Microsoft® Office Excel® 2007 Inside Out

*Mark Dodge and
Craig Stinson*

To learn more about this book, visit Microsoft Learning at
http://www.microsoft.com/MSPress/books/9702.aspx

9780735623217
Publication Date: January 2007

Microsoft® Press

Working with External Data

Microsoft Office Excel 2007 is a superb tool for analyzing data, but before you can do any analysis, you have to get the data into Excel. In many cases, the information you need resides somewhere "outside"—on a server, on a Web site, in an XML file, or perhaps in a database program such as Oracle or Microsoft Office Access. Office Excel 2007 supports a wide variety of data formats, including SQL Server (and SQL Server Analysis Services), Access, dBase, FoxPro, Oracle, Paradox, and various kinds of text files. We'll look at some techniques for retrieving external data in this chapter.

Using and Reusing Data Connections

An Office Data Connection (.odc) file is a small XML file that records information about how a workbook connects to an external data source. Such information can include the location and type of the external data, a query specification (if the connection is designed to retrieve a subset of the external source), and details about how to log on to the external server. ODC files are designed to facilitate the reuse of external connections.

Often the simplest way to import data from an external source is to execute an ODC file—a connection that either you or someone else has already established. To see what connections are available, click the Data tab on the Ribbon, and then click Existing Connections in the Get External Data group. A dialog box comparable to the one shown in Figure 23-1 appears.

In Figure 23-1, the Show list, at the top of the dialog box, is set to display all available connection files. You can use this list to restrict the dialog box to connections that are already open on your computer, connections that are already in use in the current workbook, or connections that are available on your network. If a connection file that you're looking for doesn't appear in the Existing Connections dialog box, click Browse For More. This will invoke the Windows Vista search facility, which will gather connection files from various locations on your computer.
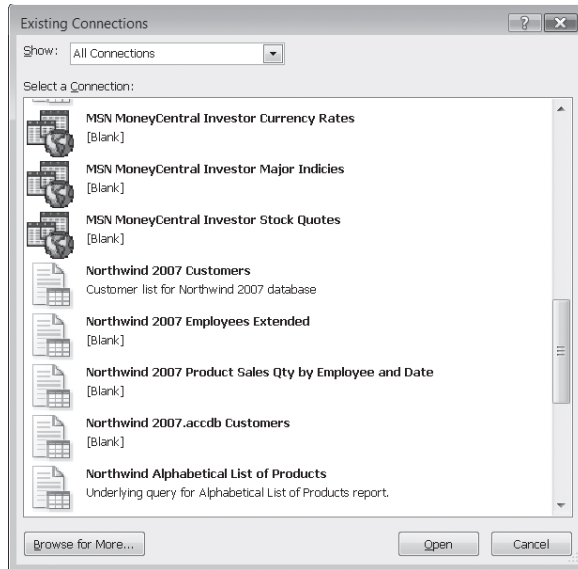
Figure 23-1 The Existing Connections dialog box lists connection files that are already established for you.

You can distinguish the various types of connection files that appear in the Existing Connections dialog box by their types of icons. In the examples that follow, we'll assume that you're opening one that looks like Northwind 2007 Customers in Figure 23-1. This connection file enables Excel to import data from an Access database. If you open one of the Web query connections (the three included in Excel 2007 begin with "MSN MoneyCentral Investor"), the dialog boxes you see will be somewhat different from the ones described here. (For more about Web queries, see "Using a Web Query to Return Internet Data" on page 779.)

To open a connection file, double-click it in the Existing Connections dialog box. The Import Data dialog box, shown in Figure 23-2, appears. In this dialog box, you indicate where you want the data to go and whether you want an ordinary table or a PivotTable.

If you accept the default settings, Excel creates a table at the current cell location. (For information about creating a PivotTable, see Chapter 22, "Analyzing Data with Pivot-Table Reports.") The resulting table behaves like any other Excel table (see Chapter 21, "Managing Information in Tables"), except for a crucial difference: The table remains linked to its external source, letting you refresh the data (update it with any changes that have occurred in the external source) on demand or at regular time intervals.
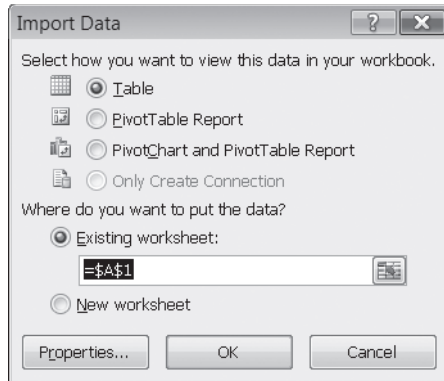
Figure 23-2 By default, Excel renders imported data as a table. Using the Import Data dialog box, you can create a PivotTable (or PivotTable and PivotChart) instead.

Setting Refresh Options

To specify how you want the data refreshed, you can click Properties in the Import Data dialog box (see Figure 23-2). Alternatively, after the table (or PivotTable) has been created, select a cell within it, click the Data tab, and then click Properties. In the External Data Properties dialog box that appears, click the Connection Properties button (to the right of the Name box):
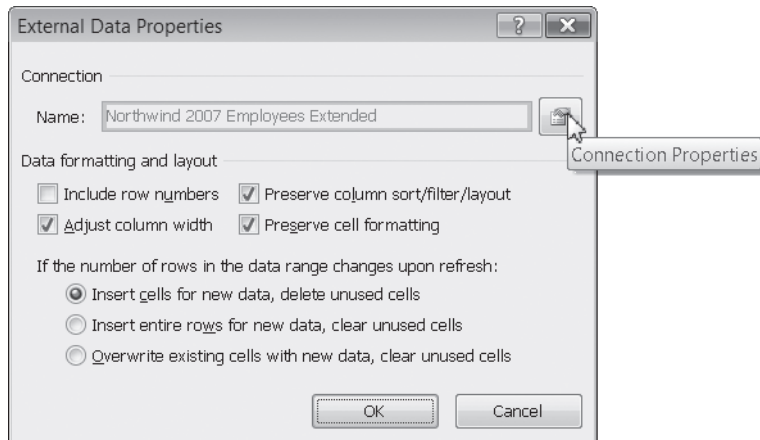
These steps bring you to the Connection Properties dialog box, shown in Figure 23-3. Your refresh options appear on the Usage tab in this dialog box.
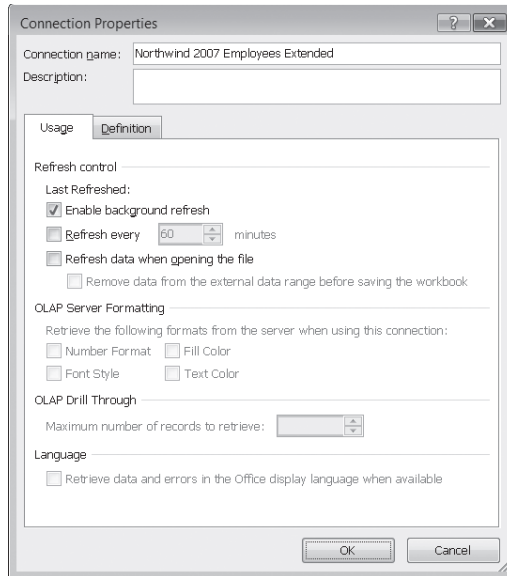
Figure 23-3 You can set your connection to refresh the imported data at regular time intervals.

The check boxes in the Refresh Control area in this dialog box are not mutually exclusive. You can have Excel refresh your data whenever you open the file as well as at regular time intervals. The Enable Background Refresh check box, selected by default, means you can do other work in Excel while the refresh is in progress. Note that this option is not available with online analytical processing (OLAP) queries.

If you select the Refresh Data When Opening The File check box, an additional option to remove the data from your worksheet when you close the file becomes available. You might as well select this check box as well, because Excel is going to refresh the data when you reopen the file anyway.

## Requiring or Not Requiring a Password to Refresh

If connecting to your external data requires a password, Excel, by default, will require that you supply the password again whenever you refresh. If that's a burdensome obligation, click the Definition tab in the Connection Properties dialog box. Then clear the Save Password check box.

Refreshing on Demand

In addition to requesting a refresh at regular time intervals, you can refresh the data whenever the need arises. Right-click a cell within the table, and then click Refresh. Alternatively, click the Data tab, click the arrow next to Refresh All, and then click Refresh. (Or simply click Refresh All; this refreshes all connections open in the current workbook.)

Opening an Entire Access Table in Excel

To import an entire table created in Access (as opposed to a specific set of records from that table), click the Data tab, and then click From Access in the Get External Data group. Windows Vista will launch a search for files with the extensions .mdb, .mde, .accdb, and .accde. When you find the Access file you're looking for, select it, and then click Open. You'll be presented with the Select Table dialog box, which is shown in Figure 23-4. (Drag the lower-right corner of the Select Table dialog box if you need to see more of the Description column.)
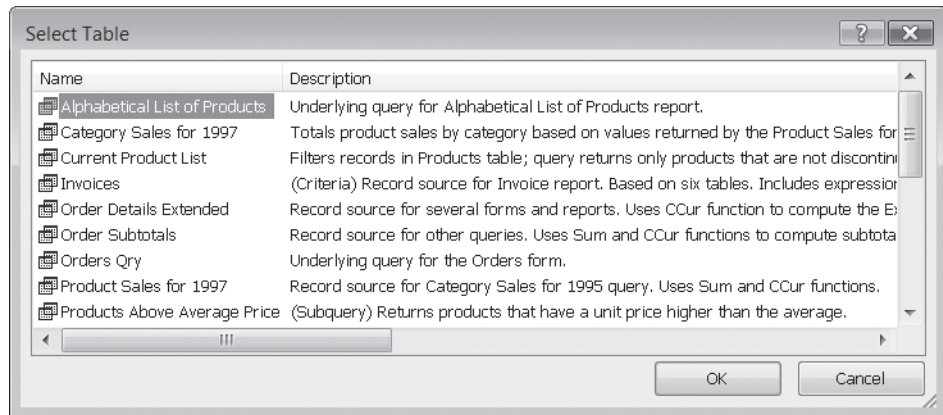
Figure 23-4 When you click the From Access command to open an Access file, the Select Table dialog box asks you to choose which table you want to import.

For information about importing selected records from an Access (or other) database using Microsoft Query, see "Using Microsoft Query to Import Data" on page 759.

The Select Table dialog box actually lists not only tables from your Access file but certain views as well. (The items shown in Figure 23-4, for example, are all views from the Northwind database that Microsoft supplies as a sample file with Access.) If you open any Access view in Excel, you get all the records currently displayed by Access in that view. If you open a table, you import all the records in that table. In either case, the data you import becomes either a table or a PivotTable in Excel, depending on how you complete the Import Data dialog box (see Figure 23-2).

You can set refresh parameters for your imported Access data the same way as for any other data connection. For details, see "Setting Refresh Options" on page 747.

Chapter 23

> **Note**
>
> You can also import an Access table by clicking the Microsoft Office Button, clicking Open, and then selecting Access Databases from the list beside the File Name box. (The resulting table or PivotTable behaves as if you had used the From Access command on the Ribbon.) You cannot save an Excel range as an Access table, however.

Working with Data in Text Files

Excel can read data in fixed-width as well as delimited text files. (A *delimited* file is one that uses some particular character or combination of characters to mark the boundaries between fields. A *fixed-width* file is one that uses space characters—as many as necessary—to achieve field alignment.) You can either open a text file (by clicking the Microsoft Office Button and clicking Open) or import it (by clicking the Data tab and then clicking From Text in the Get External Data group). If you want to be able to maintain a refreshable link to the source file, you need to do the latter.

When you open a comma-separated-values (.csv) file, Excel parses the data immediately into columns. If you open or import any other kind of text file, Excel presents the Text Import Wizard, described next.

> **Note**
>
> When you ask Excel to open or import a text file, the program looks for files with the extensions .prn, .txt, and .csv. If you want a file with a different extension, select All Files in the list to the right of the File Name box. Excel determines a file's type by its content, so it doesn't matter what the extension is.

Using the Text Import Wizard

With the Text Import Wizard, you can show Excel how to parse your text file. You get to tell the program what character or character combination (if any) is used to delimit columns, what kind of data appears in each column, and what character set or language was used to create the original file. You can also use the Text Import Wizard to exclude one or more rows at the top of your file—an option that's particularly useful if your file begins with some kind of noncolumnar descriptive information.

