How to Excel at data extraction

AC3321  Data extraction is the Autodesk® AutoCAD® software feature that enables you to liberate your drawing data for reuse. You can use the Data Extraction tools to export data from a drawing into formats such as Microsoft® Excel® and Access®, and then link the data back into AutoCAD. You can also link data directly to a table in the drawing. These Data Extraction tools offer the serious power user an incredibly powerful suite of tools for exchanging data, eliminating data entry errors, and increasing efficiency and effectiveness. This class explains the workings of data extraction using real-world examples of how it should be applied in a production environment. The class covers the full range of system capabilities, including extracting, linking, and advanced table formatting.

Learning Objectives
At the end of this class, you will be able to:

- Extract data from drawings for reuse in tables or other drawings
- Link drawings to Excel spreadsheets
- Create tables from extracted data
- Improve reliability of data by minimizing the potential for human error

About the Speaker
With more than 25 years of engineering and design experience, Martin has worked in a variety of roles including Civil Infrastructure Designer both in the UK and Australia on some of the largest infrastructure projects at the time. In his role as CAD Manager Martin has helped bridge the capability gap in business, providing and managing systems that deliver significant capability and performance to design teams.

Martin is the Founder of Duke Systems L.L.C. Duke Systems is a CAD consultancy, with global experience, located in the San Francisco Bay Area able to support all of your CAD requirements.

An accomplished and confident presenter, Martin has presented at Autodesk University in 2010, 2011 and 2012. And at Bentley Learn in 2013

martin.duke@duke-systems.com
Introduction
Data Extraction is an important new feature that enables the user to liberate their drawing data for reuse, either within the drawings, or to an external file such as an Excel spreadsheet or Access database. Data Extraction is an important tool because it enables the user to convert graphical type information into tabular data and at the same time also merge that data with other external data or summarize it or manipulate it in other ways.

Data Extraction can be used in two main separate ways.

- It can be used to extract attributes from blocks within drawings
- It can be used to extract properties from objects within the drawings

Data extraction can be used both within the main drawing, or can be used on other external drawings. Output can either be to a table within a drawing or to an external file.

Extracting data from drawings for reuse in tables or other drawings

In AutoCAD the commands and functionality around Data Extraction can be found on the Insert Palette on the Linking & Extraction Panel of the standard ribbon.

The data extraction process can either be started from the ribbon with the Extract Data button from the command line with the DATAEXTRACTION command. Whichever command you chose the AutoCAD will initiate an 8 step wizard to guide the user through the process.
**Data Extraction Wizard steps**

- **Page 1**
  
  The start page lets the user either create a new data extraction, or modify an existing one. When creating a new extraction a previous extraction can be used as a template.

- **Page 2**
  
  The Define Data Source page lets the user define where the objects that will be queried will come from. The data can either come from the current drawing or from external drawing files, or a sheet set.

  The Settings button additionally allows the user to refine origin of the source data.
How to Excel at Data Extraction

Page 3

The Select Objects page presents the user with a list of available objects types in the drawing data set.

At this point is important to have a clear distinction between extracting attribute data and doing materials take off.

Page 4

Select Properties page defines which properties of the selected objects you wish to extract.

The Category Filter on the right side of the panel is an easy way to filter out the properties into specific categories.

Page 5

The Refine Data page allows the user to format and manipulate the data into the required presentation style.

Also the user can also chose to link their data to some previously defined external data.
Link External Data

The Linking to External Data panel lets the user define a relationship between the extracted data and some external data in a spreadsheet. The external data is then included with the extracted data in a table.

Page 6

The Choose Output page lets the user select how they would like to output the data, to either a spreadsheet or to a table on the drawing.

Page 7

The Table Style Page lets the user allocate a table style for the data and fine tune any extra formatting of the table.
Link your data to Excel spreadsheets

Linking your data to an Excel spreadsheet occurs on Page 5 on the wizard. There are two main parts to the panel here:

1. Data Linking: The process of selecting a Microsoft Excel spreadsheet with the purpose of merging all or some of its contents with extracted drawing data
2. Data Matching: Creating a relationship between two sets of data; the data in a column from the extracted drawing data and the data in a column from an Excel spreadsheet.

