupted data

Log file: C:\Documents and Settings\Andrew\My Documents\Visual Studio 2008\Projects\Publicati

Log

```
1/9/2008 1:37:56 PM: Checking for interrupted data
1/9/2008 1:38:19 PM: Beginning harvesting
1/9/2008 1:38:19 PM: Adding language restriction: eng
1/9/2008 1:38:19 PM: Initializing the database
1/9/2008 1:38:19 PM: Writing publication types to database
1/9/2008 1:38:19 PM: Reading people from sample-input.xls and writing them to the database
1/9/2008 1:38:19 PM: Getting publications for Doe (X0000001), number 1 of 2
1/9/2008 1:38:20 PM: Wrote 3 publications, average write time 73.4 ms
1/9/2008 1:38:20 PM: Getting publications for Smith (X0000002), number 2 of 2
```

v1.0.23 Reading publication 56 of 143 (70 ms average)

Table of Contents

1 Introduction

1.1 Purpose

The purpose of this document is to serve as a guide to people who want to use the Publication Harvester software. It should give them all of the information necessary to install, configure and use the software.

1.2 Scope

This document contains step-by-step instructions to show users how to install, configure and use the Publication Harvester software on a machine running Windows XP. It covers:

- Downloading and installing the .NET Framework 2.0
- Downloading and installing MySQL 5.0
- Downloading and configuring MySQL Connector/ODBC
- Creating a new MySQL database
- Creating an ODBC data source
- Installing the Publication Harvester software
- Preparing the input files
- Using the Publication Harvester to retrieve publications from PubMed
- Using the Publication Harvester to generate reports

1.3 System Overview

The purpose of the Publication Harvester is to generate an accurate count of publications for a set of people, using a set of possible name variations for that individual, and recording author position carefully.

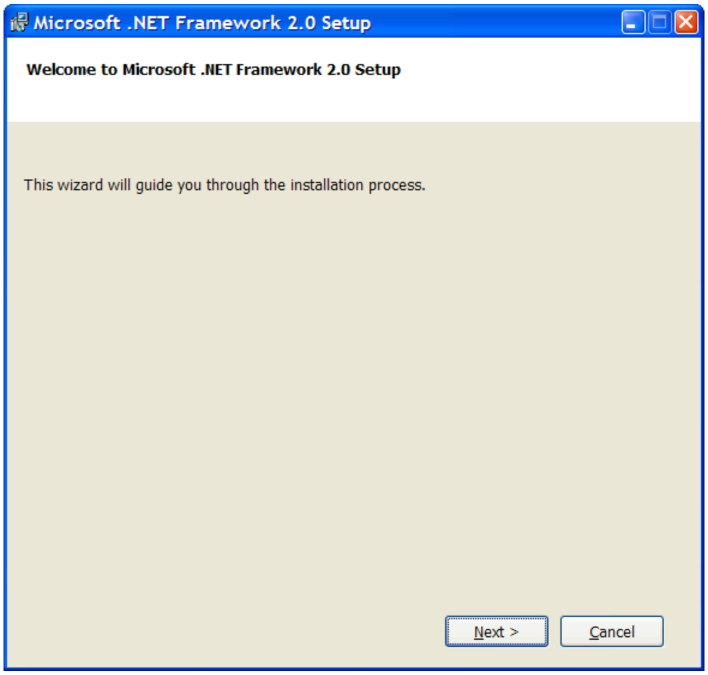
2 Installation

This section describes how to install the Publication Harvester software. It covers:

- Downloading and installing the .NET Framework 2.0
- Downloading and installing MySQL 5.0
- Downloading and configuring MySQL Connector/ODBC
- Creating a new MySQL database
- Creating an ODBC data source
- Installing the Publication Harvester software

2.1 Download and Install the .NET Framework 2.0

The Publication Harvester requires the .NET Framework 2.0, which can be downloaded from the Microsoft website. Go to http://www.microsoft.com/downloads/details.aspx?FamilyID=0856EACB-4362-4B0D-8EDD-AAB15C5E04F5 and click on "Download" to download the file **dotnetfx.exe**. Save this file to disk and double-click on it. The following window will pop up:

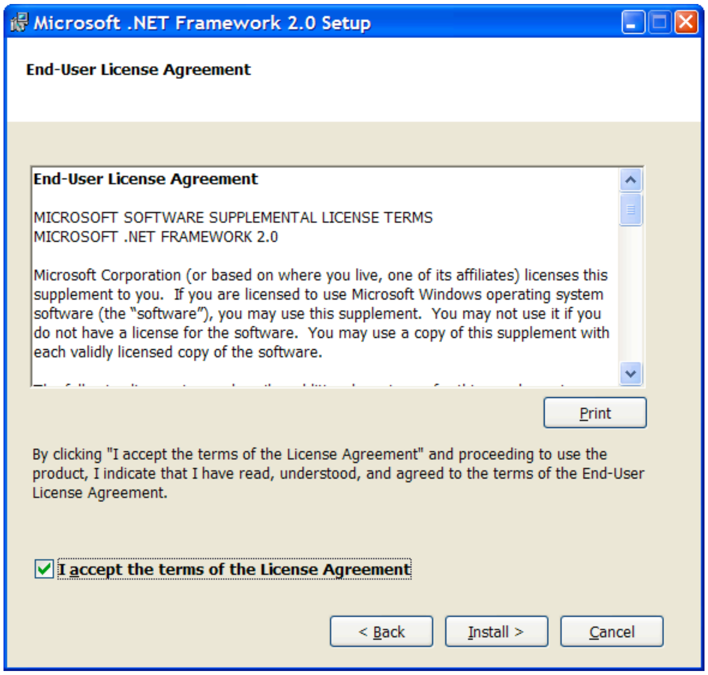


Microsoft .NET Framework 2.0 Setup

Welcome to Microsoft .NET Framework 2.0 Setup

This wizard will guide you through the installation process.

Click "Next >".


Microsoft .NET Framework 2.0 Setup

End-User License Agreement

End-User License Agreement

MICROSOFT SOFTWARE SUPPLEMENTAL LICENSE TERMS
MICROSOFT .NET FRAMEWORK 2.0

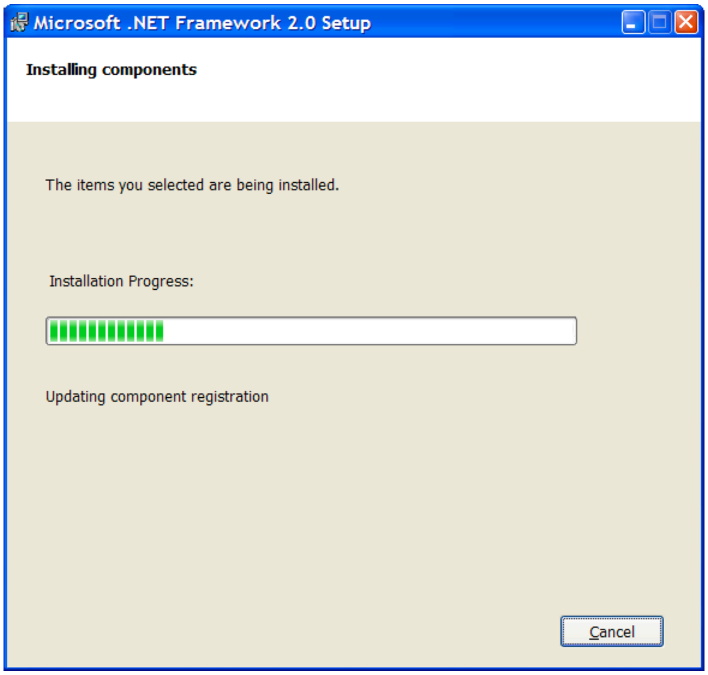
Microsoft Corporation (or based on where you live, one of its affiliates) licenses this supplement to you. If you are licensed to use Microsoft Windows operating system software (the "software"), you may use this supplement. You may not use it if you do not have a license for the software. You may use a copy of this supplement with each validly licensed copy of the software.

By clicking "I accept the terms of the License Agreement" and proceeding to use the product, I indicate that I have read, understood, and agreed to the terms of the End-User License Agreement.

☑ **I accept the terms of the License Agreement**

Check the checkbox and click "Install >".

The software will install automatically:


Microsoft .NET Framework 2.0 Setup

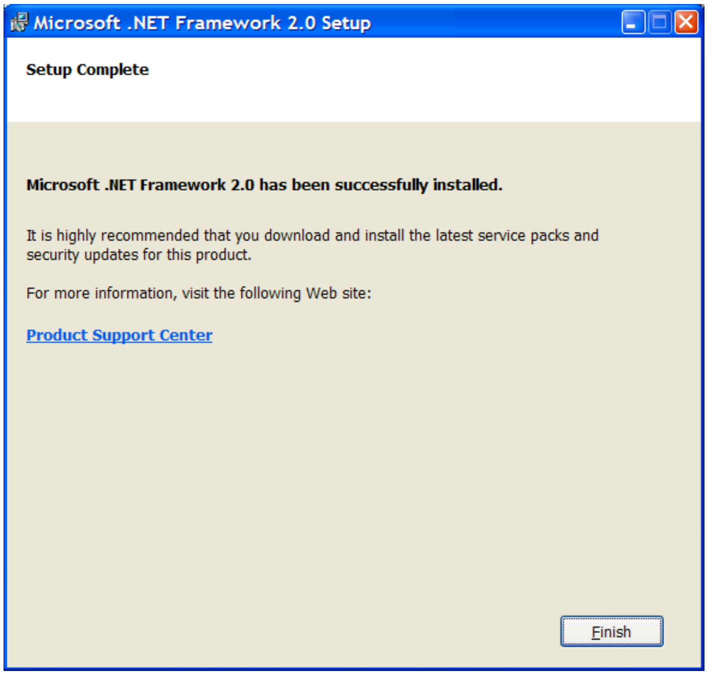
Installing components

The items you selected are being installed.

Installation Progress:

Updating component registration

When the installation is complete, the following window will pop up:



Microsoft .NET Framework 2.0 Setup

Setup Complete

Microsoft .NET Framework 2.0 has been successfully installed.

It is highly recommended that you download and install the latest service packs and security updates for this product.

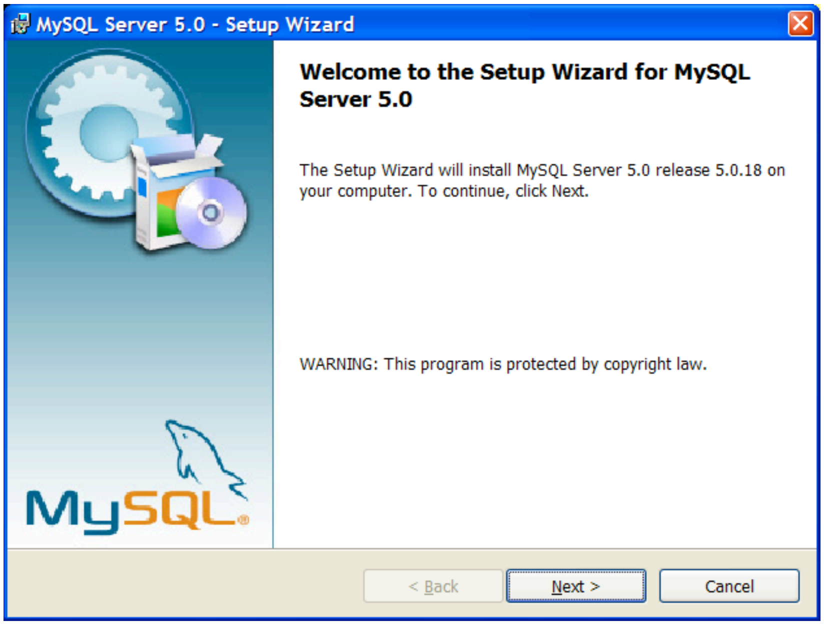
For more information, visit the following Web site:

Product Support Center

Click “Finish”.