The first page of the wizard, shown in Figure 23-5, presents a preview of the data that Excel is about to import. It also indicates the best estimation of whether your file is delimited or fixed-width. You'll find that the wizard is usually correct with this first

guess—but if it's mistaken, you can set it straight. (If you're not sure, just go to the second page. When you get there, you'll know whether the program was wrong, and you can return to the first page to fix the problem.)
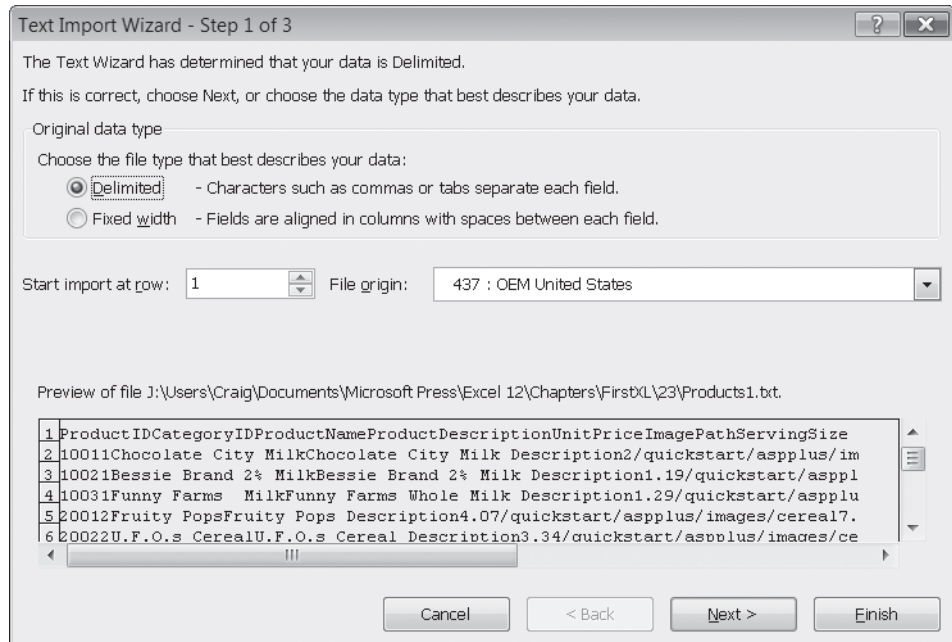

Figure 23-5 You can use the Text Import Wizard to tell Excel how to parse your text file.

While you're still on the first page of the wizard, use the Start Import At Row text box to eliminate any header rows that you can live without in Excel. Header rows make it hard for Excel to parse your file correctly, so you can help the program (and yourself) by lopping them off here. Click Next.

The second page of the Text Import Wizard looks something like either Figure 23-6 or Figure 23-7, depending on whether your file is delimited or fixed-width. In both cases, the vertical lines in the Data Preview section show how Excel proposes to split your file into columns. The Data Preview section regrettably shows a paltry 5 rows at a time and 65 characters per row. You cannot make it show more, but you can look at other parts of the file by using the scroll bars.

If your file is delimited, the second page of the wizard indicates what character Excel regards as the delimiter. In Figure 23-6, for example, the program has correctly divined that the file in question is tab-delimited. Most of the time, Excel gets this right. If it does not in your case, you can select a different check box and see the effect immediately in the Data Preview section. You can also select more than one check box to indicate that your file is delimited by multiple characters. If you select two or more check boxes, Excel breaks to a new column whenever it sees any of your choices.
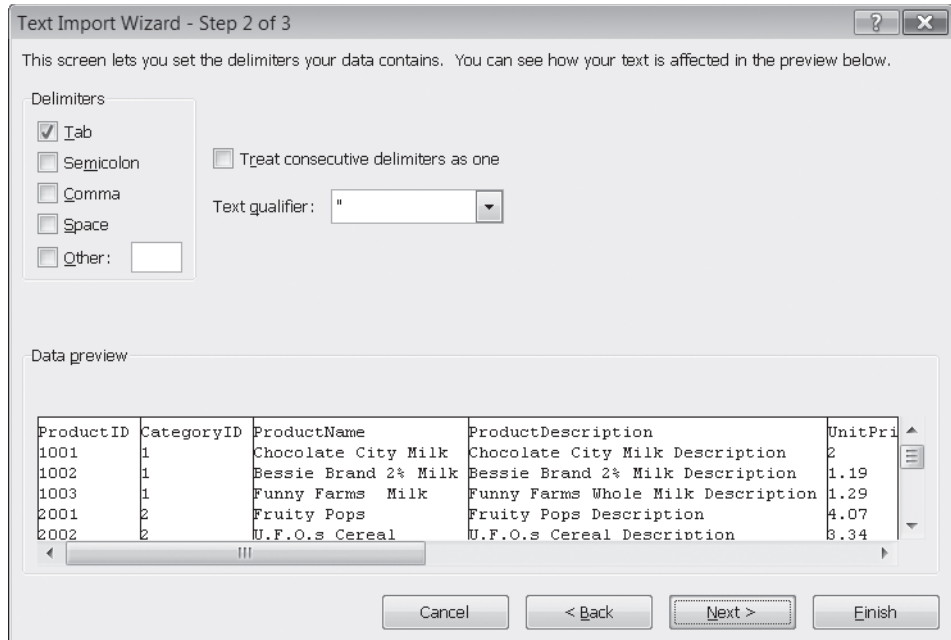
Figure 23-6 If your file is delimited, the second page of the wizard indicates what character Excel has recognized as the delimiter, and the Data Preview section shows how Excel will parse your file.

A separate check box lets you stipulate that Excel should regard consecutive delimiting characters as a single delimiter. You'll find that this option sometimes saves the day with tab-delimited files. The original creator of the file might occasionally have used two or more tabs to skip to the next column when the current column's contents were short. That strategy could disrupt your alignment in Excel unless you tell the program to treat consecutive delimiters as a single delimiter.

Excel is much more likely to introduce errors when trying to parse a fixed-width file. In the file shown in Figure 23-7, for example, the program initially fails to recognize that the first column break should occur between ID and Category to create a Product ID column and a Category ID column. It also erroneously plants a column break in the middle of the Category ID heading. Fortunately, it's easy to fix parsing problems of this sort. (However, if the file is long and the mistakes are many, catching them all is likely to become a trial; if you have a choice between opening a fixed-width file and an equivalent delimited file, by all means go with the latter.)
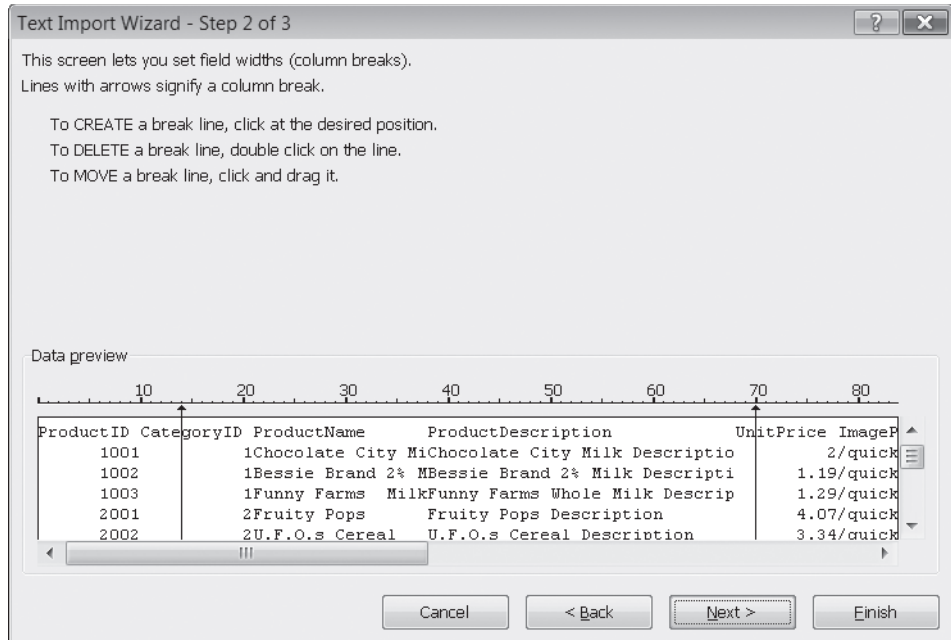
Figure 23-7 If your file is fixed-width, be sure to look at the Data Preview box ; you can fix any mistakes by manipulating the vertical bars.

To fix parsing errors, drag vertical lines to the left or right to reposition the column breaks. To create a column break where one doesn't yet exist, click once at the appropriate place. To remove a column break that shouldn't be there at all, double-click it. When you have finished, click Next.

The third page of the wizard, shown in Figure 23-8, lets you specify the data type of each column. Your choices are limited to General (which treats text as text, numbers as numbers, and dates in recognizable formats as dates), Text (which treats everything as text), Date, and Do Not Import Column (Skip). Excel initially assigns the General description to all columns, and you'll probably want to override that presumption in some cases. For example, if your file happens to have a text field that begins with a hyphen, Excel will regard the hyphen as a minus sign and attempt to turn your text into a formula. You can avoid errors by indicating that the field is Text.
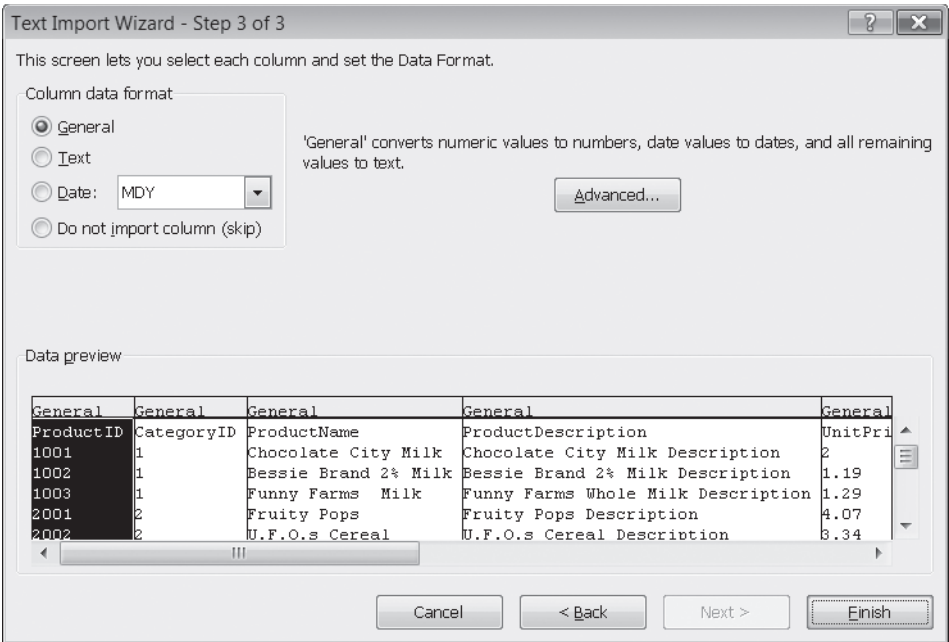
Figure 23-8 You can use the third page of the wizard to control the data type of each column.

Opening dBase Files

Excel no longer gives you the option of saving files in any of the dBase formats, but you can still open such files. To open a dBase file, click the Microsoft Office Button, click Open, and then select dBase Files in the list to the right of the File Name box:
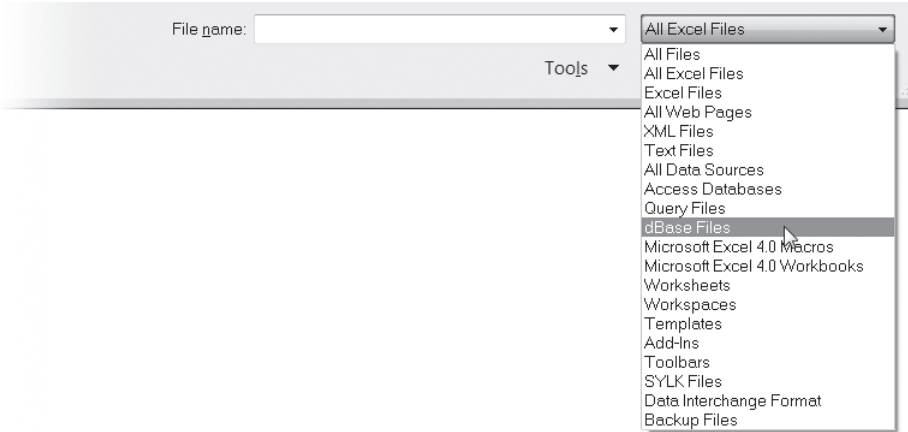
An imported dBase file becomes an unrefreshable range in Excel.

