DATA BLENDING FOR BUSINESS INTELLIGENCE (BI)

10 Things You Need to Know about Data Preparation

Based on “Data Preparation: An Essential part of BI Visualization,”
Easyl White Paper by Rich Julius
The Analytics Landscape:

1. Businesses that engage in BI do better.

Bain’s *Big Data: The Organizational Challenge* report reveals that companies that use analytics are **TWICE AS LIKELY** to have top quartile financial performance than those that do not and **FIVE TIMES MORE LIKELY** to make decisions “much faster than the competition.”

It’s not a fly-by-night trend.

A.T. Kearney forecasts that worldwide spending on “Big Data” will **GROW AT A RATE OF 30% (CAGR)** from now until 2018, when the market will be $114 billion year.

Cloud-based Business Intelligence, a subset of this larger market, is expected to **GROW AT 31%**.

---

2. [www.atkearney.com/analytics/featured-article/-/asset_publisher/FNSUwH960jy7/content/beyond-big-the-analytically-powered-organization/10192](http://www.atkearney.com/analytics/featured-article/-/asset_publisher/FNSUwH960jy7/content/beyond-big-the-analytically-powered-organization/10192)
Even though the goal is to plug-and-play for BI visualization, you can’t just point your BI tool to different data sources and expect magic. That’s the myth that gets in the way of BI adoption – it doesn’t work without data prep. **So what do you do?** You have three choices for data preparation:

The reality is, if you give data preparation short shrift, everything that comes after it is a waste.

David Dietrich
InFocus, The Global Services Blog

### Choices For Data Preparation
[more about these later]:

- You engage IT to build an expensive, time-consuming data warehouse, which is 50% obsolete the day it goes live,
- You blend data manually with Excel, spending 80% of your time in manual prep and 20% of your time actually analyzing that data. Or,
- You go for an easy-to-use, self-serve data prep tool like Easyl. No more manual data prep, no investment in a monolithic data warehouse. Presto data.
Data preparation is the work required to give BI tools access to all the right data relevant to running a business, setting strategy, and making informed decisions.

Why You Need Data Prep:
BI Tools don’t blend data very well. They can do basic tasks, but if your data isn’t optimized for reporting, it’s like putting a whole pineapple in a blender: rind, eyes, core, leaves and all. No wonder you can’t use the results.

Data prep tools enable you to optimize your data using joins and maps that make it more palatable for consumption by BI tools, meaning more accurate analytics.

There are requirements for effective data preparation.
Data prep involves accessing data needed for BI from different locations.

You can pretend that your data is in just a few systems-of-record, like your CRM and ERP systems. Or you can admit that a big part of the BI puzzle comes from everything from Google Analytics to Excel files strewn across the desktop landscape.

You need a tool that can reach data wherever it’s hiding.
Each data source stores things differently. Really differently.

Your data comes in all shapes and sizes from relational data to the Google Analytics cube; from cloud API data like Salesforce, Marketo, and Eloqua to NoSQL, NewSQL; and spreadsheets. You can’t achieve meaningful data integration without data preparation.

Here’s another example. In this case, analyzing the Lifetime Value of a Customer is dependent on data from four different locations:

<table>
<thead>
<tr>
<th>RESPONSE PATTERNS TO DIGITAL MARKETING (Digital Marketing Database)</th>
<th>HISTORICAL PURCHASES (General Ledger Database)</th>
<th>ESTIMATES PENDING (CRM)</th>
<th>ORDERS IN PROGRESS (ERP)</th>
<th>CUSTOMER LIFETIME VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESPONSE PATTERNS TO DIGITAL MARKETING (Digital Marketing Database)</td>
<td>HISTORICAL PURCHASES (General Ledger Database)</td>
<td>ESTIMATES PENDING (CRM)</td>
<td>ORDERS IN PROGRESS (ERP)</td>
<td>CUSTOMER LIFETIME VALUE</td>
</tr>
</tbody>
</table>
Standardizing data to ensure consistency and quality is mandatory.