2.2 Download and Install MySQL 5.0

The Publication Harvester software uses MySQL 5.0 to store its harvested data. MySQL is a free and open source database. The latest version can be found here: http://dev.mysql.com/downloads/. To download the latest version of MySQL 5.0, go to that page and follow the link labeled “Generally Available (GA) release (recommended).” Find the “Download” link labeled “Windows Essentials (x86)”. Download the file (it will end with .MSI) and double-click on it. The following window pop up:



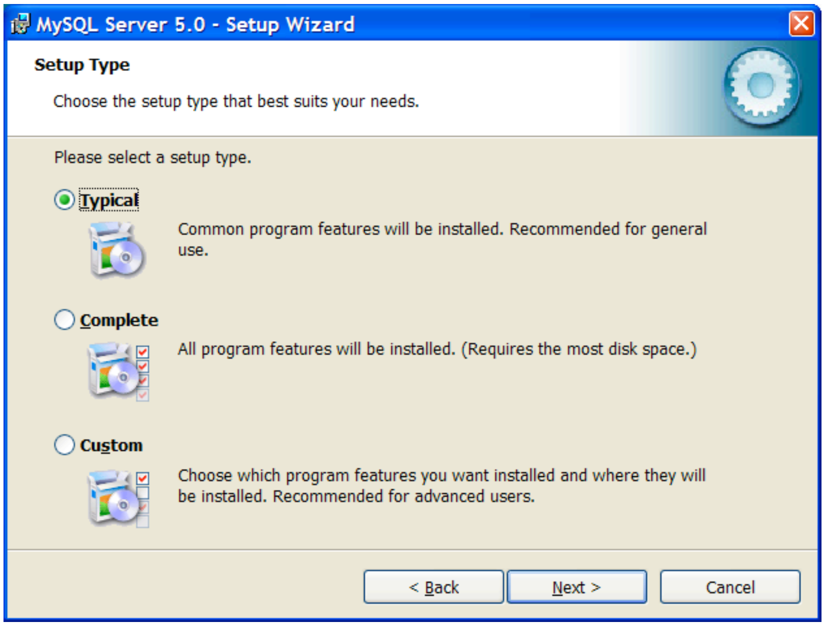
MySQL Server 5.0 - Setup Wizard

Welcome to the Setup Wizard for MySQL Server 5.0

The Setup Wizard will install MySQL Server 5.0 release 5.0.18 on your computer. To continue, click Next.

WARNING: This program is protected by copyright law.

Click “Next >”. The following window will pop up:



Setup Type

Choose the setup type that best suits your needs.

Please select a setup type.

● **Typical**

Common program features will be installed. Recommended for general use.

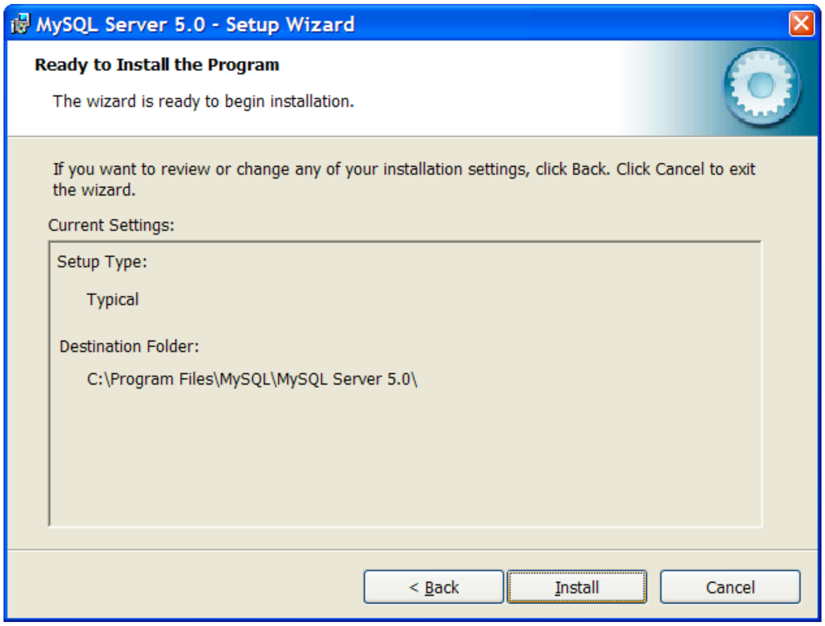
○ **Complete**

All program features will be installed. (Requires the most disk space.)

○ **Custom**

Choose which program features you want installed and where they will be installed. Recommended for advanced users.

Make sure that "Typical" is selected and click "Next >".



Ready to Install the Program

The wizard is ready to begin installation.

If you want to review or change any of your installation settings, click Back. Click Cancel to exit the wizard.

Current Settings:

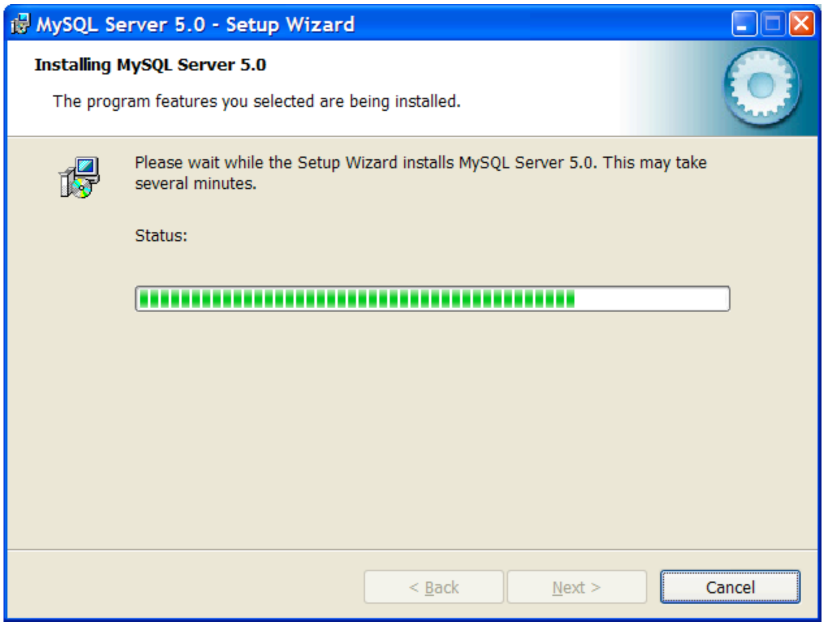
Setup Type:

Typical

Destination Folder:

C:\Program Files\MySQL\MySQL Server 5.0\

This will begin the MySQL Server 5.0 installation:



MySQL Server 5.0 - Setup Wizard

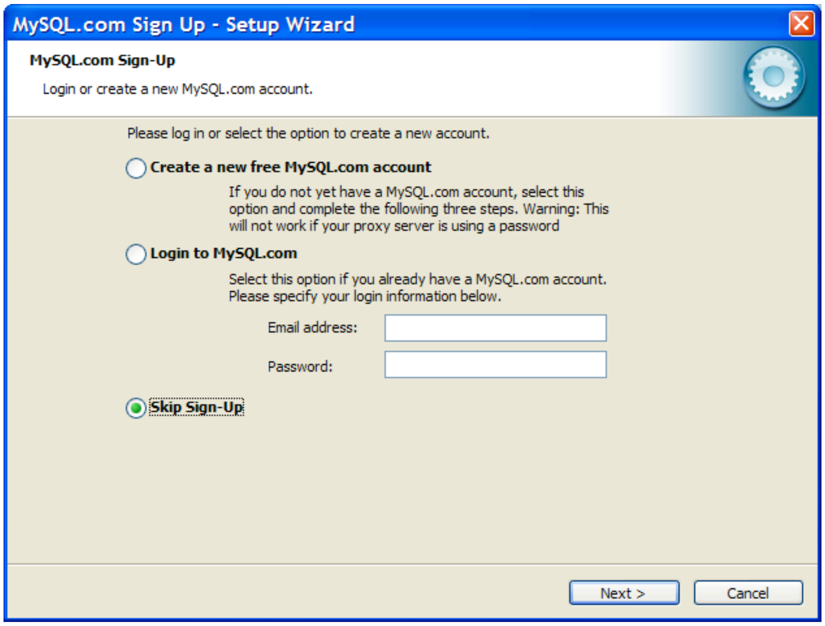
Installing MySQL Server 5.0

The program features you selected are being installed.

Please wait while the Setup Wizard installs MySQL Server 5.0. This may take several minutes.

Status:

When the installation is done, the software will display this window:



MySQL.com Sign Up - Setup Wizard

MySQL.com Sign-Up

Login or create a new MySQL.com account.

Please log in or select the option to create a new account.

○ **Create a new free MySQL.com account**

If you do not yet have a MySQL.com account, select this option and complete the following three steps. Warning: This will not work if your proxy server is using a password

○ **Login to MySQL.com**

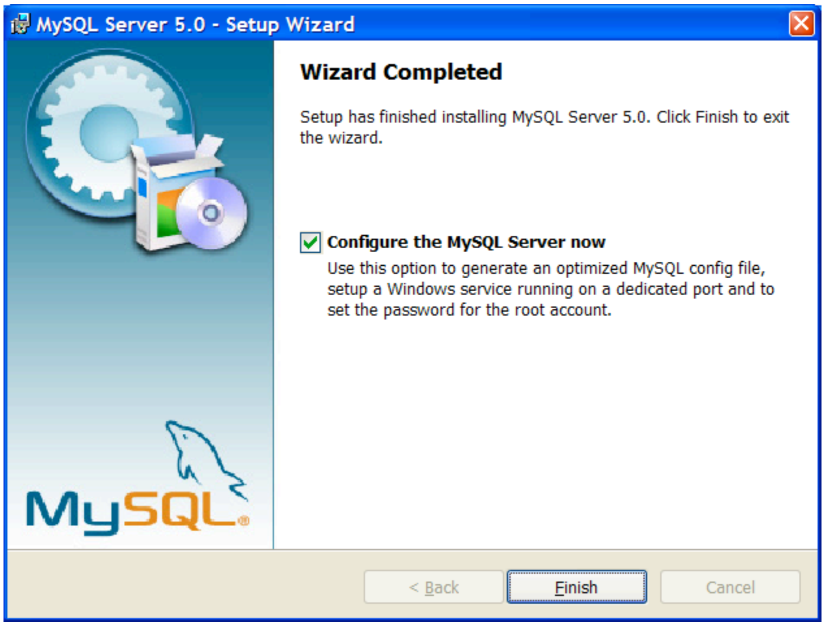
Select this option if you already have a MySQL.com account. Please specify your login information below.

Email address:

Password:

● **Skip Sign-Up**

Click “Skip Sign-Up” and select “Next >”.



MySQL Server 5.0 - Setup Wizard

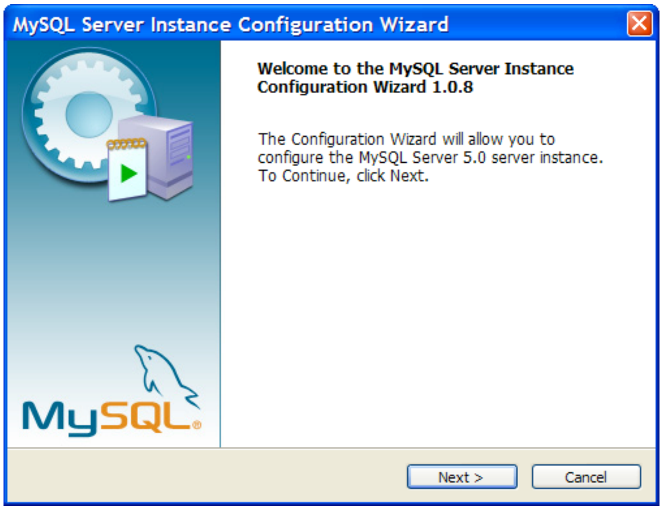
Wizard Completed

Setup has finished installing MySQL Server 5.0. Click Finish to exit the wizard.

☑ **Configure the MySQL Server now**

Use this option to generate an optimized MySQL config file, setup a Windows service running on a dedicated port and to set the password for the root account.

Make sure that the “Configure the MySQL Server now” box is checked and click “Finish”:

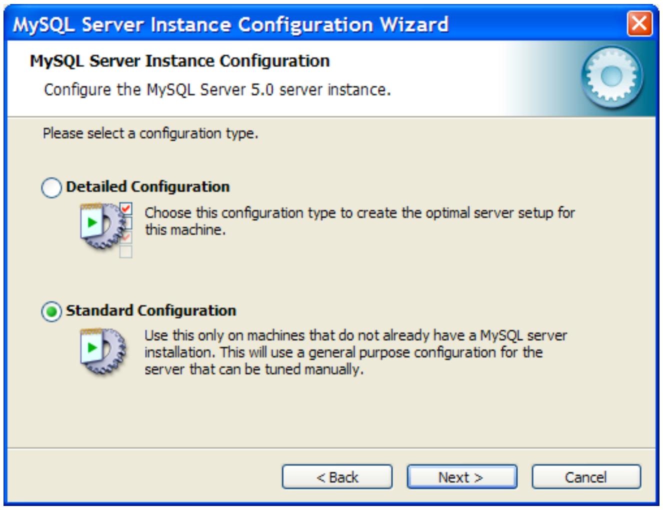


MySQL Server Instance Configuration Wizard

Welcome to the MySQL Server Instance Configuration Wizard 1.0.8

The Configuration Wizard will allow you to configure the MySQL Server 5.0 server instance. To Continue, click Next.

Click “Next >”:



MySQL Server Instance Configuration Wizard

MySQL Server Instance Configuration

Configure the MySQL Server 5.0 server instance.

Please select a configuration type.