The third page also includes an Advanced button. By clicking it, you can change the way the wizard handles commas and periods in numeric data. By default, Excel uses the settings specified in the Regional And Language Options item of Control Panel. If your text file was created under other assumptions, you'll need to make some adjustments in the Advanced Text Import Settings dialog box. Click OK, click Finish, and then click OK to import the text file.

Parsing Clipboard Text

Occasionally, when working with text data, you might find long text strings that you need to break into separate columns. This can happen, for example, if you paste text into Excel from the Clipboard. To parse such data, select it, click the Data tab, and then click Text To Columns in the Data Tools group.

Working with XML Files

Excel can open, import, and export XML data in any structure. To open a list that has been saved in XML, click the Microsoft Office Button, and click Open—just as you would to open an ordinary Excel workbook. With the Open dialog box set to display all Excel files, your XML files will be included. But if you're having trouble finding the file you want (because of all the other Excel files in the same folder), select XML Files in the list beside the File Name box.

When you open your file, Excel presents the Open XML dialog box, shown in Figure 23-9.
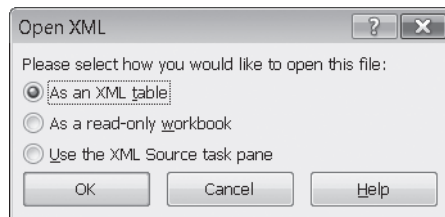
Figure 23-9 When you open an XML data file, Excel presents this dialog box. Choose the first option to open all elements of the XML structure or the third option to work only with particular elements.

As Figure 23-10 shows, the result of opening an XML file using the As An XML Table option is a table that presents each element of the source file, in order, as a table column. All records in the source file are included in the resulting list.
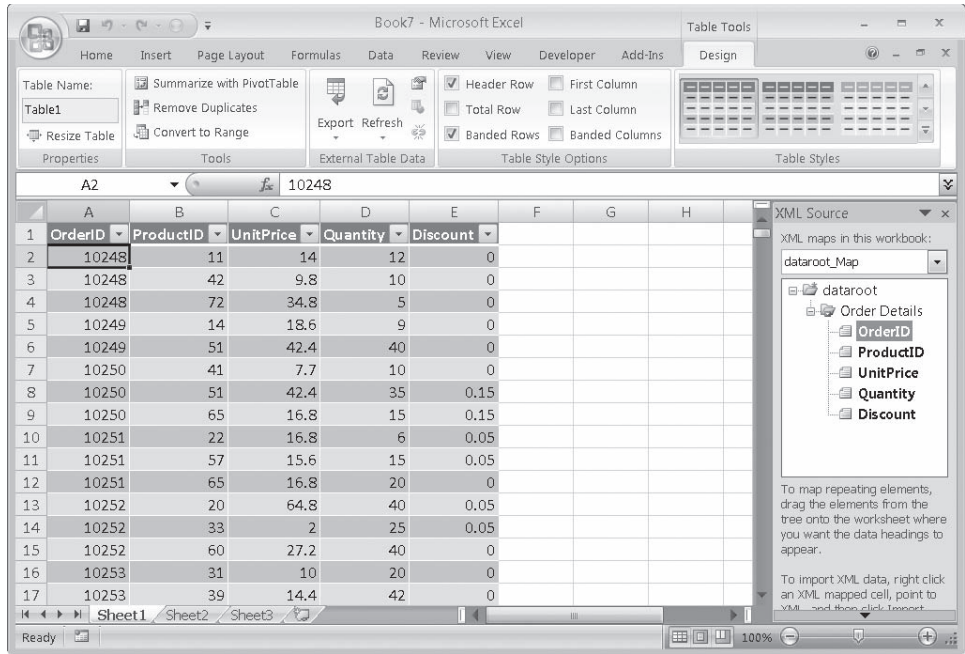
Figure 23-10 Each element of the opened XML file maps to a column in the resulting XML table.

TROUBLESHOOTING

Excel reports a problem with the specified XML or schema source.

When you open an XML file, Excel looks for an associated schema file, which defines the structure of the XML data. If it doesn't find one, or if it finds errors in the associated schema file, you will see an error message. If you click OK, Excel will infer the structure from the data it sees. In many cases, particularly with files that are not particularly complex, this works out fine. You can forget about the error message after you have clicked OK. If Excel can't infer the structure of your file, you will need to fix the schema (or provide one).

After Excel opens the XML file, it can display an XML Source task pane, which shows how the elements of the source file map to columns in the table. If the task pane isn't visible and you want to see it, right-click a cell in the table, click XML, and then click XML Source.

Opening an XML file by the method just described (clicking the Microsoft Office Button and clicking Open) creates a new workbook. If you want to create an XML table on an existing worksheet, click the Data tab, click From Other Sources (in the Get External

Data group), and click From XML Data Import. After you have selected your file in the ensuing dialog box, Excel will ask you where to put the incoming data.

You can refresh an XML table the same way you would an imported text or Access table. For details, see "Setting Refresh Options" on page 747.

Creating an Ad Hoc Mapping of XML Elements to Table Columns

Opening an XML file using the As An XML Table option (see Figure 23-9) might be fine for a relatively simple XML structure. But if your structure is not simple, it's likely you'll be interested in only certain portions of the XML data. In such cases, it's usually more effective to open the file using the third option, Use The XML Source Task Pane. When you do this, Excel presents the XML structure in the XML Source task pane, without creating a table—as shown in Figure 23-11.
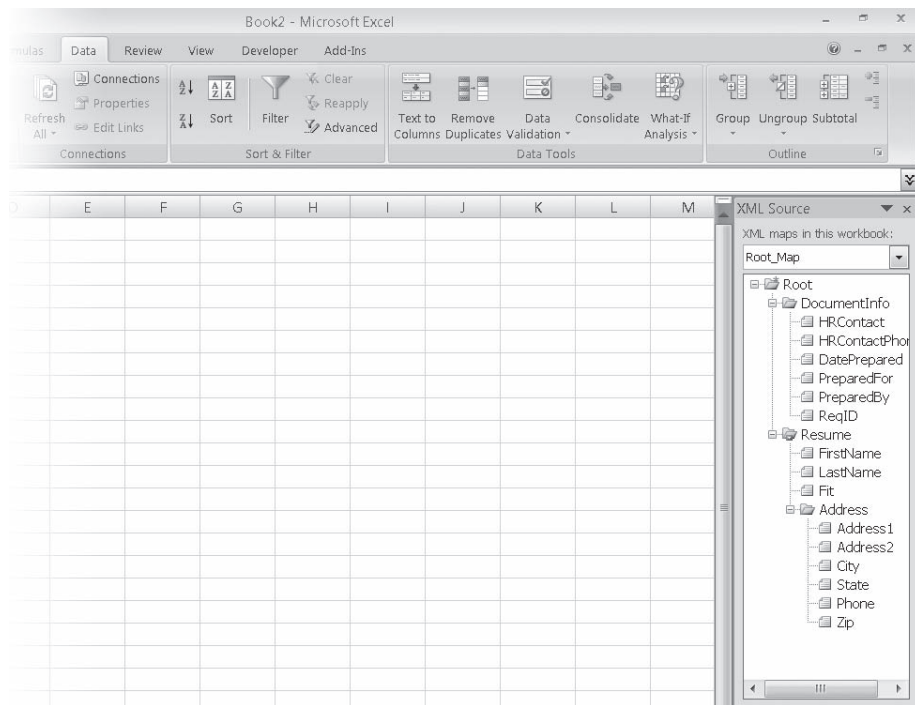
Figure 23-11 When you open an XML file using the XML Structure task pane to map elements to table columns, no data appears until you drag XML elements from the task pane to the worksheet and then refresh.

In the example shown in Figure 23-11, the data file consists of ratings and contact data for a set of job applicants, along with contact information about the human-resources person who conducted each interview. If you're reviewing this data, you might be interested in only the HRContact field from the DocumentInfo element, the LastName

and FirstName fields from the Resume element, and perhaps some additional fields pertaining to individual applications. To create a table on your worksheet that displays only the fields you care about, you can Ctrl+click the headings of interest in the XML Structure task pane and then drag the selected set onto the worksheet. (Excel calls this process of associating XML elements with table headings *mapping*.) The result might look like Figure 23-12.
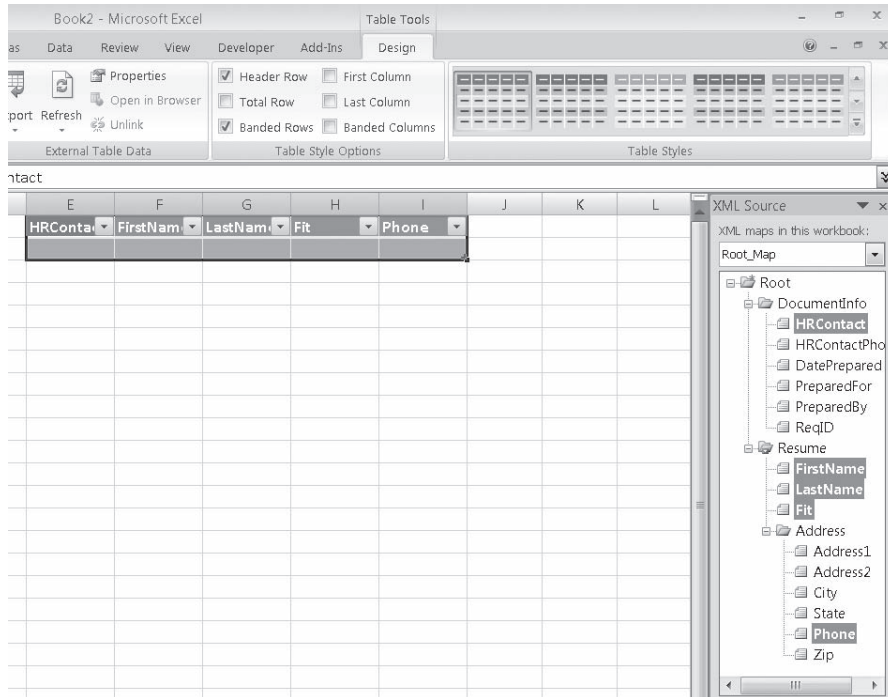
Figure 23-12 You can drag headings from the XML Structure task pane to the worksheet to create a table showing only particular columns.

To populate the table after you have it structured, click a cell in the table header row, click the Data tab (or Design tab), and then click Refresh All. (If you have other tables open and you want to refresh only this one, click the small arrow beneath Refresh All, and then click Refresh.)

Importing XML Data Using an Existing XML Structure

Populating the table by clicking Refresh All, as just described, brings in data from the file whose structure you've imported into the XML Structure task pane. As an alternative, you can right-click a cell in the table header, click XML, and then click Import. You'll then be prompted for the name of an XML data file.

The Import command lets you bring in records from any file whose structure is reflected in the XML Source task pane. Importing is particularly useful when you have a number of identically structured XML files. For example, if each member of your

human-resources staff created a separate file of interviewee data with each file built on the same XML schema, you could examine each one in turn with the help of the Import command.

> **Note**
>
> If you perform successive imports of two or more identically structured files, each import replaces the previous one. If instead you want to import several files at once, use the Import procedure as described, and then Ctrl+click each file you want to import.

Using Microsoft Query to Import Data

If you don't have a connection already set up for the data you need, you can create one with the help of Microsoft Query, a versatile querying tool included with the 2007 Microsoft Office system. Query generates statements in SQL and passes those statements to the data source while shielding you from the need to master SQL. If your query is relatively simple, you might not need to interact directly with Query; instead, you can formulate your request by means of a four-step wizard that acts as a front end to Query.