Data quality and normalization involves everything from validating addresses to blending data that has different formats, to ensuring records in one system are reliably matched to records in another source. That’s when the process gets complicated.

For example, one system uses US date format, while another uses European. State names are used in one system, while abbreviations are used in another. And is “Bill Smith” in your CRM system the same person as “William T. Smith” in your ERP system?

<table>
<thead>
<tr>
<th>ERP</th>
<th>CUS</th>
<th>CTRY</th>
<th>REV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acme, Ltd.</td>
<td>CUST</td>
<td>Great Britain</td>
<td>$ 64,228.90</td>
</tr>
<tr>
<td>Big Corp.</td>
<td>CTRY</td>
<td>United States</td>
<td>$ 354,254.00</td>
</tr>
<tr>
<td>Central Co.</td>
<td>CTRY</td>
<td>United Kingdom</td>
<td>£ 423,113.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CRM</th>
<th>Customer</th>
<th>Country</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acme</td>
<td>CUS</td>
<td>UK</td>
<td>£ 42,345.00</td>
</tr>
<tr>
<td>Big, Inc.</td>
<td>Country</td>
<td>USA</td>
<td>$ 354,254.00</td>
</tr>
<tr>
<td>Central Co.</td>
<td>England</td>
<td>UK</td>
<td>$ 643,132.00</td>
</tr>
</tbody>
</table>
7. Must-have benefits of data preparation.

If you’re not getting these benefits from your data prep process, it’s time to explore your options for faster, better BI support.

- Analytics are based on clean, digestible data
- BI visualization is more complete and accurate, preventing flawed analysis
- Repeatable processes speed up the BI cycle for fast answers
- Data wrangling is reduced up to 75%, thanks to data blending recipes and templates
- Knowledge workers enjoy self-serve access to data with automated data prep tools
- Data mining is easier and more flexible using an iterative approach for further analysis without always starting from scratch

Ultimately, the big benefit of data preparation is **BETTER RESULTS FASTER**.
There are the three different ways for IT and Business users to perform data preparation.
Manual data preparation is the traditional way.

Manual data prep is error-prone, tedious, and not very repeatable, i.e., done from scratch almost every time. It involves manually pulling data from a variety of sources and dumping the results into Excel. By the time you are done with all the VLookups and transformations, your Excel file is an unwieldy and poorly-documented data blending application.

In addition, it cannot be easily leveraged when you need to add new data sources, dimensions, or answer new questions.

Manual data prep is the reason why analysts spend 80% of their time preparing data, and only 20% of their time analyzing the results. By now, manual data prep smacks of the olden days.
Data Preparation is also done through large data warehouse projects.

At the other end of the spectrum are large-scale IT-driven data warehouses (or data marts). Software developers, IT professionals and database administrators collaborate to deliver a solution that puts the data through a series of extractions, transformations, filters, and corrections as specified by the particular BI use case.

IT data warehouse projects are expensive, time-consuming, and aren’t known for their flexibility. They are great at querying systems-of-record, but not so great at blending data from all the different data sources typically used by marketing departments. By the time the data rolls out, many of the questions have changed, and the solution has to be re-tooled.
The new generation of self-serve data preparation tools changes everything.

These self-serve tools make complex data blending easier for knowledge workers and analysts without a lot of IT hand-holding.

Moreover, the work is iterative, automated, and often templated, so analysts can easily shift gears and expand on their knowledge to answer new questions and accommodate new data sources as business evolves.

The new self-serve tools are more nimble, enabling analysts to experiment and do data blending in hours, not weeks or months, minus the need for monumental and expensive data warehouse initiatives.

An A.T. Kearney/Carnegie Mellon University study showed that organizations that do analytics well (“the leaders”) tend to foster a culture of cross-functional collaboration and co-creation of BI assets.  

The slog of data preparation is over. What was slow and inefficient, weighed down by days of Excel and email, is now fast and brilliant. For everyone. Business analyst, decision-maker, IT pro, data integrator -- Progress Easyl puts you ahead of the pack. High-performance data preparation is now in your hands.

Learn more

Download the white paper