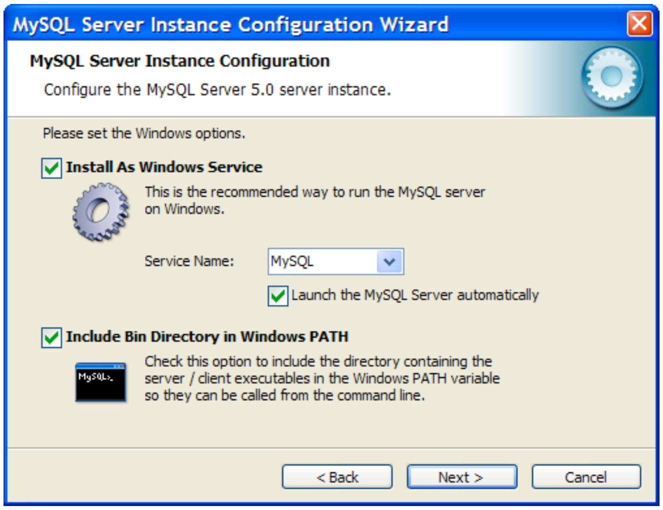
○ **Detailed Configuration**

Choose this configuration type to create the optimal server setup for this machine.

● **Standard Configuration**

Use this only on machines that do not already have a MySQL server installation. This will use a general purpose configuration for the server that can be tuned manually.

Select "Standard Configuration" and click "Next >":



MySQL Server Instance Configuration Wizard

MySQL Server Instance Configuration

Configure the MySQL Server 5.0 server instance.

Please set the Windows options.

☑ **Install As Windows Service**

This is the recommended way to run the MySQL server on Windows.

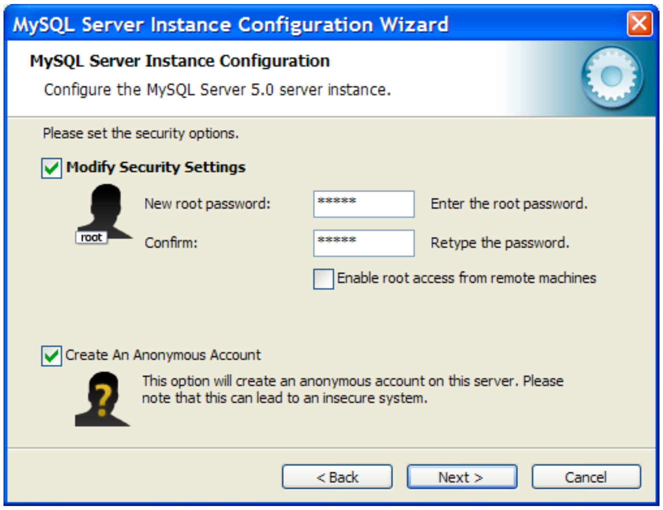
Service Name: MySQL

☑ Launch the MySQL Server automatically

☑ **Include Bin Directory in Windows PATH**

Check this option to include the directory containing the server / client executables in the Windows PATH variable so they can be called from the command line.

Check both boxes and select "Next >":



MySQL Server Instance Configuration Wizard

MySQL Server Instance Configuration

Configure the MySQL Server 5.0 server instance.

Please set the security options.

☑ **Modify Security Settings**

New root password: ***** Enter the root password.

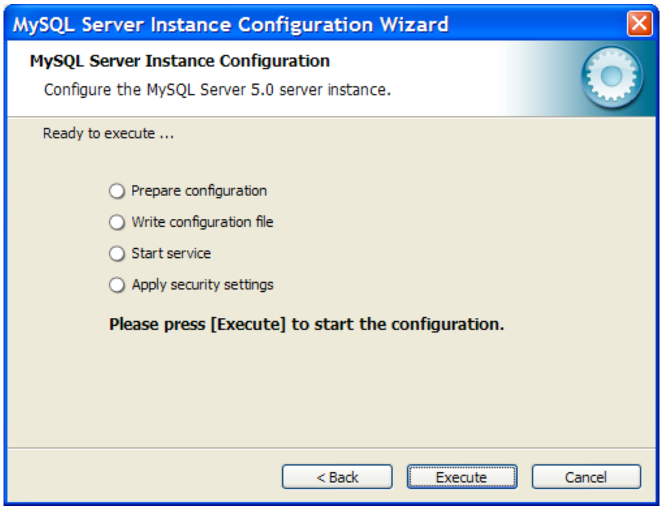
Confirm: ***** Retype the password.

□ Enable root access from remote machines

☑ Create An Anonymous Account

This option will create an anonymous account on this server. Please note that this can lead to an insecure system.

You must now enter a password for your MySQL server. Enter the password in both text boxes, and make sure that both the "Modify Security Settings" and "Create An Anonymous Account" checkboxes are checked. Click "Next >":



MySQL Server Instance Configuration Wizard

MySQL Server Instance Configuration

Configure the MySQL Server 5.0 server instance.

Ready to execute ...

○ Prepare configuration

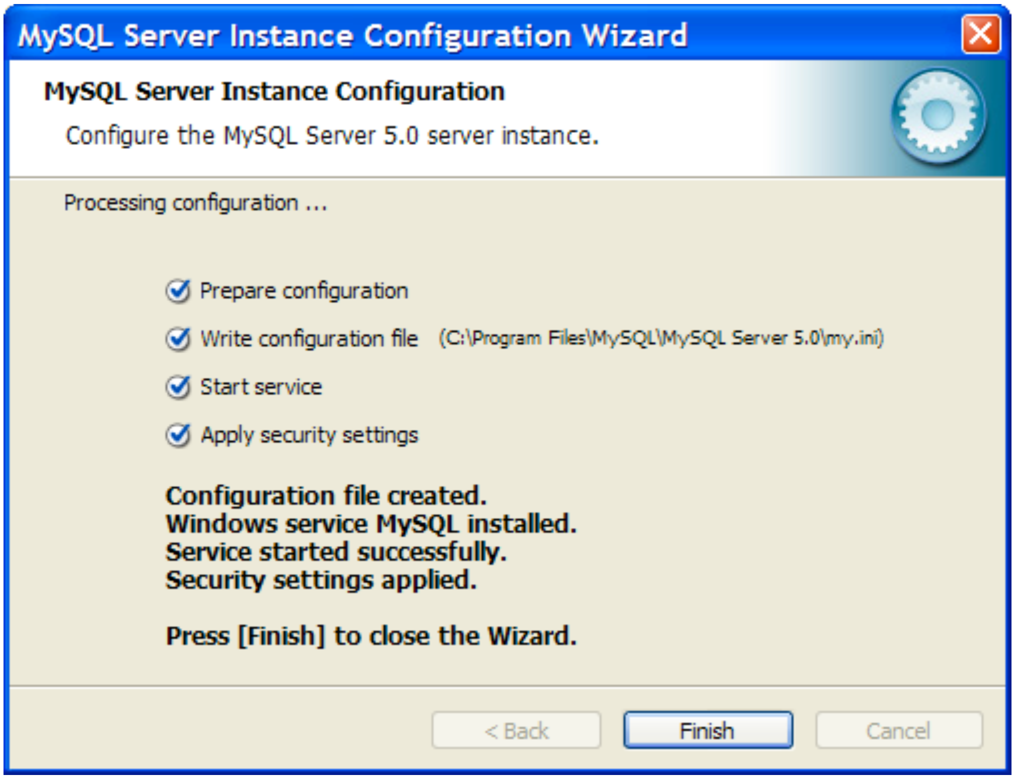
○ Write configuration file

○ Start service

○ Apply security settings

Please press [Execute] to start the configuration.

Click “Execute” to configure the MySQL server:



MySQL Server Instance Configuration Wizard
MySQL Server Instance Configuration
Configure the MySQL Server 5.0 server instance.

Processing configuration ...

- ☑ Prepare configuration
- ☑ Write configuration file (C:\Program Files\MySQL\MySQL Server 5.0\my.ini)
- ☑ Start service
- ☑ Apply security settings

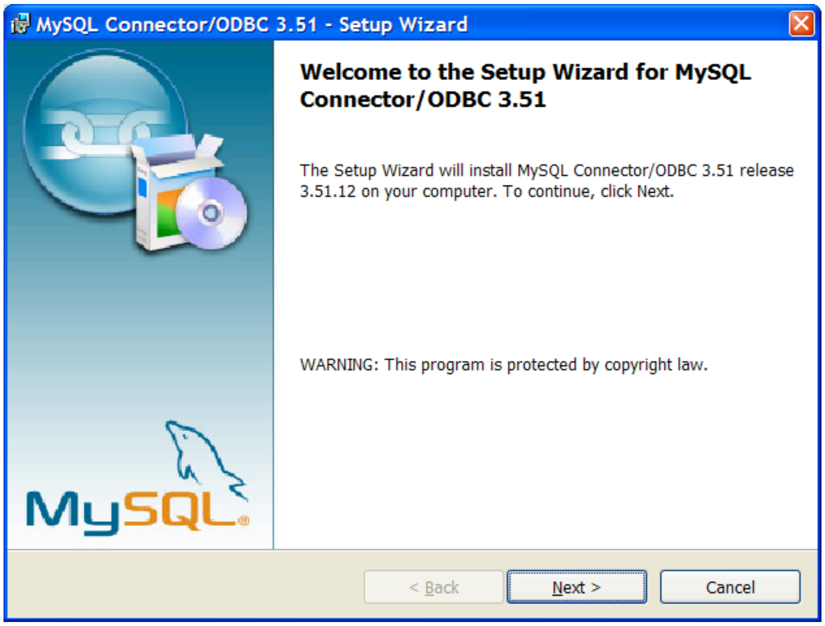
Configuration file created.
Windows service MySQL installed.
Service started successfully.
Security settings applied.

Press [Finish] to close the Wizard.

Click “Finish”. MySQL 5.0 is now installed.

2.3 Download and configure MySQL Connector/ODBC

The Publication Harvester uses a piece of software called the Connector/ODBC. To install this, go to http://www.mysql.com/downloads and look for the link labeled “Connector/ODBC 3.51 -- Generally Available (GA) release” (the version number may be different). Click on that link. That page will have a section labeled “Windows Downloads”. Find the row in that section labeled “Driver Installer (MSI)” and click on the “Pick a mirror” link. Scroll down this page and click on any of the links labeled “HTTP” or “FTP”. This will link to a file that ends with “.msi”. Save that file and double-click on it. The following window will pop up:

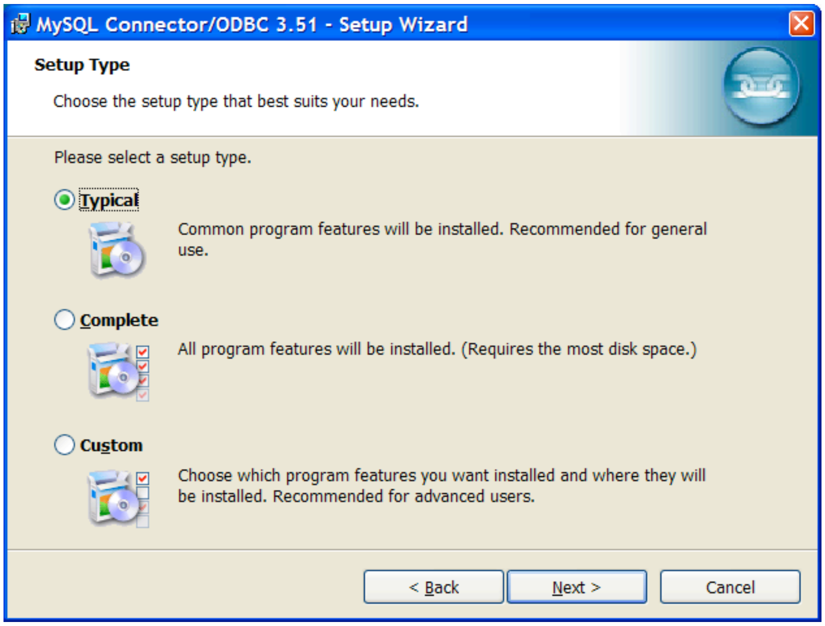


Welcome to the Setup Wizard for MySQL Connector/ODBC 3.51

The Setup Wizard will install MySQL Connector/ODBC 3.51 release 3.51.12 on your computer. To continue, click Next.

WARNING: This program is protected by copyright law.

Click "Next >". The following window will pop up:



Setup Type

Choose the setup type that best suits your needs.

Please select a setup type.

● Typical

Common program features will be installed. Recommended for general use.