The first step in creating a query is to form a connection to the data source. Click the Data tab, click From Other Sources (in the Get External Data group), and then click From Microsoft Query. The Choose Data Source dialog box, shown in Figure 23-13, appears.
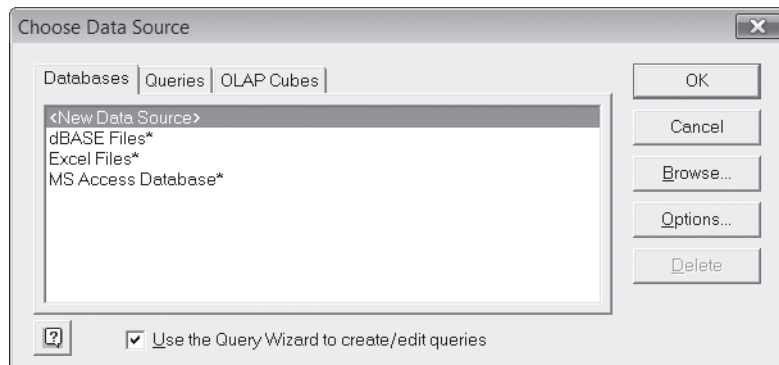
Figure 23-13 The first step in using Microsoft Query is to choose a data source.

You can query a separate Excel file (extracting particular records from a table in that file), a dBase file, or an Access file by selecting one of the options on the Databases tab. To edit an existing query (a .dqy file that has already been created), click the Queries tab. To work with OLAP data, click the OLAP Cubes tab. Otherwise, click

<New Data Source>, and click OK. If you click <New Data Source> and click OK, the Create New Data Source dialog box prompts you to supply a name for the new query, identify the driver for the type of database you are going to query, supply logon information for your connection to the external source, and select the table in the external database you want to use. Click OK to save your changes.

In the following sections, we'll assume for the sake of simplicity that you're going to work with an Access file. After double-clicking MS Access Database in the Choose Data Source dialog box (or selecting that entry and clicking OK), you will see a Select Database dialog box:
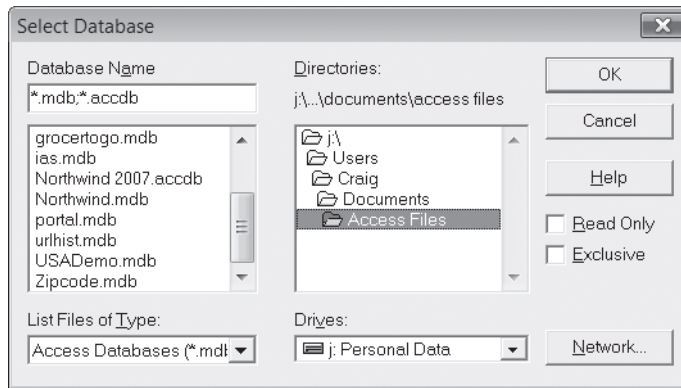
We'll choose the file Northwind.mdb (a sample database that Microsoft included with earlier versions of Access) for this example. After we double-click that file in the Select Database dialog box, the opening page of the Query Wizard, shown in Figure 23-14, appears.
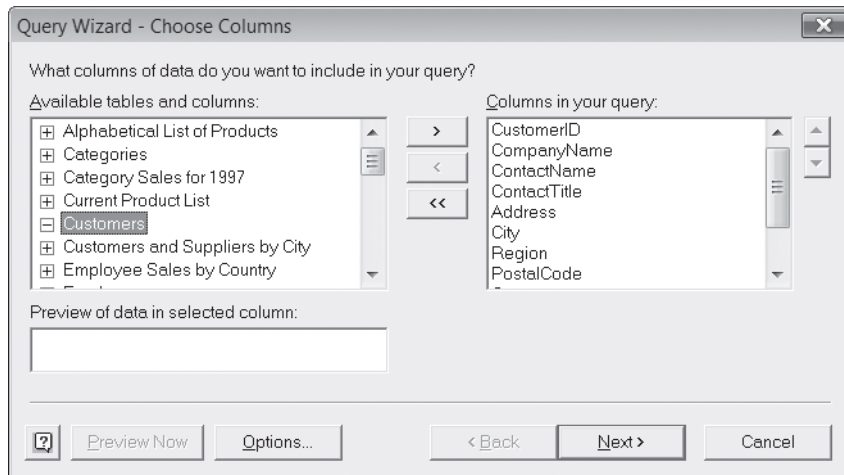
Figure 23-14 The Query Wizard, a friendly front end to Microsoft Query, begins by asking you to choose the columns of data that you want to include in your query.

Choosing Tables and Fields (Columns)

On the first page of the wizard, you see a list of tables on the left and selected fields (Query refers to them as *columns*) on the right. Outline controls (plus signs and minus signs) appear to the left of table names. Your job is to pick the particular fields, from one or more tables, that you want to include in your query.

To add a field to your query, click the plus sign beside the name of the table to which it belongs. This expands the table to reveal its fields. Then select the field, and click the right arrow button to add those fields to your query. (To add all fields from a given table, you can select the table name and click the right arrow button.)

If you add fields from a second or subsequent table to your query, Query performs a *join* operation on the selected tables, if it can. Query joins related tables when it recognizes a primary key field in one table and a field with the same data type (and typically, but not necessarily, the same field name) in the other table. For an example of a query that involves two joined tables, see "Working Directly with Microsoft Query" on page 764.

Filtering Records

After specifying tables and fields and clicking Next, you arrive at the Query Wizard – Filter Data page, shown in Figure 23-15. Here you can specify one or more filter criteria. This is an optional page; if you skip it, Query returns all records from the selected tables.
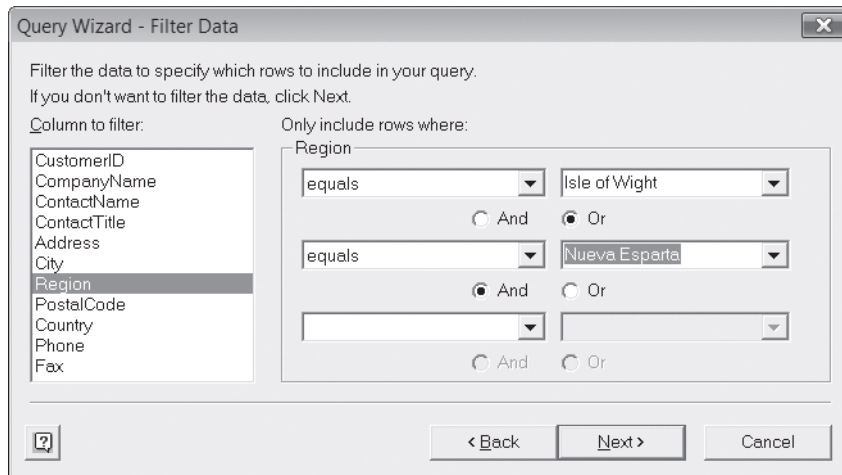
Figure 23-15 Filters, specified on the second page of the wizard, select the records that meet your criteria.

A filter criterion has three components: a field name, a relationship, and a value. You can specify as many as three criteria for each field, connected by And or Or. The list at the left side of the wizard page includes all the names of your selected fields. The lists in the center include available relationships, and the lists at the right include all the

available values for the selected field. Figure 23-15 shows how the second page of the wizard would look if you wanted to see only those records in which the Region field equaled either Isle of Wight or Nueva Esparta.

INSIDE OUT **For More Relationships, Use Microsoft Query Directly**

The Query Wizard offers a long list of relational operators for building filtering criteria. If you use Microsoft Query directly, four additional relationships are available: Is One Of, Is Not One Of, Is Between, and Is Not Between. These additional operators work with two or more values—something the wizard doesn't accommodate. For example, Is Between and Is Not Between both require two values. Is One Of and Is Not One Of can use a list of values. For more information, see "Working Directly with Microsoft Query" on page 764.

TROUBLESHOOTING

The Query Wizard won't let me get rid of a filter.

The Query Wizard is a little clumsy when it comes to letting you remove filters. It doesn't have a Delete button. Clicking Back to return to the previous screen and then clicking Next to return to the Filtering screen doesn't get it done (the previous filter is still there). Clicking Cancel either bails you out of the entire edit process or takes you to Microsoft Query, neither of which is what you probably want. To get rid of a criterion, open its relationship list, and select the blank entry at the top of the list.

If you filter on two more different fields, you'll find that when you select the second field, the wizard removes the first criterion from view. You can tell that you've applied a criterion to a field, however, by looking at the left window on the page. Filtered fields appear there in bold.

Note

Because the wizard accepts up to three criteria per field, you can use it to generate some pretty marvelous filters. But it's a whole lot easier to see what you're doing if you use the full Query interface for multifield filtering. For details, see "Working Directly with Microsoft Query" on page 764.

Sorting Records

After you finish filtering and click Next, the wizard presents its Query Wizard – Sort Order page, shown in Figure 23-16. Sorting is optional, of course. If you decline, Query returns records in the order in which they're stored in the external database field.
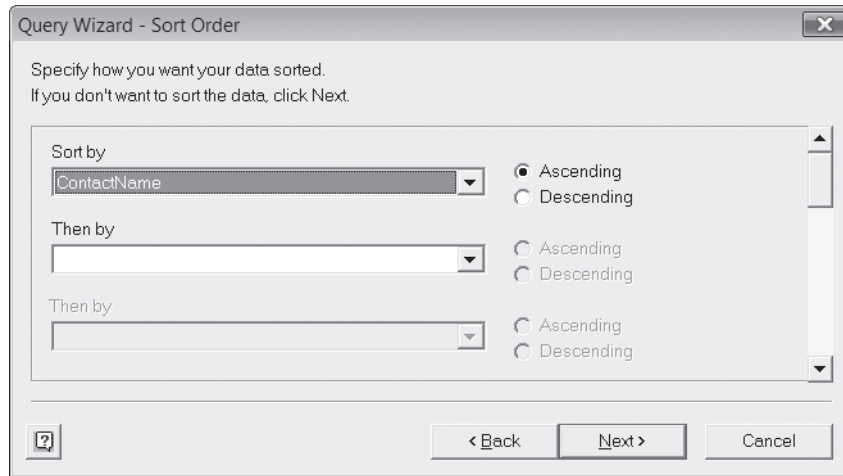

Figure 23-16 Use the Sort Order page to arrange the records that are returned to Excel.

To sort, begin by clicking the Sort By list. There you'll find the name of each field in the table you're querying. Select a field, and then select the Ascending or Descending option to the right of the list. You can sort on as many fields as you want. To remove a sort item, select the blank entry at the top of the list. In Figure 23-16, we've asked for records sorted in ascending order by ContactName. Click Next.

Saving the Query or Moving to Microsoft Query

The Save Query button, on final page of the wizard (see Figure 23-17), lets you name and save your query as a DQY file. The resulting DQY file encapsulates all the selections you've made in the construction of your query—your choice of tables and fields, your filters, and your sorting specifications. Note that this is different from the ODC file you might have made earlier. An ODC file records information required to achieve a connection with an external data source; a DQY file records query specifications.

The View Data Or Edit Query In Microsoft Query option on the last page of the wizard lets you move to the full Microsoft Query for further processing. For information about why you might want to do this and how to use Query, see the following section. If you don't want to move on to Query, select Return Data To Microsoft Office Excel, and click Finish. Click OK in the Import Data dialog box to import the data based on your query.
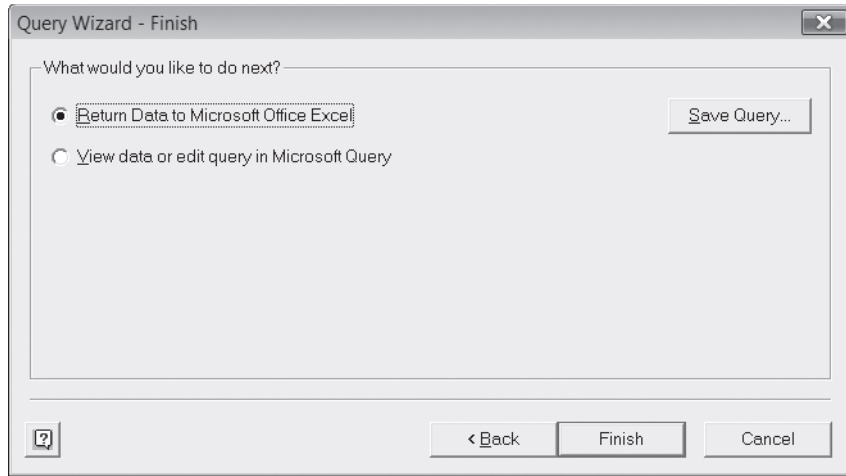
Figure 23-17 Save the query if you want, and indicate whether you want to return directly to Excel or go on to the full Microsoft Query for further processing.

