Executive Summary
Supply chain data can vary in quality from good to downright misleading. When these data are being used as the basis for business decisions, the quality of the resulting strategic directions can vary from "on-target" to "I'm not really sure." With government regulations such as Sarbanes Oxley, "I'm not sure" has legal implications.

Data in the supply chain is often a bewildering collection of instrument measurements, laboratory measurements, material transactions, inventories and events. These data are measured and reported separately but together provide a coordinated picture of the business transactions and material positions in the supply chain. Data validation generally requires developing coordinated sets of data. To do this the data must be physically or logically consolidated into a central repository. After collection, there are several tools, or applications, that have proven successful in validating data, including: balancing, catch tracking, measurement conciliation and data reconciliation. These applications can pinpoint errors and inconsistencies and make possible corrective actions that enhance overall quality and confidence in the raw information.
# Table of Contents

- How Wrong Can We Be? ................................................................. 4
- A Strategy to Improve Consistency .................................................. 4
- Data Quality Toolkit ..................................................................... 5
- Is It Worth Gambling? .................................................................... 6
# Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1:</td>
<td>Single Point of Access</td>
<td>4</td>
</tr>
<tr>
<td>Figure 2:</td>
<td>Plant Model</td>
<td>5</td>
</tr>
<tr>
<td>Figure 3:</td>
<td>Error Detection Summary</td>
<td>6</td>
</tr>
<tr>
<td>Figure 4:</td>
<td>Material and Energy Balance</td>
<td>6</td>
</tr>
</tbody>
</table>
How Wrong Can We Be?
Inaccuracies in data can be caused by problems with the raw measurement or more subtle inconsistencies in timing or interpretation. Obviously, poor meter calibration problem can cause a value to be in error, but an accurate value incorrectly placed in time can also cause errors. The following list of data issues provides the scope for any data quality program.

- Measurement accuracy - has to be controlled using auditable maintenance and calibration practices
- Missing measurements or transactions - are a significant cause of information quality degradation
- Inappropriate time stamps on values - these errors are often reporting problems, but also can also be cause by incorrect handling of time zones or daylight savings
- Blurry data standards - does the data value have the correct units of measure, is it a compensated value, an uncompensated value, etc.?
- Poorly defined data - are the data values instantaneous or are they already an average; If so what is the period of the average?
- Change management problems - has the value been changed, and by whom? Has data been compensated differently in the past such that it can no longer be used as a reliable comparison with future data?
- Outdated value - has the source actually updated the value or is it just displaying what it had last time?
- Improper source - is the measurement being measured in the location where it's expected to be measured?
- Measurements misinterpretation - the measurements may be correct but with the process operating in an alternative mode it may need to be interpreted differently.

A Strategy to Improve Consistency
Data don’t get better on their own. There needs to be a strategy. This strategy requires that data be characterized and accessible. In other words you know what they are and you can get at them. The first step in improving data quality is organization. Several tools are available for making these improvements, including: characterization, creating a central repository, conversion and aggregation.

On its own a single data point is hard to validate. But when analyzed as a component of a large data set it can be compared for consistency. However, for data values to be compared they must all be on the same basis. This means the same unit of measure, the same method of measurement, at the same reference conditions, applying to the same geographical scope and on the same timescale. Data from different sources in general will violate some if not all of these rules. The information needs to be logically characterized so that it can be processed appropriately for comparison.

To compare the monthly planned quantity of a specific feedstock in thousands of pounds that must be transferred to a unit, to the actual quantities of various feedstocks measured in gallons processed daily from various tanks, requires several levels of processing. Transactions at the tank level need to be aggregated to the unit, transactions at the individual feedstock level need to be aggregated to the feedstock types and daily transactions need to be summed for the month converting the units of measure. This can only be done if every data value is characterized.
Data must also be accessible from a single point of access (Figure 1). This doesn't imply that the data must reside in the same database but just that it must be accessible in a form that allows comparison queries to be executed. For the most effective use of the analysis tools, the data need to logically or if necessary physically reside in a single coordinated data structure. This coordination should be able to bring together real-time and transactional data into a single view.