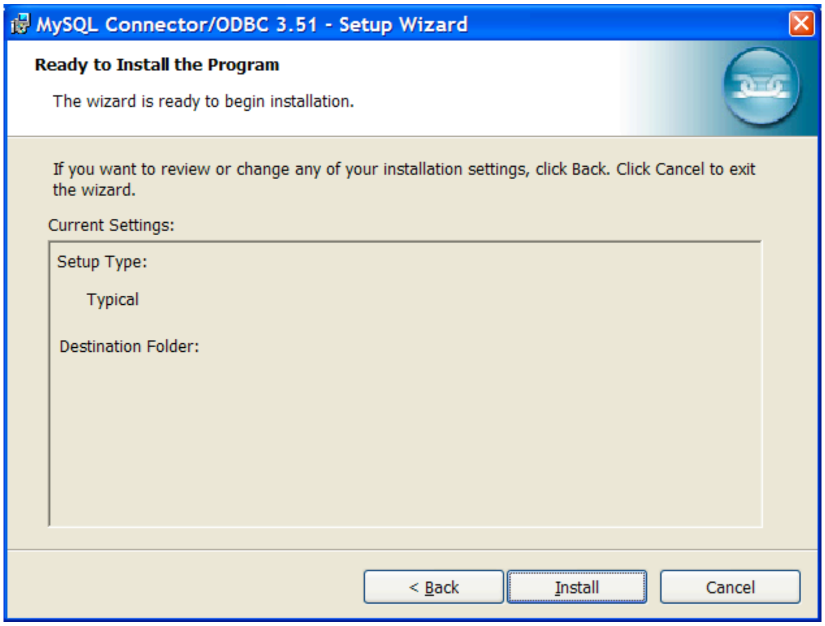
○ Complete

All program features will be installed. (Requires the most disk space.)

○ Custom

Choose which program features you want installed and where they will be installed. Recommended for advanced users.

Select "Typical" and click "Next >":



MySQL Connector/ODBC 3.51 - Setup Wizard

Ready to Install the Program

The wizard is ready to begin installation.

If you want to review or change any of your installation settings, click Back. Click Cancel to exit the wizard.

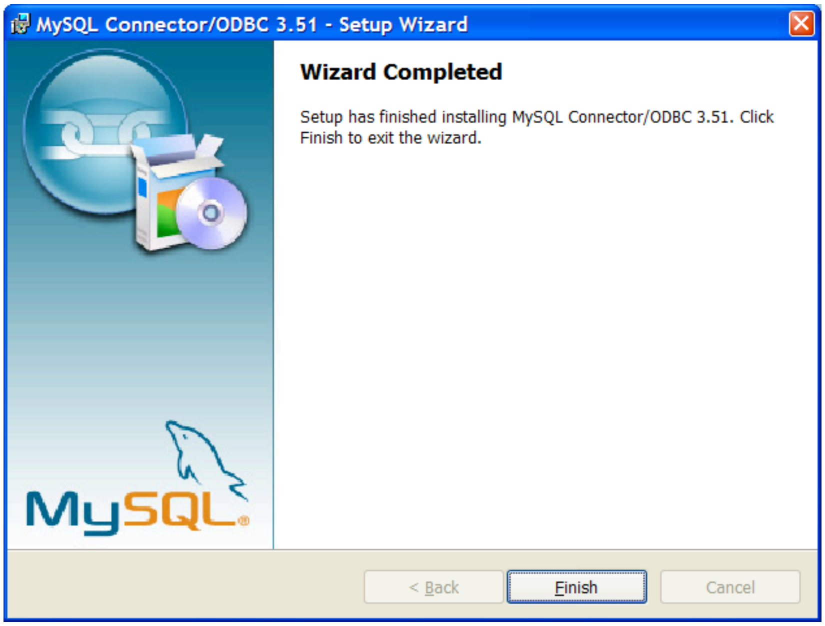
Current Settings:

Setup Type:

Typical

Destination Folder:

Click "Install". When the installation completes, the following window will pop up:



MySQL Connector/ODBC 3.51 - Setup Wizard

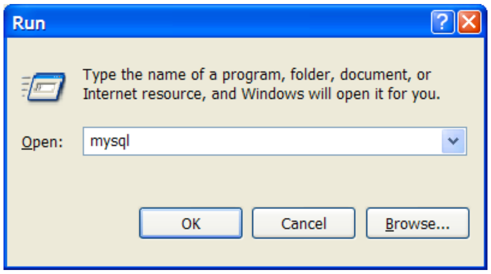
Wizard Completed

Setup has finished installing MySQL Connector/ODBC 3.51. Click Finish to exit the wizard.

Click "Finish." The MySQL Connector/ODBC is now installed.

2.4 Create a new MySQL database

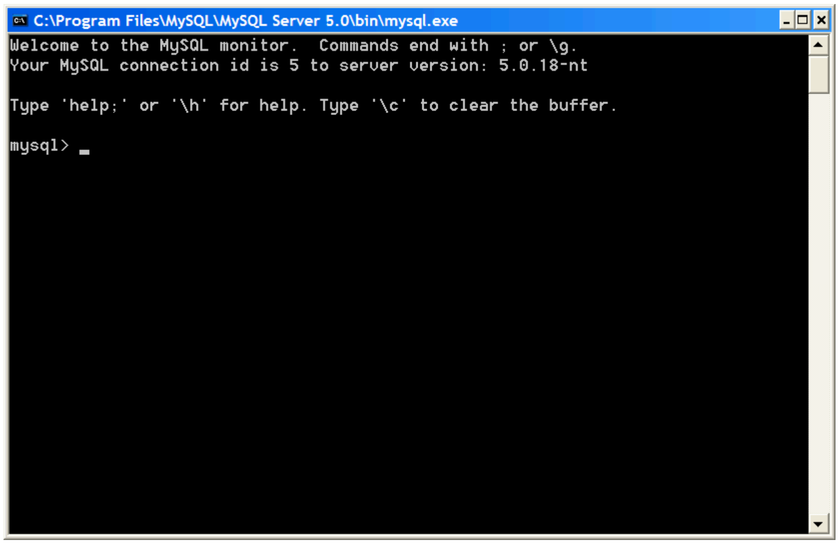
The Publication Harvester requires that you create a MySQL database for it to use. You will give this database a name. It doesn't matter what name you give it. In this example, we'll use the name "Publication Harvester". To create the database, first select "Run" from the Windows "Start" menu:



Type the name of a program, folder, document, or Internet resource, and Windows will open it for you.

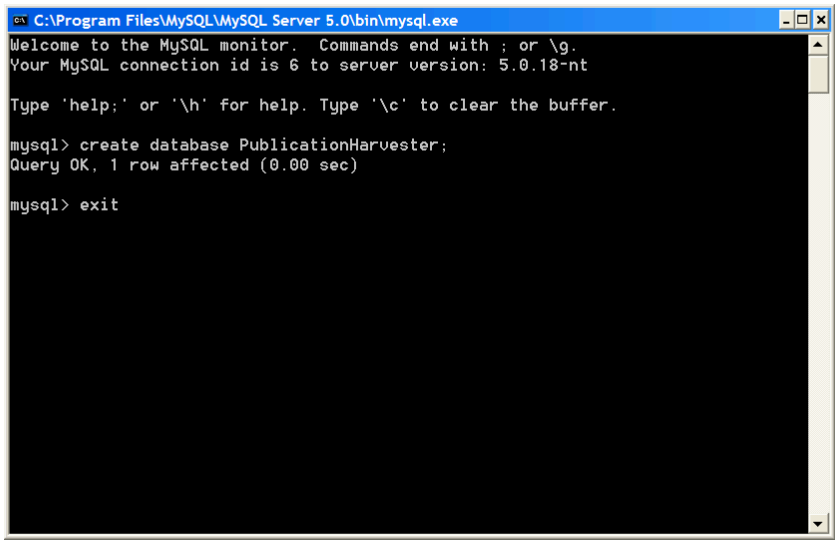
Open: mysql

Type “mysql” and click OK. The following window will pop up:



```
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 5 to server version: 5.0.18-nt

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

mysql> _
```

Type “create database PublicationHarvester;” and press enter. The database is now created.



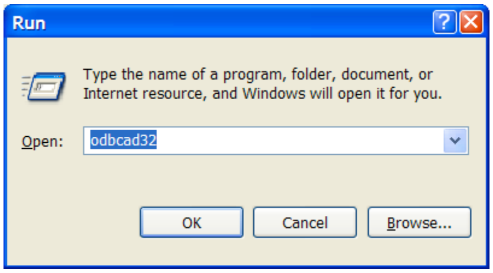
```
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 6 to server version: 5.0.18-nt

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

mysql> create database PublicationHarvester;
Query OK, 1 row affected (0.00 sec)

mysql> exit
```

Type “exit” and press enter to close the window.

2.5 Create an ODBC Data Source

The Publication Harvester connects to the MySQL database using a feature of Windows called Open Database Connectivity, or ODBC. ODBC allows you to store information about how to connect with a database. Before you can run the Publication Harvester, you must create an ODBC data source using the ODBC Data Source Administrator.

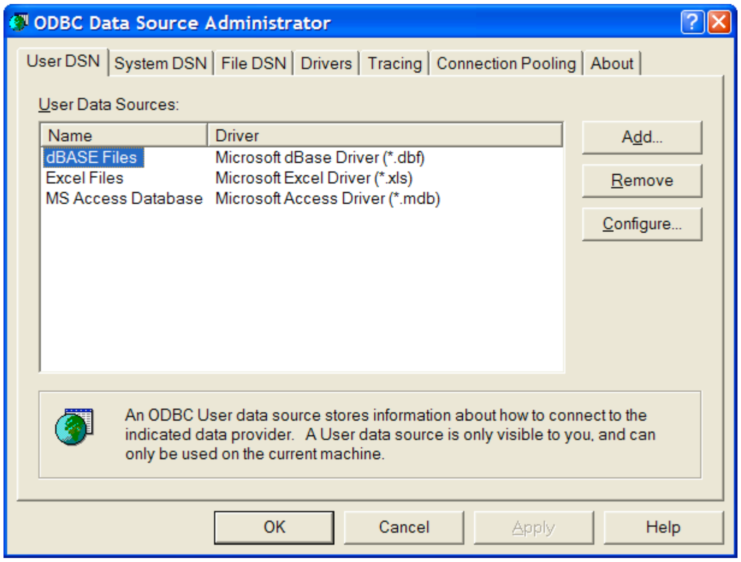
Select “Run” from the Windows Start menu:



Type the name of a program, folder, document, or Internet resource, and Windows will open it for you.

Open: odbcad32

Enter “odbcad32” and click “OK”. The following window will pop up:

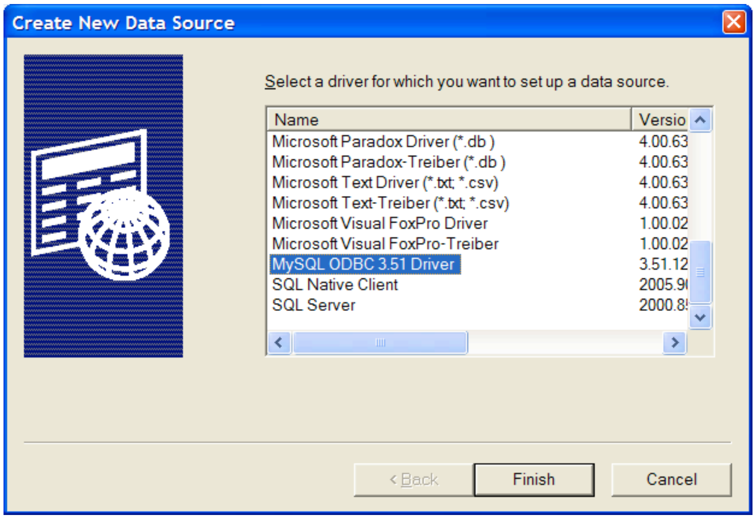


User Data Sources:

Name	Driver
dBASE Files	Microsoft dBase Driver (*.dbf)
Excel Files	Microsoft Excel Driver (*.xls)
MS Access Database	Microsoft Access Driver (*.mdb)

An ODBC User data source stores information about how to connect to the indicated data provider. A User data source is only visible to you, and can only be used on the current machine.

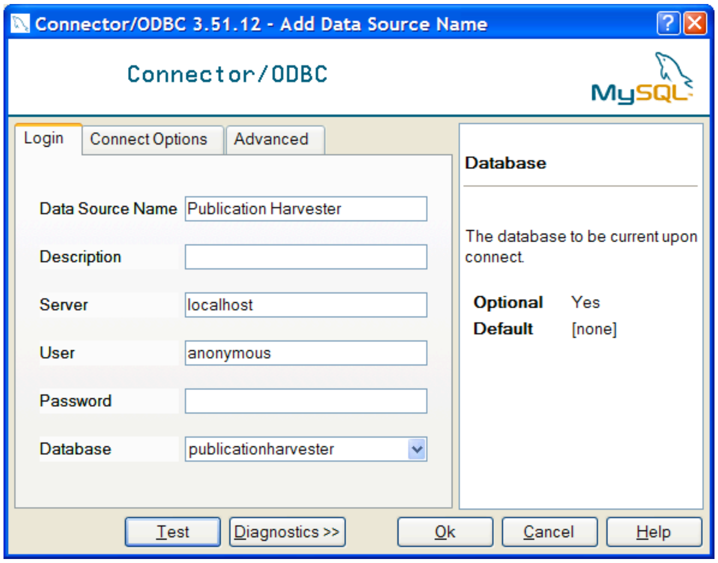
Click the “Add…” button:



Select a driver for which you want to set up a data source.