## Working Directly with Microsoft Query

The Query Wizard is an ideal tool for creating relatively simple queries, but it doesn't provide access to all the power of Microsoft Query. You'll need to work directly with Query if your query uses criteria involving calculations (other than simple comparisons) or if you want to create a query that prompts the user for one or more parameters when run. Query, but not the Query Wizard, also lets you do the following:

- Filter on the basis of fields that you don't intend to import into Excel—that is, fields that are not included in the *result set*, the records that meet your current criteria.

- Filter using Is One Of, Is Not One Of, Is Between, or Is Not Between.

- Limit the result set to unique entries.

- Perform aggregate calculations, such as totals or averages.

- Create your own joins between tables.

- Edit a query's SQL code.

Getting to Query

If you have already stored the query you want to edit in a DQY file, you can open it in Microsoft Query using either of the following methods:

- Click the Data tab, and then click Existing Connections. Your query will appear in the list, alongside your ODC files, marked by a distinctive icon—two intersecting blue rectangles.
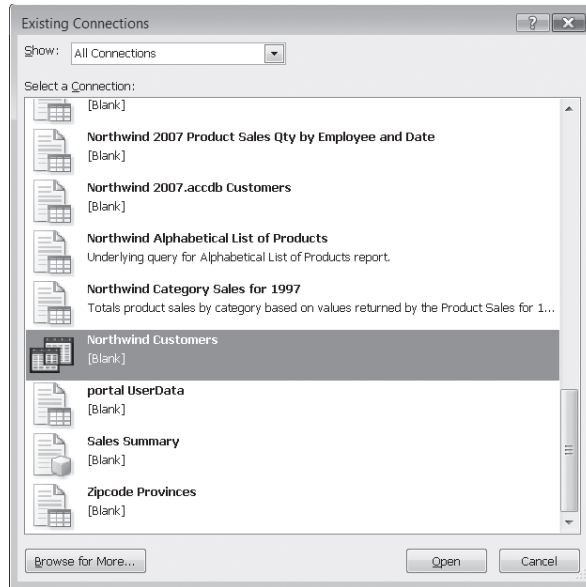
- Click the Data tab, and then click From Other Sources. On the menu that appears, click From Microsoft Query. In the Choose Data Source dialog box, click the Queries tab. Your query should appear there:
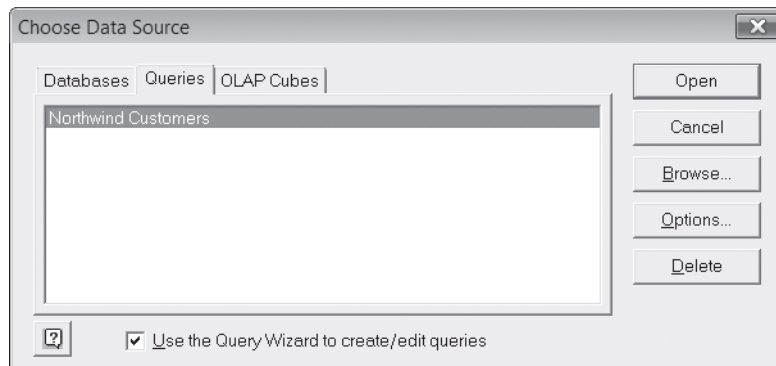
If you have just finished creating your query in the Query Wizard and want to open it in Microsoft Query for further editing, select View Data Or Edit Query In Microsoft Query on the final page of the wizard (see Figure 23-17), and click Finish.

Figure 23-18 shows Query with a query against three tables from Northwind.mdb. The tables are Products, Categories, and Suppliers. The Products table is joined to the Categories table in the CategoryID field and to the Suppliers table in the SupplierID field. The query shows selected fields from these tables, revealing products by category and supplier (CompanyName), along with some price and inventory information.
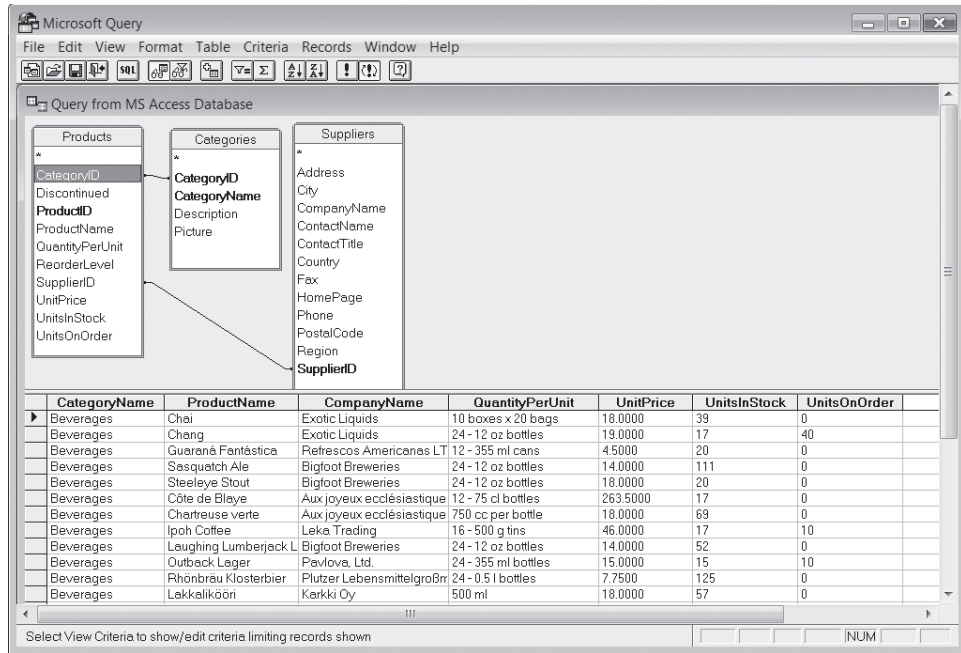
Figure 23-18 We're using Query to edit a query against three tables in Northwind.mdb.

Note that the Query window is divided horizontally into two panes—an upper pane for tables and a lower one for data. The tables pane shows a window for each table that's currently involved in the query. The data pane shows the *result set*—the collection of records that meet the criteria. (At the moment, we haven't defined any filters, so all records in the three tables are included.)

Shortly, you'll see that Query can also accommodate a third pane, in which you specify filtering criteria. All these panes, as well as the individual table windows, are independently sizable and movable. We've bumped the data pane down a bit from its default position to make more room for the Suppliers and Products tables, and we've stretched the windows in which those tables are displayed so that we won't have to scroll to see all their fields. You'll find that Query seldom gives you an ideal window layout when it starts, so you'll want to manipulate it to get the view you need.

Adding and Removing Tables

To add a table to the data pane, click Table, and then click Add Tables. The Add Tables dialog box lists all the tables available in the data source you're using. To add a table, select it, and click Add. You can add as many as you like before closing the Add Tables dialog box. To remove a table, select it in the table pane, and then click Table, Remove Table.

Working with Joins

If Query doesn't already have your tables joined appropriately, you can create your own joins by dragging. If you click a field in one table and then drag to a field in another, Query creates a join based on those fields and draws a line to indicate that it has done so. You can create, inspect, and modify joins by double-clicking any join line or by clicking Table, Joins. Figure 23-19 shows the Joins dialog box for the query shown in Figure 23-18.
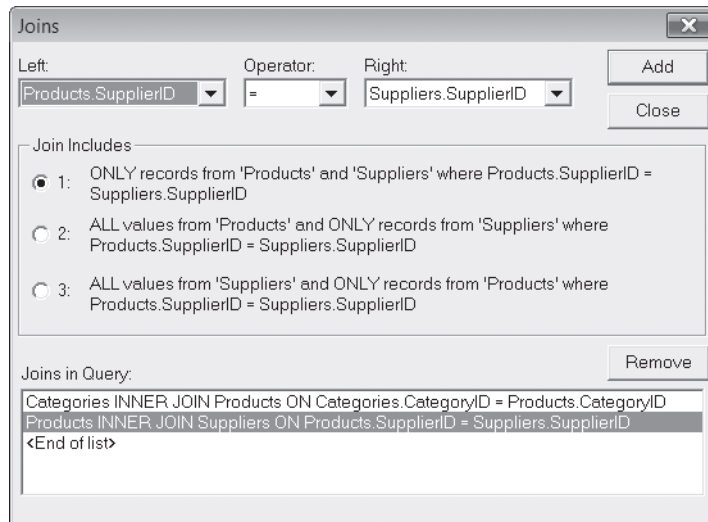
Figure 23-19 The Joins dialog box tells you exactly how your tables are joined and lets you modify the joins or create new ones.

If you're not sure what's joined to what or what the effect of a join is, it's a good idea to visit the Joins dialog box. The Join Includes area in the dialog box provides a pretty clear description of what's happening. By working with the Left, Operator, and Right fields, you can also modify the ways in which your tables are joined.

Adding, Removing, and Moving Fields

To add a field to your data pane, double-click it in a table window. To add all fields to the data pane from a table, double-click the asterisk at the top of the table window.

To remove a field, select its heading (this action selects the entire field), and press Delete. To move a field from its current location, first select its heading, and then drag it to the position you want.

INSIDE OUT **Hide Selected Fields Without Removing Them from the Query**

If you find yourself scrolling horizontally a lot but don't want to rearrange your fields, you can hide fields that you temporarily don't need to see. Select a field, and then click Format, Hide Columns. To redisplay a hidden field, click Format, Show Columns; select the field in the Show Columns dialog box; and then click Show.

Renaming Fields

By default, Query uses the names of your fields as field headings. If these field names are short and cryptic, you might want to supply different headings.

Select the column you want to change, and then click Records, Edit Column. In the Edit Column dialog box, type a new heading in the Column Heading box, and then click OK.

Sorting the Result Set

Query initially displays records in the order in which they are stored in the external data source. You can change their order by clicking Records, Sort. Figure 23-20 shows the Sort dialog box with the CategoryName field selected. (The dialog box, like others in Query, qualifies field names with the tables to which they belong; it says Categories.CategoryName because the CategoryName field is part of the Categories table.)
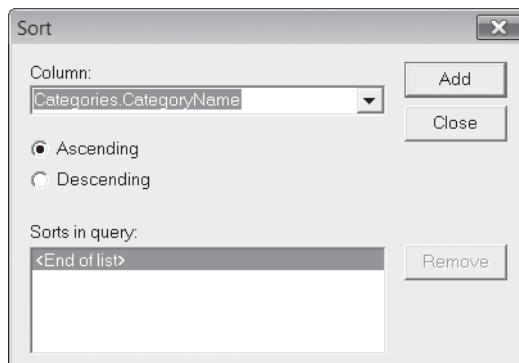
Figure 23-20 The Sort dialog box displays the current sort order and lets you add fields from a list.

The Sorts In Query section in the dialog box indicates what sort specification, if any, is currently in effect. In Figure 23-20, the list is empty, indicating that the result set is currently unsorted. The Column list at the top of the dialog box lists all the table fields available for sorting. When you add a field to the Sorts In Query list, Query performs the sort immediately but leaves the dialog box open in case you want to sort on additional fields. You can sort on as many as you please.

For multiple-field sorts, sort first on the most important sort field. Then sort on your secondary field, and so on. Figure 23-21 shows the result set sorted first by Suppliers.CompanyName and then by Products.ProductName. (The Asc abbreviation in the Sorts In Query list indicates ascending sorts.) The records now are alphabetized by supplier, with records of a common supplier alphabetized by product name.
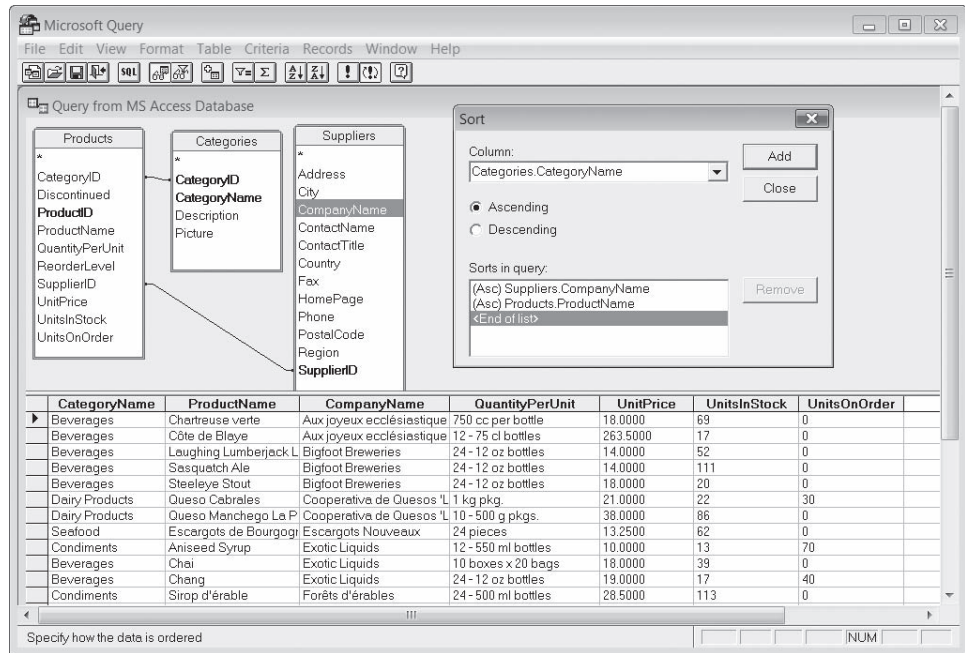
Figure 23-21 We've sorted first by the supplier's company name and then by product name.