Any attempt at data consistency will involve the use of conversion and aggregation. Trying to make disparate systems use the same units of measure, same naming, same time aggregation, same geographical aggregation and the same material aggregation is a losing battle (Figure 2). Even if you get it right at the beginning it won't last. This is where characterization is important. Knowing the exact characteristics of the data values enables conversion and aggregation routines to be used to bring consistency to the information.

Any data quality initiative must therefore have the capability for units of measure conversion, and aggregation over different time and intermediates, so that data reported at differing levels in the organization can be combined.

**Data Quality Toolkit**

With consistent information comes the ability to apply analysis methodologies, such as balancing, measurement conciliation, batch tracking and data reconciliation (Table 1).

Balancing is the application of simple volume, mass or energy balances to envelopes around sections of the facility (Figure 3). The objective is to identify areas where data mismatches occur. Typically, the analysis can be exception based or can balance by commodity or owner. Balancing will show errors across a plant area but cannot determine the exact location of the erroneous measurement. Further investigation and drilldown is needed to pinpoint the problem.

Conciliation is a process of mediating between two measurements of the same quantity based on predefined accuracies. The rules for this mediation can be as defined by a standards organization or by business rules and agreements. Rules can be as simple as, if meter A is different from meter B, use meter B’s value. Alternately, more complex adjustments can be made using a statistical comparison of the meter accuracies. The adjustments made on each side of the conciliation process can be tracked and analyzed to provide statistical information on the relative meter performance.

In one example, a reactor thermocouple measured against another thermocouple and a calculation of the adiabatic flame temperature. Any deviation was a sign of either mechanical failure of one of the thermocouple or poor installation. Mechanical failure meant an increase in resistance with a corresponding creeping rise in reported temperature.

Batch Tracking involves following batches of material along the supply chain from inventory locations through transportation systems to distribution. At each location compositions, physical qualities and quantities are calculated for comparison to measured qualities and quantities. This process can be applied to any supply chain configuration: continuous flow or with discrete transactions, with or without predefined batches. It requires no additional data input other than that normally captured to run the business: measurements, lab analyses, inventories and movements.

Benefits are significant. It provides predictions of qualities and quantities at any time and for any location in the supply chain. This enables continuous validation of material balances and laboratory analyses. It is important to understand the method used to measure a value; there should be an agreement between the methods and accuracy of measurements compared.
Data reconciliation is the application of statistical or heuristic methods to generate a complete and consistent mass balances across an entire process flow sheet. Every measurement of inventory and flow is adjusted, subject to applied constraints, to produce a completely balanced result. The adjustments to each specific measurement that had to be made to get a balance are compared to their accuracy tolerance. This relative measure shows which values were statistically in error and need to be examined. This is a powerful analysis tool because it pinpoints specific measurements in error, it provides estimates for unknown flows and it provides accuracy statistics on measurements and balances.

Although Data Reconciliation is a powerful tool it is not often taken advantage of to its fullest extent. This is because the data collection and clean up required to make it effective are not easily put in place. It is not a tool that can be implemented in isolation; it needs a data quality strategy behind it.

**Is it Worth Gambling?**

Would you bet your business on your data? Whether you like it or not you are doing that every day. Application of a coordinated data integration strategy with some data analysis tools will put you on a continuous improvement path toward high inherent raw data quality. Validated, audited, consistent information with a single version of the truth is the key foundation of successful business decisions.

### Table 1. Each method has its merits.

<table>
<thead>
<tr>
<th></th>
<th>Balancing</th>
<th>Tracking</th>
<th>Reconciliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area imbalance</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Node imbalance</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Specific measurement error</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Movement with no quantity</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Negative inventory</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Statistical measurement error</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Property discrepancies</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Ownership problems</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown flows</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Figure 3. A typical material and energy balance begins with a process flow diagram.