Name	Versio
Microsoft Paradox Driver (*.db)	4.00.63
Microsoft Paradox-Treiber (*.db)	4.00.63
Microsoft Text Driver (*.txt; *.csv)	4.00.63
Microsoft Text-Treiber (*.txt; *.csv)	4.00.63
Microsoft Visual FoxPro Driver	1.00.02
Microsoft Visual FoxPro-Treiber	1.00.02
MySQL ODBC 3.51 Driver	3.51.12
SQL Native Client	2005.9
SQL Server	2000.8

Scroll down to “MySQL ODBC 3.51 Driver” and click “Finish”. The following window will pop up:



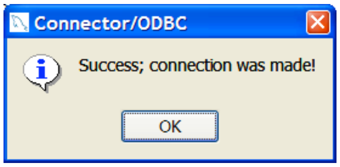
Data Source Name	Publication Harvester
Description	
Server	localhost
User	anonymous
Password	
Database	publicationharvester

Database

The database to be current upon connect.

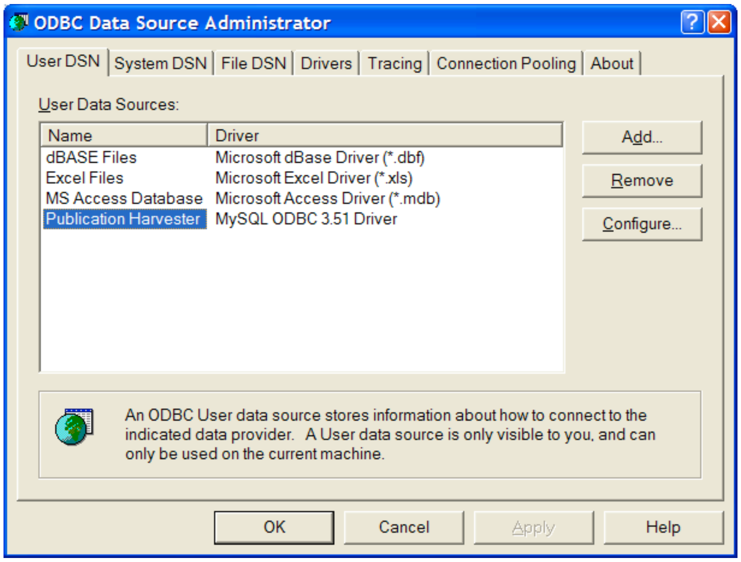
Optional	Yes
Default	[none]

Enter “Publication Harvester” as the data source name, “localhost” as the server, and “anonymous” as the password. Click the down-arrow at the right-hand side of the database box to display the list of databases. The “PublicationHarvester” database that you created in section 2.4 should be in the list (in all lowercase). Select it. Click “Test” to make sure the database works. The following window should pop up:



Success; connection was made!

Click “Ok”. The ODBC data source has now been created, and should show up in this list:



User Data Sources:

Name	Driver
dBASE Files	Microsoft dBase Driver (*.dbf)
Excel Files	Microsoft Excel Driver (*.xls)
MS Access Database	Microsoft Access Driver (*.mdb)
Publication Harvester	MySQL ODBC 3.51 Driver

An ODBC User data source stores information about how to connect to the indicated data provider. A User data source is only visible to you, and can only be used on the current machine.

Click “OK” to close the window.

2.6 Install the Publication Harvester

The Publication Harvester installer can be downloaded from http://www.stellman-greene.com/PublicationHarvester/bin/latest/. Download the ZIP file that contains the installer for the latest version, extract it, and run setup.exe. This will install the program into the start menu – it will be listed under “Publication Harvester”.

In addition, there are three sample files that can be downloaded from the Publication Harvester website to help you get started:

- sample-input.xls -- sample input People file
- sample-JIFs.xls -- sample JIF file for generating reports
- sample-pubtypes.csv -- sample publication types file

The Publication Harvester program will pop up as soon as the installer completes. You can run it again at any time by selecting it from the Start menu.

3 Using the Publication Harvester

Now that the Publication Harvester has been installed, it can be used to retrieve publications. To do this, a set of input files must be prepared. Then those input files can be used to tell the software whose publications to retrieve and what to do with them. Once the publications have been retrieved, the software can generate reports.

3.1 Prepare the Input Files

The Publication Harvester uses two input files: the people file, which contains a list of people to retrieve; and the publication type categories file, which divides the publication types into a set of categories so articles can be classified by publication type.

3.1.1 People File

The People File contains a list of people, and information which tells the software how to retrieve the data from PubMed for those people. It is provided as a Microsoft Excel 8.0 file, with the first row containing column headings. The file contains the following columns:

- setnb (text): The unique identifier for the person
- first (text): The person's first name
- middle (text): The person's middle name or initial [may be blank]
- last (text): The person's last name
- name1 (text): The PubMed-formatted name which will appear in the author list of a publication returned by an NCBI query. Only publications that have this name (or the name in column name2, name3 or name4) will be added to the Publications table.
- name2 (text, optional): Another PubMed-formatted name. If more than one name is provided, the software will look for publications which match any of the names.
- name3 (text, optional): PubMed-formatted name
- name4 (text, optional): PubMed-formatted name
- medline_search1 (text): A search term which will be used to execute the PubMed search. For example:

```
("van eys j"[au] OR "vaneys j"[au] OR "eys jv"[au])

("tobian l"[au] OR "tobian l jr"[au] OR "tobian lj"[au])

(("reemtsma k"[au] OR "reemtsma kb"[au]) AND 1956:2000[dp])

("guillemin rc"[au] OR ("guillemin r"[au] NOT (Electrodiagn
Ther[ta] OR Phys Rev Lett[ta] OR vegas[ad] OR lindle[au])))
```

A sample people file can be downloaded from the following URL: http://www.stellman-greene.com/PublicationHarvester/sample-input.xls.

Note: In addition to an Excel file, the People file may also be CSV format – which is useful if you need to harvest publications for over 65,535 people, the maximum size of an Excel file.

3.1.2 Publication Type Categories File

The Publication Type Categories file is used by the software to divide articles into categories based on publication type. Each PubMed article has a publication type. The software must either discard each publication or populate the Publications.PubTypeCategoryID column based on that publication type. The PublicationTypeCategories.csv is a comma-delimited text file which the software uses to determine how to process the publication types.

PublicationTypeCategories.csv contains two columns:

- PublicationType (string): The publication type that appears in a PubMed article
- PubTypeCategoryID (text): This will typically be 1, 2, 3, 4 or 0. This contains the numeric category, or "bin," into which the software must classify the any article with the type specified in the PublicationType column. If this column contains 0, the software ignores any publication with the type specified in the PublicationType column.

A publication may contain several publication types. Normally, the Publication Harvester only reads the first publication type. However, there are some publication types (like "Review") that always occur as a second or third publication type. To specify that this category should override the first type in a citation, specify a negative publication type. So if the category "Review" should be given "bin" 2 but should always override the first publication type, then the publication types file should contain a value of "-2" for this category.

A sample people file can be downloaded from the following URL: http://www.stellman-greene.com/PublicationHarvester/sample-pubtypes.csv.

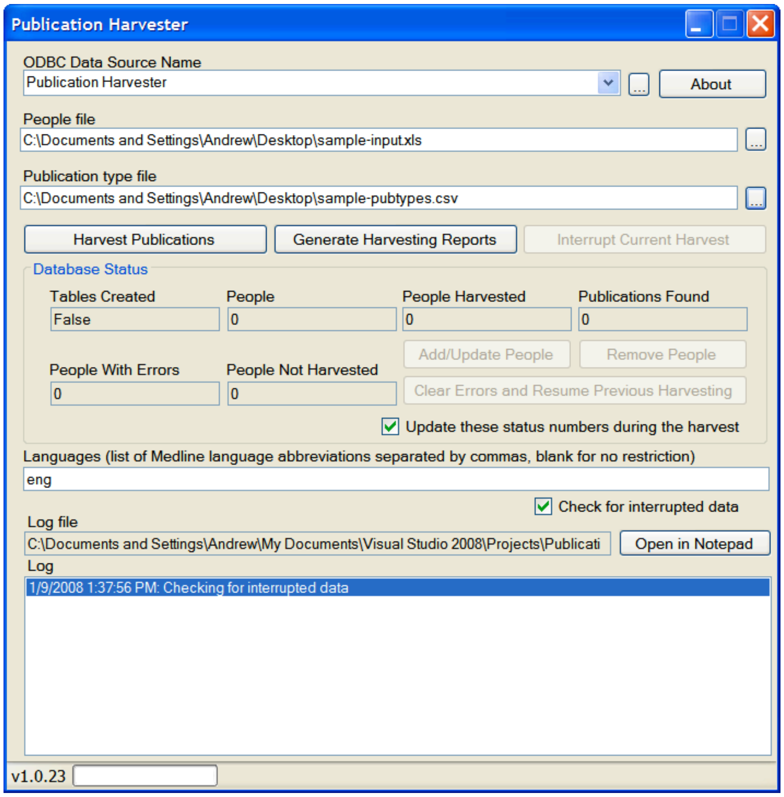
3.2 Retrieve Publications from PubMed

To retrieve publications from PubMed, first launch the Publication Harvester (by double-clicking on the PublicationHarvester program downloaded in the section 2.6). The main form of the Publication Harvester will pop up.

Click on the ODBC Data Source field and select the data source created in section 2.5. If you click the "…" button next to that field, it will pop up the ODBC Data Source Administrator.

Click on the "…" button next to the "People file" field and browse to the People file (see section 3.1.1). Then click on the "…" button next to the "Publication type file" and browse to the Publication Type Categories file (see section 3.1.2).

Once these forms are filled in, the Publication Harvester window will look like this:



ODBC Data Source Name
Publication Harvester

People file
C:\Documents and Settings\Andrew\Desktop\sample-input.xls

Publication type file
C:\Documents and Settings\Andrew\Desktop\sample-pubtypes.csv

Tables Created: False
People: 0
People Harvested: 0
Publications Found: 0
People With Errors: 0
People Not Harvested: 0

☑ Update these status numbers during the harvest

Languages (list of Medline language abbreviations separated by commas, blank for no restriction)
eng

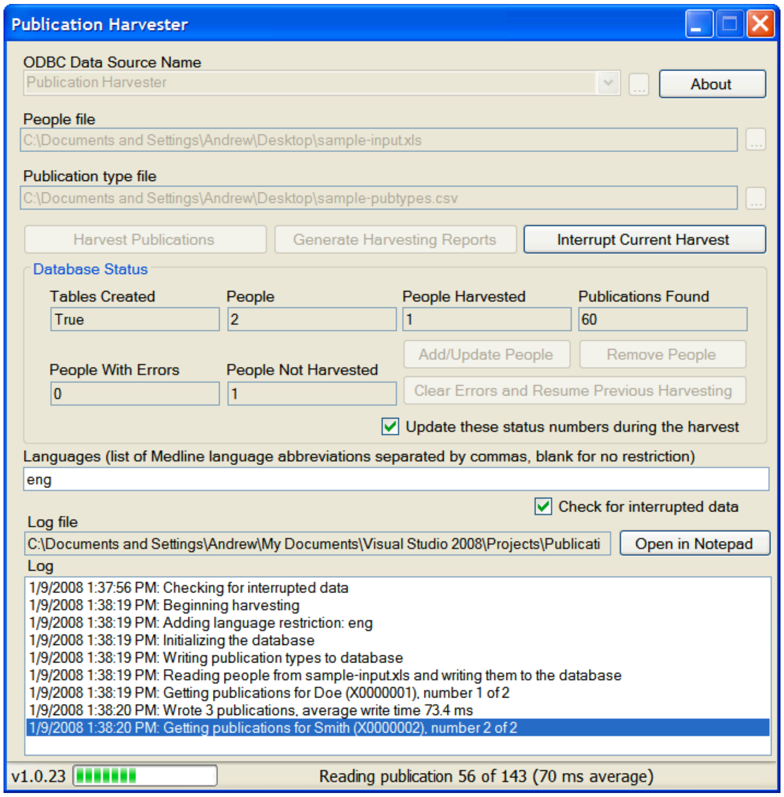
☑ Check for interrupted data

Log file
C:\Documents and Settings\Andrew\My Documents\Visual Studio 2008\Projects\Publicati

Log

```
1/9/2008 1:37:56 PM: Checking for interrupted data
```

To begin harvesting data, click the “Harvest Publications” button. The software will connect to the PubMed server and begin retrieving publications.