When you click Add in the Sort dialog box, Query adds your new sort field above the currently selected field in the Sorts In Query list. If you accidentally add a field in the wrong order, select it, and click Remove. Then add the field in the correct position.

Sorting with the Toolbar   The Sort icons on the Query toolbar work differently from the Sort command on the Ribbon in Excel. You can add a sort to the current list by holding down the Ctrl key when you click a Sort icon. If you do not hold down Ctrl, clicking a Sort icon replaces the current sort with the new one.

Filtering the Result Set

Query provides a variety of methods by which you can filter the result set so that it includes only the records in which you're interested. As with the Query Wizard, you create a filter by specifying one or more criteria—conditions that particular fields must meet.

Creating Exact-Math Criteria   The simplest kind of criterion is one in which you stipulate that a field must exactly equal some value. Query makes it extremely easy to create such criteria:

1.   Select a field value that meets your exact-match criterion.

2.   Click the Criteria Equals button on the toolbar:
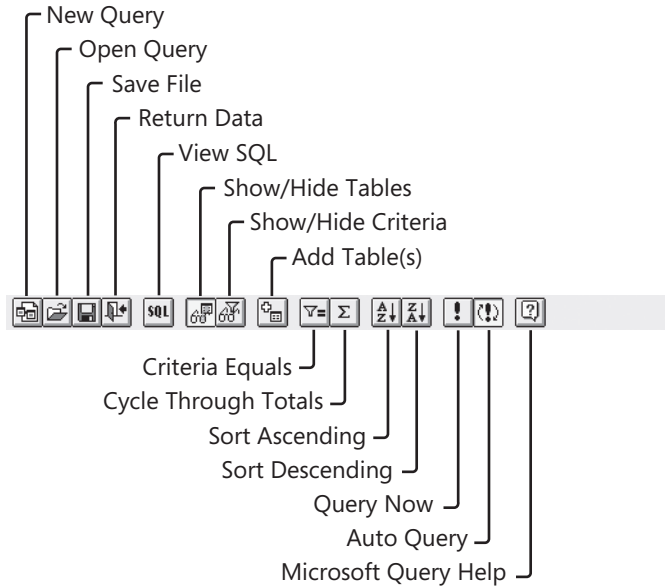
For example, suppose you want to filter the result set shown in Figure 23-21 to include only those records in which the CompanyName field is Bigfoot Breweries. To do this, select any record with the CompanyName field that already equals Bigfoot Breweries, and click the Criteria Equals button. Query responds by displaying the criteria pane (if it's not already displayed) and applying the new filter to the table, as shown in Figure 23-22.

> **Note**
>
> After you use the Criteria Equals button to specify an exact-match criterion, you can quickly switch to a different match. Type a new value in the criteria pane to replace the current one.

If you have used the Advanced Filter command (see "Using the Advanced Filter Command" on page 689), you'll notice that the criteria pane in Query looks a lot like a criteria range in an Excel worksheet. Field headings appear in the top row, and criteria are stated in subsequent rows. Although you can type new criteria or edit existing ones

directly in the criteria pane, it's not necessary, because the Query menu commands take care of entering information in the criteria pane for you. In fact, you don't need to have the criteria pane on your screen at all.
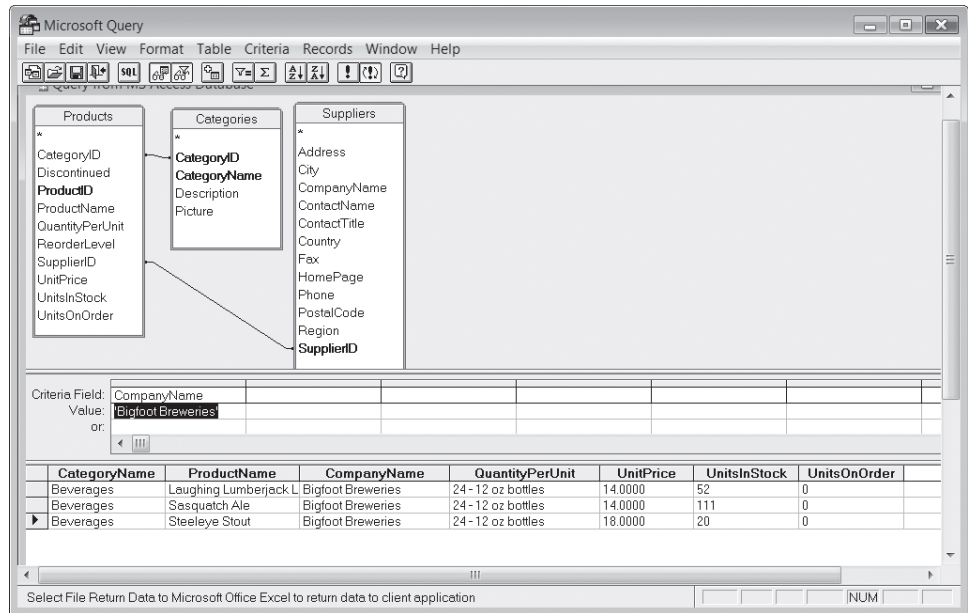
Figure 23-22 When we clicked the Criteria Equals button, Query displayed the criteria pane and applied the filter to the result set.

> **Note**
>
> To remove the criteria pane, click the Show/Hide Criteria button on the toolbar in Query, or click View, Criteria. To remove the tables pane, click Show/Hide Tables, or click View, Tables.

Using Multiple Exact-Match Criteria   To generate a query that uses exact-match criteria in two or more fields, repeat the process just described for the second and each subsequent criterion. For example, to filter the result set in Figure 23-21 to show only those records with CompanyName equal to Exotic Liquids and CategoryName equal to Beverages, select Exotic Liquids in the CompanyName field, click Criteria Equals, then select Beverages in the CategoryName field, and click Criteria Equals again. As Figure 23-23 shows, the criteria pane then shows the two criteria on the same line. Just as with an Excel criteria range, Query treats criteria on the same line as if they are connected by the AND operator.
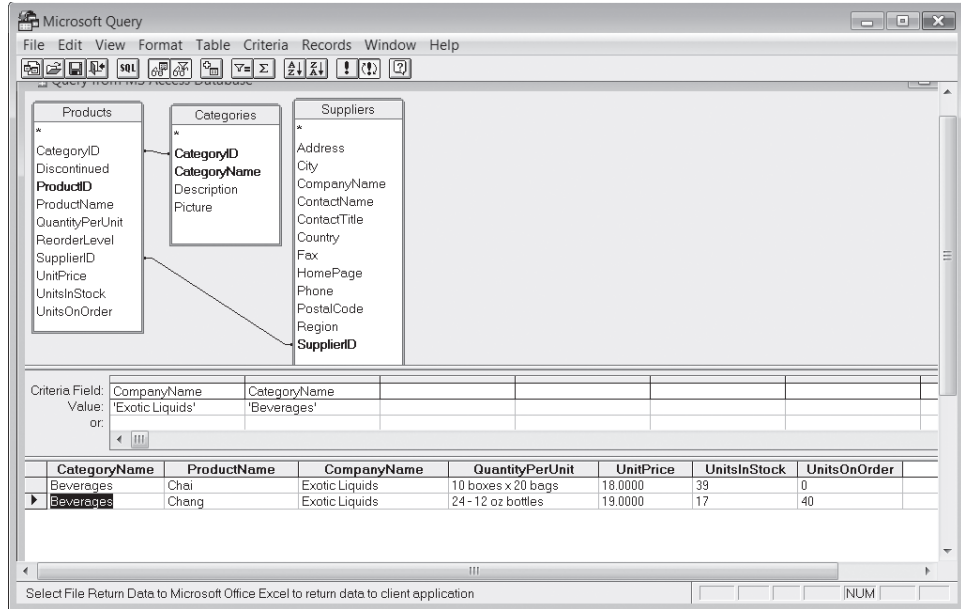
Figure 23-23 When you use the Criteria Equals button in two separate fields, the filter in Query admits only those records that meet both criteria.

Automatic Query vs. Manual Query

By default, Query updates the result set every time you add a new field to the data pane, rearrange the order of the existing fields in the data pane, change a sort specification, or change a filter criterion. (If you're working in the criteria pane, the query is executed as soon as you click away from the current criteria-pane call.) In response to these actions, Query creates a new SQL statement and executes that statement against your data source. (You can see the SQL code—and edit it, if you are inclined—by clicking the View SQL button on the toolbar.) If your data source is particularly large or network traffic is high, Automatic Query can cause annoying delays. You can turn off the Automatic Query feature so that Query executes the current SQL statement only when you ask it to do so.

You can determine whether Automatic Query is on by verifying whether the Auto Query button on the toolbar is selected (has a "pushed in" appearance). To turn the feature off, click the Auto Query button, or click Records, Automatic Query.

To execute the current query in manual mode, click the Query Now button, or click Records, Query Now.

Using Menu Commands to Specify Exact-Match Criteria   If you'd rather use menu commands than toolbar buttons, you can specify an exact-match criterion as follows:

1. Select a field value that meets your specification.

2. Click Criteria, Add Criteria. In the Add Criteria dialog box, click Add.

Removing Criteria   The simplest way to remove a filter criterion is to select the criterion's heading in the criteria pane and press Delete. To remove all criteria and restore the unfiltered result set, click Criteria, Remove All Criteria.

Specifying Comparison Criteria   To specify a comparison criterion, follow these steps:

1. Click Criteria, Add Criteria. You'll see a dialog box similar to the one shown in Figure 23-24.
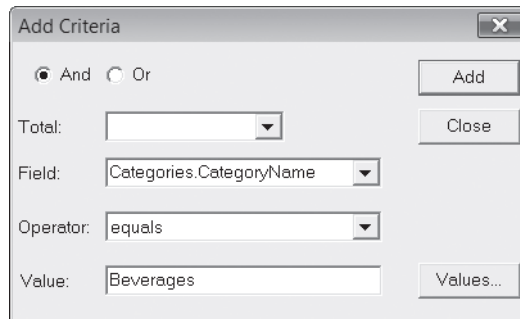
Figure 23-24 The Add Criteria dialog box lets you select fields, comparison operators, and values.

In the Add Criteria dialog box, you can construct your criteria by selecting options from various lists. For example, you can select a field from the Field list and then select an operator, such as Is Greater Than, in the Operator list. You can also enter a value in the Value text box by typing it or clicking the Values button and selecting from the list.

> **Note**
>
> For comparison criteria that don't involve computed fields, be sure the Total field in the Add Criteria dialog box is blank, as it is in Figure 23-24. For more information about the Total field, see "Filtering on Calculated Fields" on page 777.

2. When you have filled out the Field, Operator, and Value fields, click Add.

    Query responds by creating the appropriate entry in the criteria pane and, if Automatic Query is on, executing the new query. The Add Criteria dialog box remains open so you can specify more criteria.

Chapter 23

3. To add another criterion, select the And option or the Or option at the top of the dialog box, and then type the information as before.

4. When you've finished typing criteria, click Close.

Filtering on Fields That Are Not in the Result Set   Your filter criteria can be based on fields that are not currently displayed in the result set. The Field list in the Add Criteria dialog box includes all fields in all active tables, not only the fields you plan to return to Excel.

Limiting the Result Set to Unique Entries   To limit the result set to unique entries, click View, Query Properties. In the Query Properties dialog box, select Unique Values Only. You can make this selection before or after you create your filter.