Data Link

The Data Link option lets the users select an existing Data Link, or to create a new one. A data link is created through the Data Link Manager which can be accessed from the Linking & Extraction Panel. A data link is similar to an External Reference file, and in fact appears in the External References palette along with conventional Xrefs and Images. There are three steps to creating a Data Link:
Step 1 involves starting the Data Link Manager and Selecting the Create a New Excel Data Link options from the tree view. The user will be asked to supply a name for the data link before proceeding to step 2.

Step 2 involves defining the file and path type for the connection. It is always recommended to use the Relative Path option.

Step 3 allows the user to define the actual data in the spreadsheet to use, it can be either:

- An entire sheet
- A named range
- A specified range

Formatting options are also available on the extended part of the panel by click the expand button. Typically the default format options are sufficient.
Data Matching

Data matching defines the relationship between the drawing data and external data. The user is asked to select one column from each that will define a one to one relationship. Once the columns are defined the user can test the relationship with the check Match button. The match will fail if it cannot find a valid match between the drawing data and the external data.

There are two possible error messages:

No match was found between any values in the drawing data and external data.

If this is not resolved the external data will not be merged. Re-examine the data to check that the right column has been selected and that the data if numerical has the same level of precision

The data column in the external data does not contain unique values.

If there are multiple matches the program cannot choose which one to use.

Create tables from extracted data

Tables can be created by a variety of means:

- From an empty table (the default option, user must enter all the data)
- From a data link
- From object data in the drawing (Data Extraction)
The second and third options are the ones we are interested in.

Create a Table from a data link

If a data link already exists the users selects the link from the drop down list and then places the table as normal. If a data link does not exist, one can be created at this time.

Create a table from object data in the drawing (Data Extraction)

The third option when selected takes the user directly to the DATAEXTRACTION command once the OK button has been selected. From there the process is as defined above.
Improve reliability of data by minimizing the potential for human error

Using Data Extraction minimizes the potential for human error by automating the tasks usually associated with material take off and data entry processes. By having the computer perform these functions there is certainty that the results will be correct, however there are still several traps for users to be aware of.

1. Out of date data
2. Incorrect object filter or selection.

Out of date data is best avoided by ensuring that the DXEVAL variable is set correctly for your situation. The table below shows the trigger values for the system to compare the data extraction tables against the source data. If the data is out of date AutoCAD will display an update notification. The value stored in DXEVAL is a sum of the options in the table. If you wished to have the table checked on drawing open, save and plot you would set DXEVAL to 7

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No notification</td>
</tr>
<tr>
<td>1</td>
<td>Open</td>
</tr>
<tr>
<td>2</td>
<td>Save</td>
</tr>
<tr>
<td>4</td>
<td>Plot</td>
</tr>
<tr>
<td>8</td>
<td>Publish</td>
</tr>
<tr>
<td>16</td>
<td>eTransmit/Archive</td>
</tr>
<tr>
<td>32</td>
<td>Save with Automatic Update</td>
</tr>
<tr>
<td>64</td>
<td>Plot with Automatic Update</td>
</tr>
<tr>
<td>128</td>
<td>Publish with Automatic Update</td>
</tr>
<tr>
<td>256</td>
<td>eTransmit/Archive with Automatic Update</td>
</tr>
</tbody>
</table>

Incorrect object filter or selection can only be avoided by carefully defining the data you wish to work on. In most cases this is fairly easy, however if Data Extraction is being used to take quantities from a drawing then it is up to the user to ensure that the right properties are selected to select all the objects the user wishes to query and none of the objects that shouldn’t be queried.

Understand the importance of reusing data

When we use Data Extraction we are reusing data from our drawings either in tables on drawings or in an external file such as a spread sheet. Reusing data is important because by doing so we reduce the chances of making a mistake. If we take the example of a line in a drawing representing a length of pipe, by linking the quantity in a spreadsheet directly to the graphical object we ensure that if the line changes then the spreadsheet will be updated as well. Reusing data in this way gives us confidence in the validity of our data, reducing mistakes and saving time.