ODBC Data Source Name
Publication Harvester

People file
C:\Documents and Settings\Andrew\Desktop\sample-input.xls

Publication type file
C:\Documents and Settings\Andrew\Desktop\sample-pubtypes.csv

Tables Created: True
People: 2
People Harvested: 1
Publications Found: 60
People With Errors: 0
People Not Harvested: 1

☑ Update these status numbers during the harvest

Languages (list of Medline language abbreviations separated by commas, blank for no restriction)
eng

☑ Check for interrupted data

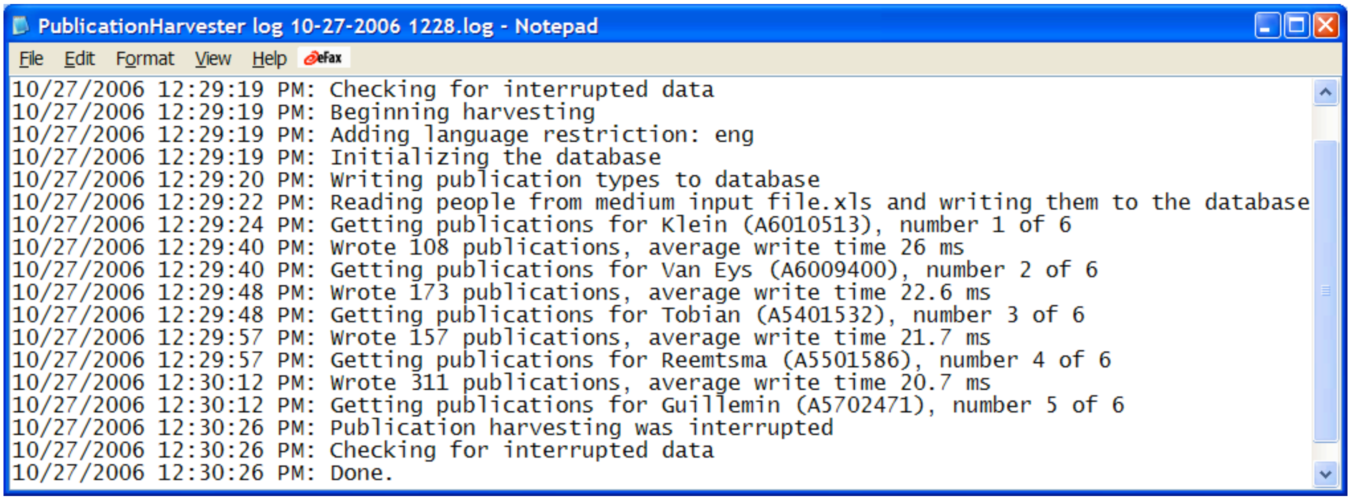
Log file
C:\Documents and Settings\Andrew\My Documents\Visual Studio 2008\Projects\Publicati

Log

```
1/9/2008 1:37:56 PM: Checking for interrupted data
1/9/2008 1:38:19 PM: Beginning harvesting
1/9/2008 1:38:19 PM: Adding language restriction: eng
1/9/2008 1:38:19 PM: Initializing the database
1/9/2008 1:38:19 PM: Writing publication types to database
1/9/2008 1:38:19 PM: Reading people from sample-input.xls and writing them to the database
1/9/2008 1:38:19 PM: Getting publications for Doe (X0000001), number 1 of 2
1/9/2008 1:38:20 PM: Wrote 3 publications, average write time 73.4 ms
1/9/2008 1:38:20 PM: Getting publications for Smith (X0000002), number 2 of 2
```

Reading publication 56 of 143 (70 ms average)

The software will run until it completes this harvesting step. Note that for very large databases (with over 50,000 people), performance may slow down due to the fact that it takes time to update the statistics in the "Database Status" box. Uncheck the "Update these status numbers during harvest" box to prevent this. (The box will not significantly speed anything up for small databases.) As it retrieves publications, it adds entries to the log at the bottom of the main window. To open this log in Notepad (and save a copy of its contents to disk), click the "Open in Notepad" button:

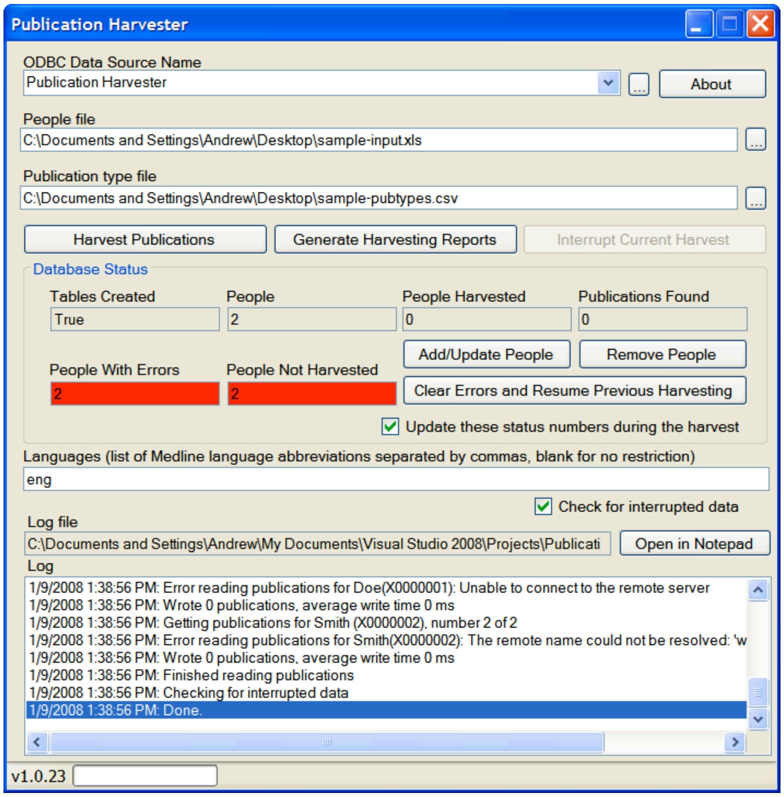


```
10/27/2006 12:29:19 PM: Checking for interrupted data
10/27/2006 12:29:19 PM: Beginning harvesting
10/27/2006 12:29:19 PM: Adding language restriction: eng
10/27/2006 12:29:19 PM: Initializing the database
10/27/2006 12:29:20 PM: Writing publication types to database
10/27/2006 12:29:22 PM: Reading people from medium input file.xls and writing them to the database
10/27/2006 12:29:24 PM: Getting publications for Klein (A6010513), number 1 of 6
10/27/2006 12:29:40 PM: Wrote 108 publications, average write time 26 ms
10/27/2006 12:29:40 PM: Getting publications for Van Eys (A6009400), number 2 of 6
10/27/2006 12:29:48 PM: Wrote 173 publications, average write time 22.6 ms
10/27/2006 12:29:48 PM: Getting publications for Tobian (A5401532), number 3 of 6
10/27/2006 12:29:57 PM: Wrote 157 publications, average write time 21.7 ms
10/27/2006 12:29:57 PM: Getting publications for Reemtsma (A5501586), number 4 of 6
10/27/2006 12:30:12 PM: Wrote 311 publications, average write time 20.7 ms
10/27/2006 12:30:12 PM: Getting publications for Guillemin (A5702471), number 5 of 6
10/27/2006 12:30:26 PM: Publication harvesting was interrupted
10/27/2006 12:30:26 PM: Checking for interrupted data
10/27/2006 12:30:26 PM: Done.
```

If there are errors – for example, if the local network goes down, preventing the Publication Harvester from contacting the NCBI server to issue PubMed queries – then those errors will appear in the log. The people will be tagged in the database with errors so that you can retry those people later (using the "Clear Errors and Resume Harvesting" button – see below). Some Medline queries will not retrieve any publications; when this happens, a warning will appear in the log.

If there are many rows in the People file, the harvester could take a long time to retrieve and process all of the publications. To make it easier to deal with large data, it is possible to interrupt and continue the harvesting operation. To do this, click the "Interrupt Current Harvest" button. You may also click the "X" button in the upper right-hand corner of the window to close the window – if you do this while harvesting publications, the software will prompt to verify that you really want to interrupt the harvest.

If the harvest is interrupted, the software will indicate this by coloring the "People with Errors" and/or the "People Not Harvested" fields red and allowing you to press the "Clear Errors and Resume Harvesting" button:



ODBC Data Source Name: Publication Harvester

People file: C:\Documents and Settings\Andrew\Desktop\sample-input.xls

Publication type file: C:\Documents and Settings\Andrew\Desktop\sample-pubtypes.csv

Tables Created	People	People Harvested	Publications Found
True	2	0	0

People With Errors	People Not Harvested
2	2

☑ Update these status numbers during the harvest

Languages (list of Medline language abbreviations separated by commas, blank for no restriction): eng

☑ Check for interrupted data

Log file: C:\Documents and Settings\Andrew\My Documents\Visual Studio 2008\Projects\Publicati

Log

```
1/9/2008 1:38:56 PM: Error reading publications for Doe(X0000001): Unable to connect to the remote server
1/9/2008 1:38:56 PM: Wrote 0 publications, average write time 0 ms
1/9/2008 1:38:56 PM: Getting publications for Smith (X0000002), number 2 of 2
1/9/2008 1:38:56 PM: Error reading publications for Smith(X0000002): The remote name could not be resolved: 'w
1/9/2008 1:38:56 PM: Wrote 0 publications, average write time 0 ms
1/9/2008 1:38:56 PM: Finished reading publications
1/9/2008 1:38:56 PM: Checking for interrupted data
1/9/2008 1:38:56 PM: Done.
```

v1.0.23

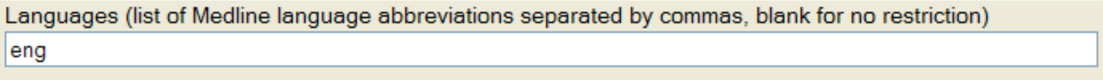
If the harvest is interrupted due to a system crash or shutdown, the software will treat it as if the harvest were interrupted properly. In this case, again push the "Clear Errors and Resume Previous Harvesting" step.

When the Publication Harvester loads a database, it immediately checks for interrupted data. If any interrupted data is found, then it displays those errors (turning the labels red to alert you that there's interrupted data or errors), and enables the "Clear Errors and Resume Previous Harvesting" button. This check can take a long time – up to a few minutes on very large databases. You can turn off this check by unchecking the "Check for interrupted data" checkbox.

Warning: Unchecking the "Check for interrupted data" checkbox can lead to unstable databases. Only use it if you are absolutely sure that the Publication Harvester software completed its last run on the database. This checkbox is very useful for quickly opening up a database so that you can generate reports, but it should not be used until all of the people in the database are completely harvested.

3.3 Retrieve Publications in Other Languages

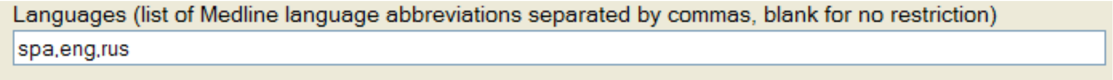
By default, the Publication Harvester will only retrieve publications that are in English, using the Languages field in the program:



Languages (list of Medline language abbreviations separated by commas, blank for no restriction)

eng

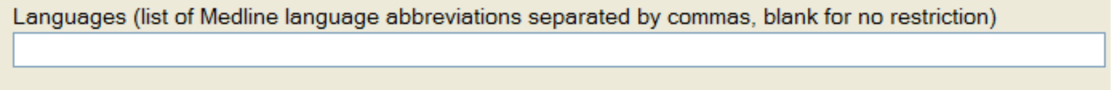
You can instruct the software to retrieve publications in different languages by changing the value in this box. To make the Publication Harvester retrieve publications in Spanish, English and Russian, set the box to contain the following value:



Languages (list of Medline language abbreviations separated by commas, blank for no restriction)

spa,eng,rus

The list of acceptable language abbreviations can be found at the MEDLINE/PubMed Language Table web page: http://www.nlm.nih.gov/bsd/language_table.html. The Languages box should contain a list of abbreviations separated by commas (with no spaces). To remove all language restrictions and harvest all publications regardless of language, leave the box blank:


Languages (list of Medline language abbreviations separated by commas, blank for no restriction)

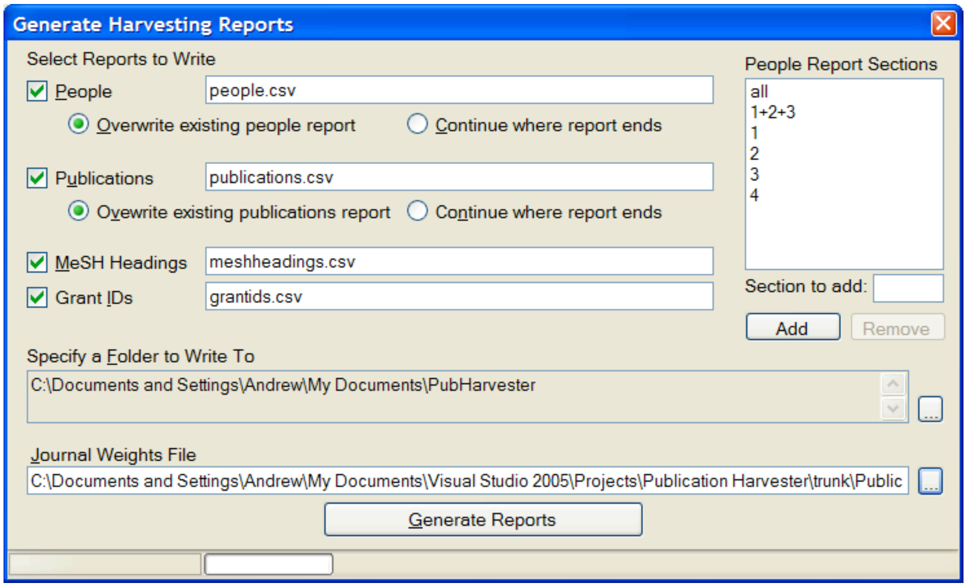
3.4 Maintain the List of People

You may change the contents of the database after it is generated by adding, updating or removing people. To add or update people, click the "Add/Update People" button. The software will prompt you for an Excel file in the same format as the People file (see above, section 3.1.1) containing the people to add or update. Any person in this file who is not in the database will be added; any person who is already in the database will be updated with information in the specified file. Any publications with any of these people will be disassociated with them. (They will not be removed from the database, however, since they may also be associated with other people.)