Comparing Fields   Your comparison criteria can compare the value in one field to that in another. For example, to display records where UnitsInStock is less than Reorder-Level, you fill out the Add Criteria dialog box as shown in Figure 23-25. Note that you have to type a field name in the Value box.
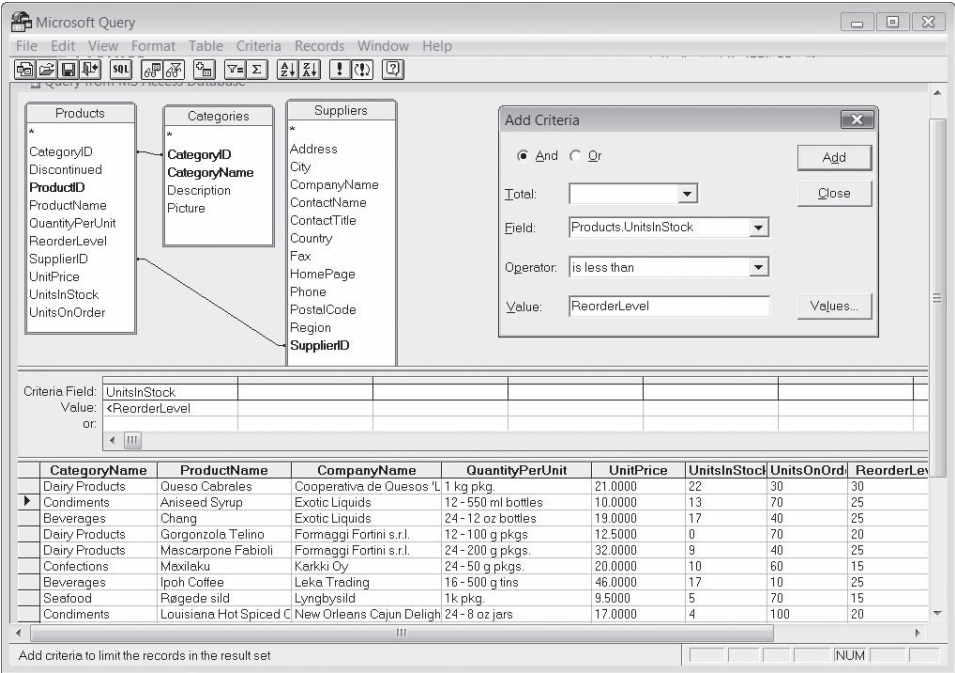
Figure 23-25 This criterion returns records where units on stock are below the reorder level.

Performing Aggregate Calculations

You can analyze your results thoroughly after you get the data back onto the Excel worksheet. If you prefer, however, you can have Query do some of the calculating for

you. With Query, you can make aggregate calculations (sums, averages, counts, and so on) the basis of filtering criteria.

Query refers to all calculations as totals, although summing values is only one of the options available. The aggregate functions that are common to all database drivers are AVG (average), COUNT, MIN (minimum), and MAX (maximum). Your driver might support additional functions.

Clicking Through the Totals   One way to perform aggregate calculations is by clicking the Cycle Through Totals button on the toolbar in Query. For example, to find the total of the UnitsOnOrder field, follow these steps:

1. Display the UnitsOnOrder field in the data pane, and remove all filtering criteria from the criteria pane.

2. Select the UnitsOnOrder field, and click the Cycle Through Totals button.

As Figure 23-26 shows, Query responds by displaying the total in the data pane.
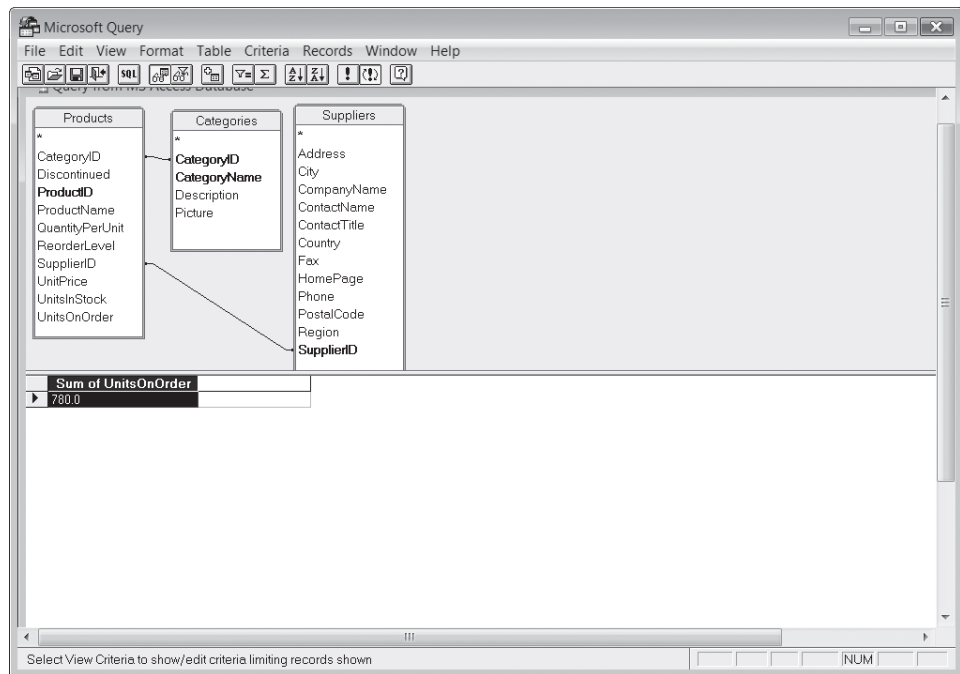
Figure 23-26 We used the Cycle Through Totals button to calculate the total units on order.

Cycling Through the Functions   In the previous example, clicking Cycle Through Totals a second time changes the aggregate function from SUM to AVG, and the number shown in the data pane changes accordingly. Successive clicks on the Cycle Through Totals button result in the count, the minimum, and the maximum. One more click returns the result set to its unaggregated state.

> **Note**
>
> Not all the aggregate functions are available for every field type.

Using Menu Commands If you prefer menus to tools, you can use the Edit Column command:

1. Click Records, Edit Column. (Alternatively, double-click the field heading.)

2. In the Edit Column dialog box (see Figure 23-27), select the function you want from the Total list.
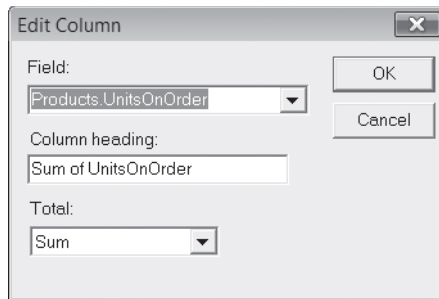
Figure 23-27 Instead of clicking Cycle Through Totals, you can click Records, Edit Column.

Aggregating Groups of Records In addition to grand totals, you can also calculate totals for groups of records. For example, to find out how many units are on order for each supplier company, do the following:

1. In the data pane, display the CompanyName field followed by the UnitsOnOrder field.

2. Select the UnitsOnOrder field, and click Cycle Through Totals.

As Figure 23-28 shows, Query displays one record for each company and shows the total units on order for each.

Using More Than One Aggregate Field You can add as many aggregate fields to your result set as you need. To display both sums and averages for a numeric field, for example, you can drag that field to the data pane twice. Click the Cycle Through Totals button to apply the function you want to each copy of the field.
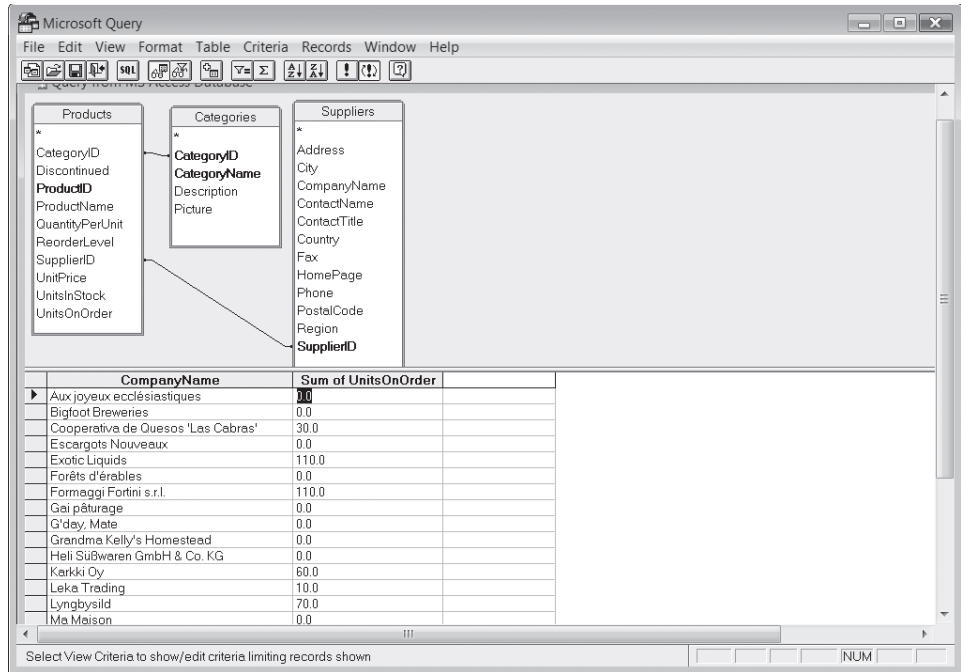
Figure 23-28 You can apply aggregate calculations to groups of records; here we calculated the total units on order per company.

Filtering on Calculated Fields   A field that performs an aggregate calculation is called a *calculated* field. You can use calculated fields as the basis for filtering criteria. To base a criterion on a calculated field in the Add Criteria dialog box, use the Total list to select the function you want. (If you're entering the criterion directly in the criteria pane, type the function name, and enclose the field name in parentheses.) Figure 23-29 shows a criterion that returns the names of companies for whom the total number of products on order is greater than or equal to 20.

Creating a Parameter-Based Query

A *parameter-based* query is one in which a filter criterion is based upon a value supplied by the user when the query is executed. To create such a query, first turn the Automatic Query feature off by clicking the Auto Query button on the toolbar. Then specify a criterion in the usual way—either by using the Add Criteria dialog box or by typing values directly in the criteria pane. Instead of typing a value, though, type a left bracket character, a prompt of your choosing, and a right bracket character. (The prompt must not be identical to the field name, although it can include the field name.) When you execute the query, either from within Query or from within Excel, a dialog box containing your prompt appears. Figure 23-30 shows a parameter-based query.
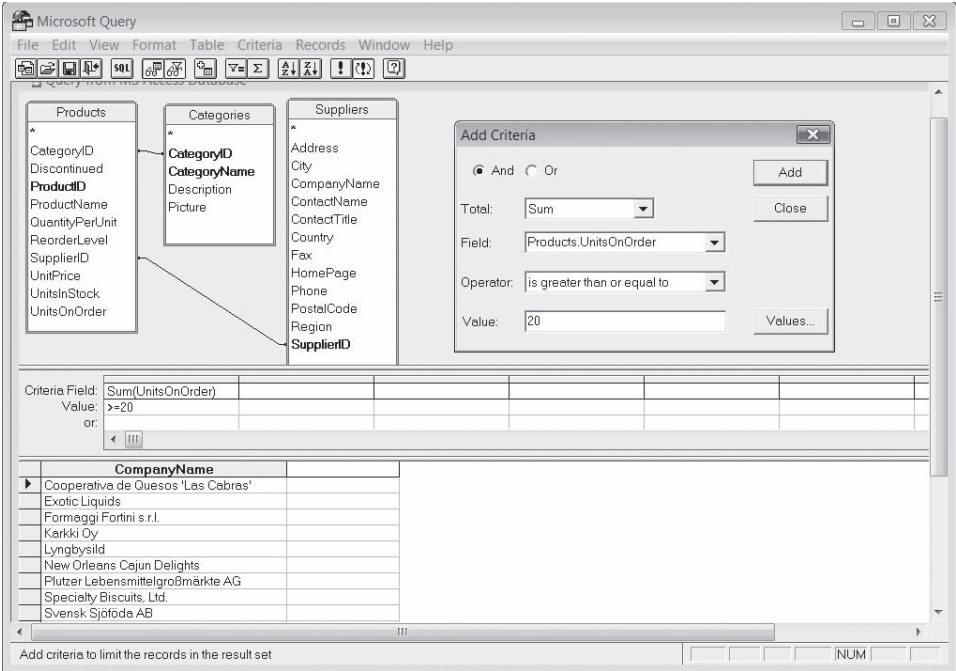
Figure 23-29 We've filtered the supplier list to show only those companies with 20 or more products on order.

