

## Be Aware of the Data Skepticism and Knowledge Are Key

**Famous futurist Alvin Toffler once said “You can use all the quantitative data you can get, but you still have to distrust it and use your own intelligence and judgment.” The skepticism you exhibit when looking at your raw data will save a lot of time and legwork for the duration of any analytical project. Being more knowledgeable allows you to improve the system, the databases and the relationships from where the data comes. The steps in this article will provide a methodical structure to make the first move.**

In the analytical world, your data is vital — and the single most important piece in your project. It is actually more important than any tools, models, or methodologies you will use. Regardless of the scope of the project, from daily operations execution (such as carrier load planning or inventory management) to strategic planning (such as logistics network design or facility design), the correct data will ultimately be what makes your project a success.

With each project, there is a new challenge to collect the data in a correct and valid form to ensure the validity of the analysis. Here are a few things to consider.

### 1. Identifying the Data Source

Understanding the data source and underlying information systems are very crucial as you structure the data request from the client. Some data might be residing in one big

data source, or sometimes pieces of data might need to be extracted from multiple sources or, the same data will be received in different formats from different divisions of the same company.

Data source should not only be defined as just the “database” or the “information systems” that the data resides in. It is also the person who will provide you the data. You should have direct access to the source for open and fast communication. It is important to convey your request to every contact that gathers the data.

Unfortunately, when a data request is received, it may go through so many departments and people that it may become distorted because:



A) It might be a long time before it comes up to the top of their to-do list and you receive any data.

B) The data might not be what you need because there was a miscommunication somewhere along the way.

## 2. Structuring the Data Request

It is important to clearly define what data is needed. Sales data, item master, customer master, inventory on hand, and shipment files are some examples. The relationships between all data files required should be explained. A specific time period for pulling the data should be identified by the client and consultant together. A sample data file with explanations for all columns can be a valuable client reference.

Unit of measures of alphanumeric data should be consistent. If not, they should be explicitly defined and a conversion table should be provided. For example, in sales data, if some items are sold in “eaches” and some items are sold in “cases” or “packs”, they should be described in each line. In addition, if the same item is sometimes sold in “eaches” and sometimes in “cases,” only one unit of measure should be used, or they should both be identified.

If any part of the data is coded (such as channel or division names, product types, carrier or payment type), a reference legend should be provided by the client. It should be clearly noted whether the date is the date the task was performed — such as sales, receipt or shipment, or if the date the data was processed might be different than when the actual work was done.

There are many types of file formats to receive the data; e.g., CVS, txt, MS Excel, MS Access, DB files. Each file type has various technical capabilities and restrictions. For example, you can only have 65,536 lines of data in one spreadsheet in Microsoft Excel, but you can apply many formulas and functions for quicker analysis. Microsoft Access enables you to import much bigger quantities of data from various other file types and create relationships among data.

When you receive the data, make sure that the characters are complete, that alphanumeric data have enough detail, such as the decimal part, and that date formats are correct and consistent. Also, the data that looks like numbers but should be text should be exported as text. An example is an item or customer number that has “0” at the beginning. If the data is exported as numbers, these “0s” disappear and there might be conflicts.

## 3. Avoiding Inherent Data Errors

The information systems work well only when the users put in the correct data. If users do not search existing customers or items before entering them into the system and do not pay attention to typos, many duplicate entries or considerable amount of wrong data will exist in the database.

**Having the latest and the greatest information systems does not guarantee that you are working with the correct data. The information systems work well only when the users put in the correct data.**

When the data is received, an initial error check saves a project before it becomes an erroneous result in analysis. Simple checks for multiple entries, unit of measure consistency, and missing data from specific time periods are some of these technical errors.

## 4. Building Assumptions

In the case of incomplete data, where pieces of the required data are not readily available for analysis, you need to make assumptions. Many companies might lack item dimensions or cost or sales price information, but this can be overcome by using category or group data assumptions.

When future plans are not yet made, past trends can be used to generate future forecast assumptions in volume and mix. If aggregation is required, correct family or category types or location of items should be available. Aggregation should be delicately handled and match the objectives of the analysis.

## 5. Data Validation

Data validation is the last step before going forward with the analysis or study. The validation process helps client management relate to the data at a high level and check the sanity of the content, rather than focus on typos and multiple entries. Representation of the detailed data in annual or monthly sales summaries, or monthly costs by function, can be compared to financial statements. For example, top ten customers or top ten selling/moving items can be validated by marketing. Daily and monthly receiving and shipping or picking and packing activity can be validated by warehouse management.

This data validation process is also an eye-opener for the project managers. It might

DATA	POSSIBLE ISSUES
<b>Sales Data</b>	<ul style="list-style-type: none"> <li>• Typos in order entry causing duplicate customer and item</li> <li>• Duplicate order or line entries for backorders</li> <li>• Invalid or inconsistent unique order numbers (e.g., repeating order numbers because of limited character availability in the system)</li> </ul>
<b>Shipment Data</b>	<ul style="list-style-type: none"> <li>• Date is not actual shipment date, but data processing date</li> <li>• Ordered quantity and shipped quantity are not distinguished</li> <li>• Sales/shipping unit of measure is not identified</li> <li>• Sales order and line detail for a shipment are not available</li> </ul>
<b>Receiving Data</b>	<ul style="list-style-type: none"> <li>• Date is not the actual receiving date, but data processing date</li> <li>• Purchasing/receiving unit of measure is not identified</li> <li>• Purchase order and line detail for a receipt are not available</li> </ul>
<b>Item Master</b>	<ul style="list-style-type: none"> <li>• Duplicate SKU entries</li> <li>• Missing or wrong dimension information</li> <li>• Inconsistent or unidentified unit of measure</li> </ul>
<b>Custom Master</b>	<ul style="list-style-type: none"> <li>• Duplicate customer entries with different values</li> <li>• Invoice address and physical/shipping address are not distinguished</li> </ul>
<b>Inventory Snapshot</b>	<ul style="list-style-type: none"> <li>• Inconsistent or unidentified unit of measure for stocking</li> <li>• On-hand and reserve quantities are not distinguished</li> <li>• Specific inventory location data is not available</li> <li>• Cycle stock vs. safety stock information is not available</li> </ul>
<b>Financial Data</b>	<ul style="list-style-type: none"> <li>• Supply chain cost broken down by fixed and variable is not available</li> <li>• Cost broken down by task in a distribution center (receiving, picking, shipping etc.) or manufacturing facility (manufacturing vs. warehousing) is not available</li> </ul>

tell a different version of the same story they see in their daily operations. It also provides the analyst with a safety confirmation from management in advance, against any confrontation along the rest of the project.

**Burcu Agar is a supply chain consultant with enVista. For more information, contact enVista at 877-684-7700 or at [inforequest@envistacorp.com](mailto:inforequest@envistacorp.com).**