To remove people, click the "Remove People" button. The software will prompt you for an Excel file containing the people to remove. Any of these people who are in the database will be removed. (Any people who are in the file but not the database will be ignored.)

3.5 Generate Reports

To generate the reports based on the harvested publications, click the "Generate Harvesting Reports" button. The following window will pop up:


Generate Harvesting Reports

Select Reports to Write

☑ People: people.csv

● Overwrite existing people report ○ Continue where report ends

☑ Publications: publications.csv

● Oyewrite existing publications report ○ Continue where report ends

☑ MeSH Headings: meshheadings.csv

☑ Grant IDs: grantids.csv

People Report Sections: all, 1+2+3, 1, 2, 3, 4

Section to add:

Specify a Folder to Write To: C:\Documents and Settings\Andrew\My Documents\PubHarvester

Journal Weights File: C:\Documents and Settings\Andrew\My Documents\Visual Studio 2005\Projects\Publication Harvester\trunk\Public

The system will generate up to four different reports: the people report, the publications report, the MeSH headings report and the Grant IDs report. By default, all four will be generated; indicate that certain reports should not be generated by unchecking them in the "Generate Harvesting Reports" window.

The People and Publications reports have fault-tolerance built in. The report generation process is very time-consuming, and if it is accidentally interrupted the software will give the user the option to either overwrite the reports or continue where the previous reporting ended. The reports will be written to the filenames specified in the four textboxes on the form. Those files will be written to the folder specified in the "Specify a Folder to Write To" box.

A Journal Weights file must be supplied to calculate some of the reports. This is a Microsoft Excel 8.0 file with two columns:

- JOURNAL_TITLE: Name of journal
- JIF: Average Journal Impact Factor

A sample JIF file can be downloaded from the following URL: http://www.stellman-greene.com/PublicationHarvester/sample-JIFs.xls

Once all of these things are specified, click "Generate Reports" to generate the reports. As the reports are generated, status will be added to the log on the main form.

3.5.1 People Report

The **People report** contains one row per person per year. A journal weights file (see section 3.5) must be provided in order to calculate the weighted publication counts – the software must prompt the user for the location of this file before the reports are run. By default, the report contains the following columns:

Field	**Type**	**Description**
setnb **(key)**	Text	Person's identifier
year **(key)**	Number	Year of transition
	All: sum of all publication types	
pubcount	Number	Total nb. of pubs in year, as queried
wghtd_pubcount	Number	Weighted total nb. of pubs in year, as queried
pubcount_pos1	Number	Total nb. of pubs in year, as queried, 1st author
wghtd_pubcount_pos1	Number	Weighted total nb. of pubs in year, as queried, 1st author
pubcount_posN	Number	Total nb. of pubs in year, as queried, last author
wghtd_pubcount_posN	Number	Weighted total nb. of pubs in year, as queried, last author
pubcount_posM	Number	Total nb. of pubs in year, as queried, middle author
wghtd_pubcount_posM	Number	Weighted total nb. of pubs in year, as queried, middle author
pubcount_posNTL	Number	Total nb. of pubs in year, as queried, next-to-last author
wghtd_pubcount_posNTL	Number	Weighted total nb. of pubs in year, as queried, next-to-last author
pubcount_pos2	Number	Total nb. of pubs in year, as queried, 2nd author
wghtd_pubcount_pos2	Number	Weighted total nb. of pubs in year, as queried, 2nd author
	1+2+3: sum of all publication type categories 1, 2 and 3	
123pubcount	Number	Total nb. of pubs in year, bins I,II & III
wghtd_123pubcount	Number	Weighted total nb. of pubs in year, bins I,II & III
123pubcount_pos1	Number	Total nb. of pubs in year, bins I,II & III, 1st author
wghtd_123pubcount_pos1	Number	Weighted total nb. of pubs in year, bins I,II & III, 1st author
123pubcount_posN	Number	Total nb. of pubs in year, bins I,II & III, last author
wghtd_123pubcount_posN	Number	Weighted total nb. of pubs in year, bins I,II & III, last author
123pubcount_posM	Number	Total nb. of pubs in year, bins I,II & III, middle author
wghtd_123pubcount_posM	Number	Weighted total nb. of pubs in year, bins I,II & III, middle author
123pubcount_posNTL	Number	Total nb. of pubs in year, bins I,II & III, next-to-last author
wghtd_123pubcount_posNTL	Number	Weighted total nb. of pubs in year, bins I,II & III, next-to-last author
123pubcount_pos2	Number	Total nb. of pubs in year, bins I,II & III, 2nd author
wghtd_123pubcount_pos2	Number	Weighted total nb. of pubs in year, bins I,II & III, 2nd author
	Publication type category 1	
1pubcount	Number	Total nb. of pubs in year, bin I
wghtd_1pubcount	Number	Weighted total nb. of pubs in year, bin I
1pubcount_pos1	Number	Total nb. of pubs in year, bin I, 1st author
wghtd_1pubcount_pos1	Number	Weighted total nb. of pubs in year, bin I, 1st author

1pubcount_posN	Number	Total nb. of pubs in year, bin I, last author
wghtd_1pubcount_posN	Number	Weighted total nb. of pubs in year, bin I, last author
1pubcount_posM	Number	Total nb. of pubs in year, bin I, middle author
wghtd_1pubcount_posM	Number	Weighted total nb. of pubs in year, bin I, middle author
1pubcount_posNTL	Number	Total nb. of pubs in year, bin I, next-to-last author
wghtd_1pubcount_posNTL	Number	Weighted total nb. of pubs in year, bin I, next-to-last author
1pubcount_pos2	Number	Total nb. of pubs in year, bin I, 2nd author
wghtd_1pubcount_pos2	Number	Weighted total nb. of pubs in year, bin I, 2nd author
Publication type category 2		
2pubcount	Number	Total nb. of pubs in year, bin II
wghtd_2pubcount	Number	Weighted total nb. of pubs in year, bin II
2pubcount_pos1	Number	Total nb. of pubs in year, bin II, 1st author
wghtd_2pubcount_pos1	Number	Weighted total nb. of pubs in year, bin II, 1st author
2pubcount_posN	Number	Total nb. of pubs in year, bin II, last author
wghtd_2pubcount_posN	Number	Weighted total nb. of pubs in year, bin II, last author
2pubcount_posM	Number	Total nb. of pubs in year, bin II, middle author
wghtd_2pubcount_posM	Number	Weighted total nb. of pubs in year, bin II, middle author
2pubcount_posNTL	Number	Total nb. of pubs in year, bin II, next-to-last author
wghtd_2pubcount_posNTL	Number	Weighted total nb. of pubs in year, bin II, next-to-last author
2pubcount_pos2	Number	Total nb. of pubs in year, bin II, 2nd author
wghtd_2pubcount_pos2	Number	Weighted total nb. of pubs in year, bin II, 2nd author
Publication type category 3		
3pubcount	Number	Total nb. of pubs in year, bin III
wghtd_3pubcount	Number	Weighted total nb. of pubs in year, bin III
3pubcount_pos1	Number	Total nb. of pubs in year, bin III, 1st author
wghtd_3pubcount_pos1	Number	Weighted total nb. of pubs in year, bin III, 1st author
3pubcount_posN	Number	Total nb. of pubs in year, bin III, last author
wghtd_3pubcount_posN	Number	Weighted total nb. of pubs in year, bin III, last author
3pubcount_posM	Number	Total nb. of pubs in year, bin III, middle author
wghtd_3pubcount_posM	Number	Weighted total nb. of pubs in year, bin III, middle author
3pubcount_posNTL	Number	Total nb. of pubs in year, bin III, next-to-last author
wghtd_3pubcount_posNTL	Number	Weighted total nb. of pubs in year, bin III, next-to-last author
3pubcount_pos2	Number	Total nb. of pubs in year, bin III, 2nd author
wghtd_3pubcount_pos2	Number	Weighted total nb. of pubs in year, bin III, 2nd author
Publication type category 4		
4pubcount	Number	Total nb. of pubs in year, bin IV
wghtd_4pubcount	Number	Weighted total nb. of pubs in year, bin IV
4pubcount_pos1	Number	Total nb. of pubs in year, bin IV, 1st author
wghtd_4pubcount_pos1	Number	Weighted total nb. of pubs in year, bin IV, 1st author
4pubcount_posN	Number	Total nb. of pubs in year, bin IV, last author
wghtd_4pubcount_posN	Number	Weighted total nb. of pubs in year, bin IV, last author
4pubcount_posM	Number	Total nb. of pubs in year, bin IV, middle author
wghtd_4pubcount_posM	Number	Weighted total nb. of pubs in year, bin IV, middle author
4pubcount_posNTL	Number	Total nb. of pubs in year, bin IV, next-to-last author
wghtd_4pubcount_posNTL	Number	Weighted total nb. of pubs in year, bin IV, next-to-last author
4pubcount_pos2	Number	Total nb. of pubs in year, bin IV, 2nd author
wghtd_4pubcount_pos2	Number	Weighted total nb. of pubs in year, bin IV, 2nd author

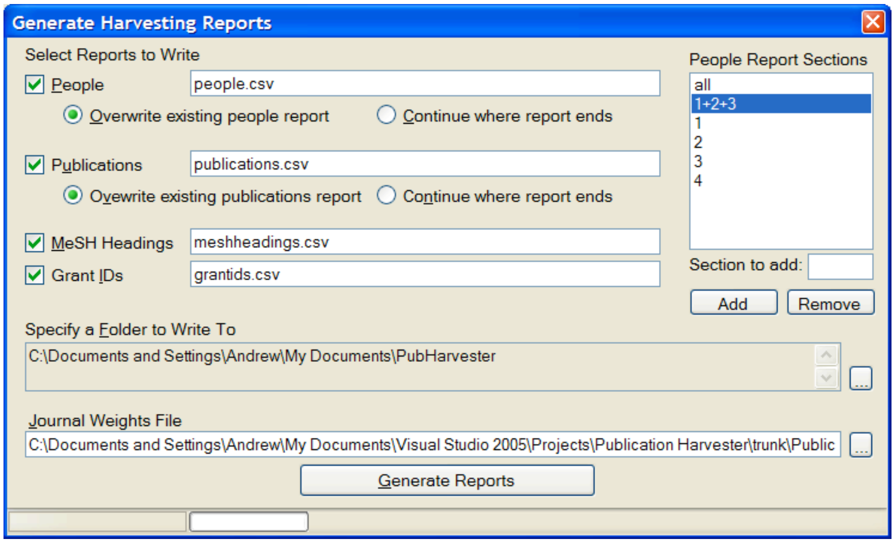
When a column is noted “as queried”, that refers to all bins (I, II, III, IV and publications without a bin).

In each of the Pubcount columns, the “bin” refers to the PublicationType value for the publication. For example, the 4pubcount_pos1 refers to any publication with a PubTypeCategoryID of 4, where the person is in the first author position (PersonPublications.PositionType = 1).

The Pubcount columns are mutually exclusive, so that within a “bin” the total pubcount is equal to the sum of the individual author pubcounts. For example, in any row, the value in 3pubcount must equal 3pubcount_pos1 + 3pubcount_posN + 3pubcount_posM + 3pubcount_pos2 + 3pubcount_posNTL.

The Wghtd_Pubcount* columns contain the count of publications for each journal for the year multiplied by the weight of the journal (from the Journal Weights input table). If a journal name cannot be matched in the Journal Weights file, the corresponding publication is ignored in the weighted pub count (i.e. it is assigned a weight of zero).