Saving a Query

To store your query specification in a reusable DQY file, click File, Save. This step is optional. If you do not save the query, you will still be able to refresh it from the data range that it creates on your Excel worksheet. You will have to re-create it if you want to use it in another workbook, however.

Returning the Result Set to Excel

To return your data to Excel, click File, Return Data To Microsoft Excel. The Import Data dialog box (refer to Figure 23-2) appears, asking you where and how you want the data returned.
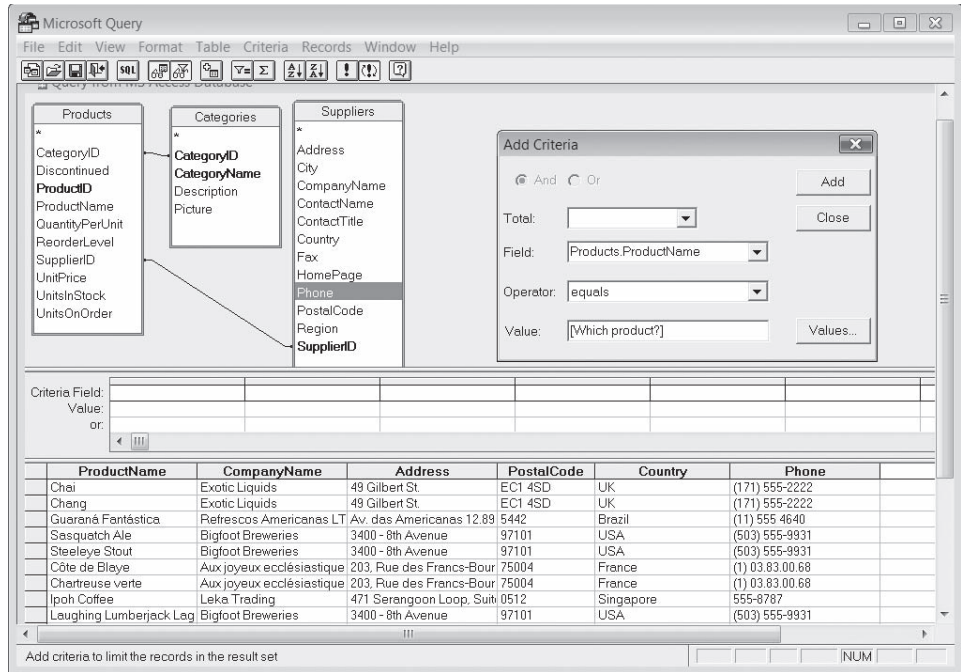
Figure 23-30 When executed, this query will prompt the user for a product name.

Using a Web Query to Return Internet Data

Web queries let you grab specific information, such as stock prices, sports scores, or your company's current sales data, from the Internet or an intranet. You can set up queries to prompt you for the data you want (for stock ticker symbols, for example) or to get the same information every time they're executed. You can try Web queries using a set of sample queries that come with Excel 2007.

The Excel graphical interface for creating Web queries lets you build a query by pointing to the data you want. You can refresh the query at any time or at regular intervals, and you can save the query in an IQY file for reuse in other worksheets. You do not need to understand how the target Web page is built to construct a query to it.

Using an Existing Web Query

To run an existing Web query—one of the samples supplied with Excel or one that you or someone else has already set up—click the Data tab, and click Existing Connections. As Figure 23-31 shows, Web queries are identified in the Existing Connections dialog box by a pair of intersecting blue rectangles and a globe. (The three Web queries supplied with Excel 2007 are all from MSN MoneyCentral Investor.)
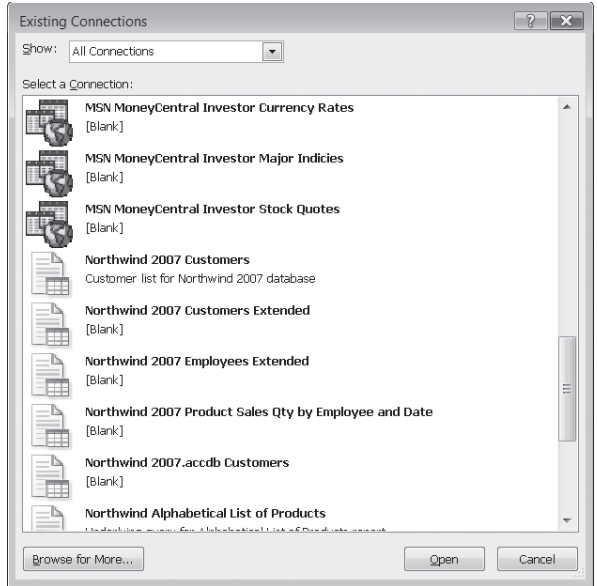
Figure 23-31 Web queries are marked by intersecting rectangles and a globe, and Excel 2007 includes three of them to try.

When you double-click a Web query, Excel prompts you to specify a location for the incoming Web data. Depending on how the query was set up, it might also prompt you for parameters. For example, if you double-click the MSN MoneyCentral Investor Stock Quotes query, after you indicate where you want the data to go, you will see the following dialog box, which prompts you to supply one or more stock symbols:
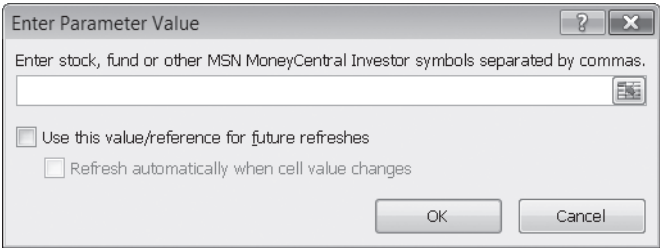
To supply parameters via the Enter Parameter Value dialog box, simply type in the box. If you prefer, you can point to a worksheet range containing your parameters. If you specify a multicell range, Excel parses the range moving across and then down.

If you use a worksheet range to feed parameters to your Web query, you can also stipulate that the query be refreshed automatically anytime the worksheet range changes. To do this, select both check boxes.

Figure 23-32 shows an example of data returned by one of the Web queries supplied with Excel 2007. Note that this query is set up to return the names of market indexes as

hyperlinks. Clicking a hyperlink takes you to a relevant page in the MSN MoneyCentral Investor Web site.

Figure 23-32 Data returned by this Web query includes hyperlinks to the MSN MoneyCentral Investor site.

Creating Your Own Web Query

Excel provides three easy ways to construct a Web query:

- Clicking the Data tab and then clicking From Web

- Copying and pasting information from your Web browser

- Right-clicking in Microsoft Internet Explorer and clicking Export To Microsoft Excel

Using the From Web Command

To create a Web query using the From Web command, follow these steps:

1. Click the Data tab, and then click From Web. The New Web Query form that appears (see Figure 23-33) is a specialized Web browser, and your home page appears in its window.
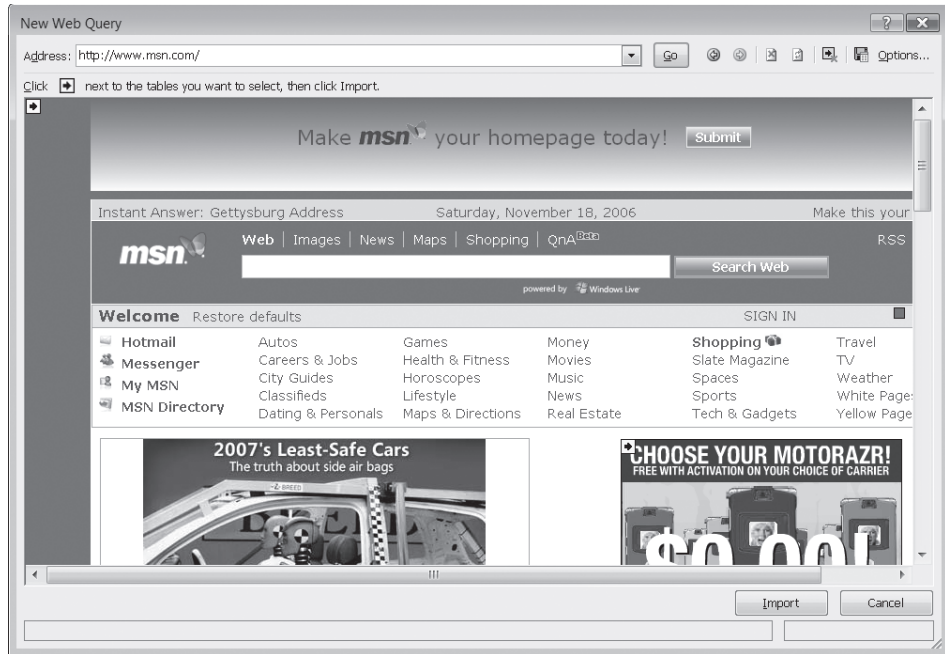
Figure 23-33 The New Web Query form is a specialized Web browser.

2. If you know the Uniform Resource Locator (URL) of the Web site you want to query, you can type or paste it into the Address field. (Unfortunately, the New Web Query form doesn't include a Favorites menu.)

3. Click Go; your Web site appears in the main window, as Figure 23-34 shows.

Yellow boxes with arrows appear along the left edge of the window. Each of these boxes represents a section of the Web site that you can import into Excel. As you rest your pointer on any of these yellow boxes, a thick bounding rectangle indicates the section of the site you will be importing if you select the yellow box. You can select any or all yellow boxes.

Saving the IQY File and Setting Formatting Options   After making your selections, you can click Import to transform your selection into a query. But before you do so, you might want to save the query (making it a reusable IQY file) or explore the menu of options. To save the query in its current form, click the Save Query command, directly to the left of the word *Options* on the toolbar in the New Web Query form. To set options, click Options. Figure 23-35 shows the Web Query Options dialog box.

Figure 23-34 We've displayed the site we want to query in the New Web Query form.
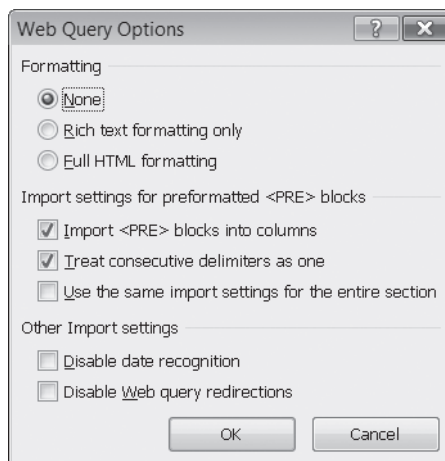
Figure 23-35 In addition to offering other features, the Web Query Options dialog box lets you control how much of the Web site's formatting Excel should preserve.

In the Formatting area of the dialog box, select None to import the data as plain text. Select Rich Text Formatting Only to preserve hyperlinks and merged cells in the Web

data. Select Full HTML Formatting to retrieve as much as possible of the original Web site's formatting. Figure 23-36 shows a Web site queried with full HTML formatting. Note the inclusion of hyperlinks in the downloaded data.

Figure 23-36 Because we queried this site using full HTML formatting, the downloaded data includes active hyperlinks and other welcome formatting characteristics.

Copying and Pasting from the Web Browser

The method for creating a Web query just described is fine if you're starting in Excel and you know the address of the target site. But you can also start from the Web browser. Select the data you want, press Ctrl+C to copy it, open a new Excel worksheet, and press Ctrl+V. You'll see a smart tag near the lower-right corner of the pasted data. Open the menu, shown here, and click Create Refreshable Web Query. Your data selection will appear in the New Web Query form.
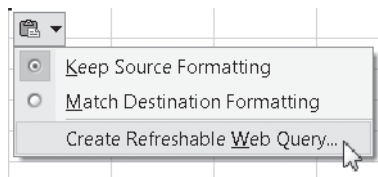
Using the Internet Explorer Export To Microsoft Excel Command

If your Web browser is Internet Explorer, you can also create a Web query by right-clicking a Web page and clicking Export To Microsoft Excel. (If you don't see this command, you probably already have something selected in Internet Explorer. Clear your selection, and try again.) The Export To Microsoft Excel command begins by creating a new instance of Excel (it does this to avoid overwriting Excel data you might already be working with). If you right-clicked something that Internet Explorer recognizes as an HTML table, it transfers that table directly onto Sheet1!A1 as a new Web query. If you clicked anywhere other than an HTML table, the command opens the New Web Query form.