By default, the People report contains two key columns (setnb, year), and five sections based on the publication type category: sum of all categories ("all"), sum of categories 1 through 3 ("1+2+3"), category 1, category 2, category 3 and category 4. The report can be customized to add or remove sections. To remove a section from the report, click on it in the "People Report Sections" list and click "Remove":



Generate Harvesting Reports

Select Reports to Write

☑ People: people.csv

● Overwrite existing people report ○ Continue where report ends

☑ Publications: publications.csv

● Ovewrite existing publications report ○ Continue where report ends

☑ MeSH Headings: meshheadings.csv

☑ Grant IDs: grantids.csv

Specify a Folder to Write To: C:\Documents and Settings\Andrew\My Documents\PubHarvester

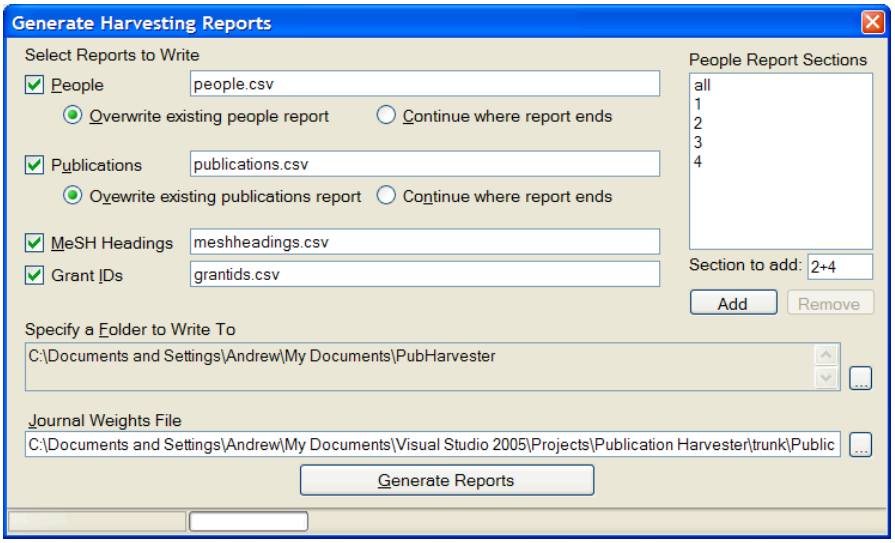
Journal Weights File: C:\Documents and Settings\Andrew\My Documents\Visual Studio 2005\Projects\Publication Harvester\trunk\Public

People Report Sections: all, 1+2+3 (selected), 1, 2, 3, 4

Section to add:

In the above example, the "1+2+3" section will be removed from the report.

To add a section to the report, enter it in the "Section to add" box and click "Add":



Generate Harvesting Reports

Select Reports to Write

☑ People: people.csv

● Overwrite existing people report ○ Continue where report ends

☑ Publications: publications.csv

● Ovewrite existing publications report ○ Continue where report ends

☑ MeSH Headings: meshheadings.csv

☑ Grant IDs: grantids.csv

Specify a Folder to Write To: C:\Documents and Settings\Andrew\My Documents\PubHarvester

Journal Weights File: C:\Documents and Settings\Andrew\My Documents\Visual Studio 2005\Projects\Publication Harvester\trunk\Public

People Report Sections: all, 1, 2, 3, 4

Section to add: 2+4

In the above example, the section "2+4" will be added to the report. This will add a section with columns named "24pubcount", "wghtd_24pubcount", etc. These columns will contain the values for all publications with categories 2 or 4.

3.5.2 Publications Report

The **Publications report** contains one row per publication, based on data in the PersonPublications table.

Field	Type	Description
setnb **(key)**	Text	Person unique identifier
pmid **(key)**	Number	Unique article identifier
journal_name	Text	Name of journal
year	Number	Year of publication
Month	Text	Month of publication
Day	Number	Day of publication
Title	Text	Article title
Volume	Text	volume number of the journal in which the article was published
Issue	Text	Issue in which the article was published
Position	Number	Position in authorship list
nbauthors	Number	Number of coauthors (including the person)
Bin	Number	From I to IV, based on the PubTypeCategoryID value
Pages	Text	Page numbers
grant_agency	Text	Agency who awarded the grant
Publication_type	Text	Publication Type from Medline

3.5.3 MeSH Headings Report

The **MeSH Headings report** contains one row per person per year per MeSH heading:

Field	Type	Description
Setnb **(key)**	Text	Person unique identifier
Year **(key)**	Number	Year of publication
Heading	Text	MeSH Heading
Count	Number	Number of MeSH headings for the person in the specified year

3.5.4 Grants Report

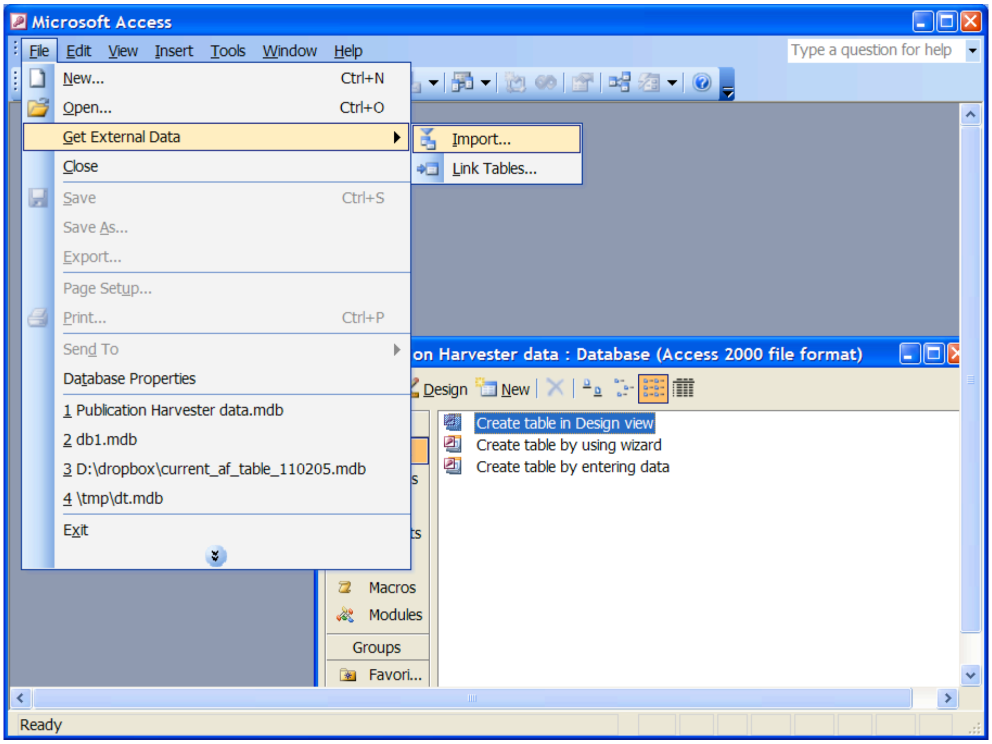
The **Grants report** contains one row per person per year per publication per grant ID, sorted by year:

Field	Type	Description
Year	Number	Year of publication
PMID	Number	Publication ID
GrantID	Text	GrantID for the publication (there may be several rows per publication, one per grant ID)

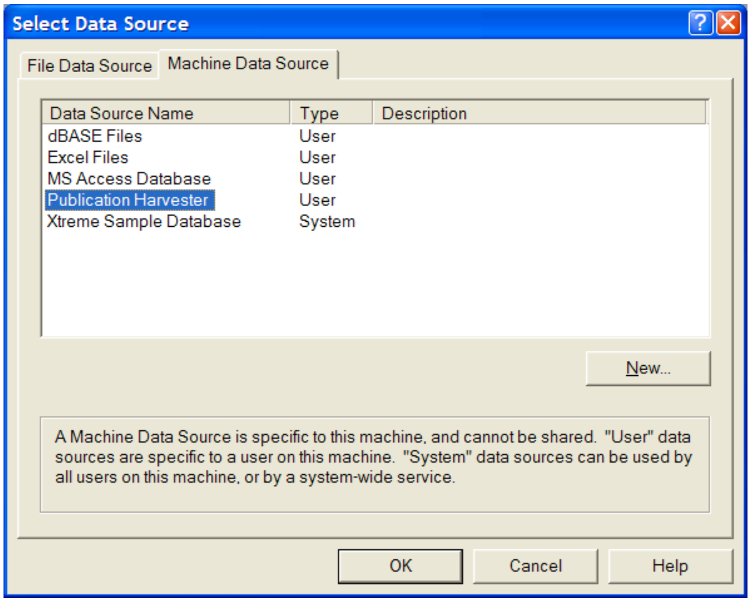
4 Importing Data into Microsoft Access 2003

Many researchers use Microsoft Access to manipulate data. The MySQL Connector/ODBC (which is required by the Publication Harvester – see section 2.3) allows Microsoft Access to easily import data from a Publication Harvester database.

To import Publication Harvester data into Microsoft Access 2003, first open the database that you would like to import data into. Then select the "File / Get External Data / Import" menu option:



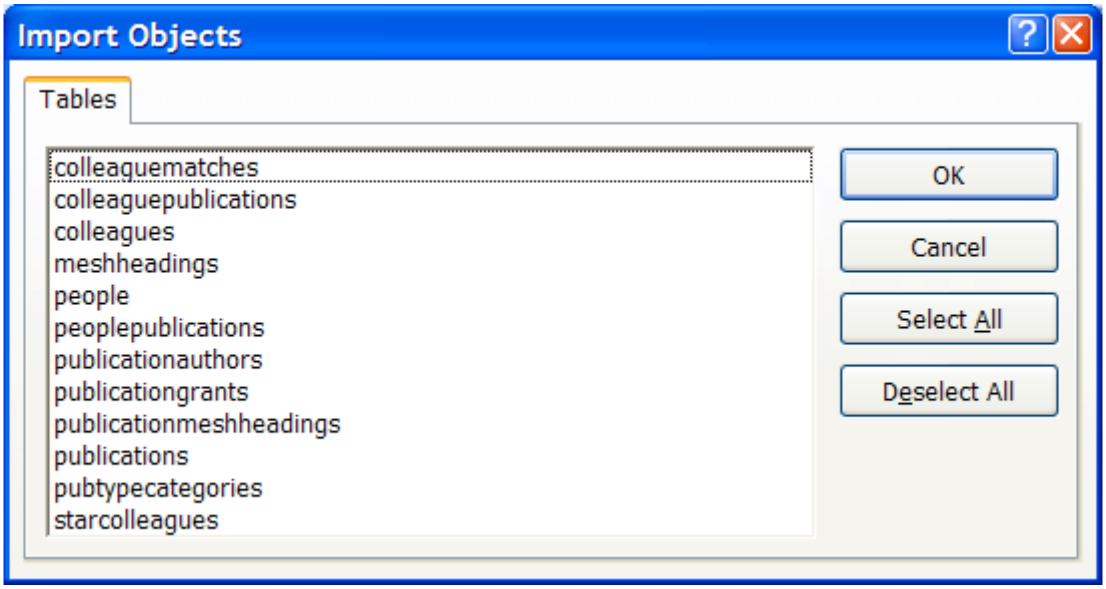
Access will pop up a "File Open…" dialog box. At the bottom of the dialog box is a dropdown list labeled "Files of type:". Select "ODBC Databases ()" from this list. This will cause the following window to pop up:



Data Source Name	Type	Description
dBASE Files	User	
Excel Files	User	
MS Access Database	User	
Publication Harvester	User	
Xtreme Sample Database	System	

A Machine Data Source is specific to this machine, and cannot be shared. "User" data sources are specific to a user on this machine. "System" data sources can be used by all users on this machine, or by a system-wide service.

Click on the “Machine Data Source” tab. The data source that you created earlier (see section 2.5) should appear in the list. (In this example, the data source is called “Publication Harvester”.) Select that data source and click “OK”. Access will then display the list of tables in the database:



colleaguematches
colleaguepublications
colleagues
meshheadings
people
peoplepublications
publicationauthors
publicationgrants
publicationmeshheadings
publications
pubtypecategories
starcolleagues

Select any tables that you would like to import into Access and click “OK”. Access will copy the data from MySQL into your Access database. (For more information on the structure and contents of the tables, see the Publication Harvester software requirements specification.)

Note that any other data management or analysis tool which uses ODBC data sources can import data in the same manner.

5 GNU Free Documentation License

Version 1.2, November 2002

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- **A.** Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- **B.** List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- **C.** State on the Title page the name of the publisher of the Modified Version, as the publisher.
- **D.** Preserve all the copyright notices of the Document.
- **E.** Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- **F.** Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- **G.** Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- **H.** Include an unaltered copy of this License.
- **I.** Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- **J.** Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- **K.** For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- **L.** Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- **M.** Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- **N.** Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- **O.** Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

6 Revision History

Date	Author	Description
28-Jan-2006	Andrew Stellman	Created initial version
30-Jan-2006	Andrew Stellman	Finished adding features, sent to Pierre for review
02-Apr-2006	Andrew Stellman	Added section on adding/updating and removing people, and updated screenshots to reflect the latest version of the software.
26-Apr-2006	Andrew Stellman	Added a section on importing data into Microsoft Access
27-Oct-2006	Andrew Stellman	Updated to version 1.0.21, added multiple language support
09-Jan-2008	Andrew Stellman	Updated to version 1.0.23, added Check for Interrupted Data checkbox and improved error handling and logging