



WHITE PAPER

Now You See It

How Integrated Data Wrangling Fuels Analytic Creativity

By Dan Woods, Chief Analyst, CITO Research



Dealing with data, by all accounts, consumes a huge chunk of analysts' time. Cleaning it, normalizing it, figuring out what to do about outliers and missing values, blending new data sources, stopping, thinking, asking IT for more data – on and on it goes. There's been a significant investment in tools to make this process easier, with companies like Trifacta, Paxata, and Alteryx among others involved. On top of this, data analysts and data scientists alike have data preparation tools of choice, often a combination of R scripts, Python programs, spreadsheets, and corporate extract, transform, and load (ETL) tools.

Right now the tools for data preparation and data visualization live on different islands of functionality and purpose. Analysts spend time on both islands, working with one tool to prepare data to bring it into a data visualization tool, then performing their analytical exploration. If, as so often happens, something about the data isn't right the first time, they have to stop everything and go back to data preparation island. Some tools provide a bridge between the islands, but the reality is, the analytical job still requires a constant traversal back and forth across the bridge to separate worlds.

Right now the tools for data preparation and data visualization live on different islands of functionality and purpose

When analysts are in the middle of discovery, stopping everything and going back to another tool is jarring. It breaks their flow. They have to come back and pick up later. Productivity plummets and creative energy crashes.



But what if there were a way to combine data preparation and analytics in one tool, to make the analytics experience into one immersive, integrated experience? What if it were possible to bring the two cultural islands of data and analytics into one place? How would it work? How could it improve the quality of insight discovery? Several vendors have tried addressing these questions. Tableau, Qlik and TIBCO Spotfire all use different forms of interactive visualizations that allow back-end data to be explored. But breaking away from the pack to integrate data wrangling tightly with analytics is TIBCO Spotfire.

So the question is, what does immersive, integrated data wrangling really do for you?

Changing Data Wrangling's Reputation: From Grunt Work to Artisanship

Whatever combination of tools is used, data preparation and wrangling is arguably the grunt work of the analytics process. Data analysts can spend days, weeks, or even months on projects just to extract, normalize and prepare data for analysis. And once they see the data, it's likely that they'll see something they missed, whether it's outliers, missing values, or simply a hankering to correlate the data they're seeing with another data source. Inevitably, they have to go back to the drawing board: back to IT, back to their ETL tool, or back to data preparation tools to work on the data again and then bring it back into the data visualization tool again. The iterative process of data preparation and data wrangling consumes about 80% of the time it takes to find analytical insights.

Data preparation and data wrangling consumes about 80% of the time it takes to find analytical insights

Integrated data wrangling is like painters with color palettes, who can adapt, mix, and create new colors at will, as they create the perfect array of colors to make their vision come to life. Like painters, analysts using immersive data wrangling tools can mix data at will, adapt it to their needs, and create new data palettes at will.

TIBCO Spotfire provides a single environment to eliminate context switching for analysts. By embedding data wrangling within Spotfire's visualization environment, data wrangling becomes organic and immersive, transforming what was fragmented, two-step grunt work and back and forth into a skillful, seamless, artistic process.





How Does Immersive Data Wrangling Work in Spotfire?

There are three main elements to immersive data wrangling in Spotfire.

- ① **Basic inline data manipulation.** For example, Spotfire lets analysts take the data they have loaded and work with it, like a color palette, pivoting, unpivoting, categorizing, grouping, splitting, and filling in missing values. Analysts can apply these techniques to a broad swath of enterprise data, thanks to the ability to connect to a wide array of data sources, both traditional and streaming.
- ② **Analysts can play and analyze at the same time.** For example, Spotfire suggests visualizations to try, based on its understanding of the data. If analysts spot data anomalies, they can continue to mix, fix, and refine the data, in place, right next to their visualization. Just as painters interact with their canvas, brush, and palette at the same time, analysts can interact with data, visualizations, and even R mathematical models, all at the same time.
- ③ **Spotfire keeps a visual, editable map of data wrangling steps** so analysts can see where they came from, where they've been, and even transformations they tried that didn't work as expected. And once they get the data wrangled just the way they like it, they can export the steps they took to make the process repeatable.

By pulling interactive data visualization and data wrangling into one environment, the vision for painless data wrangling is nearly complete for analysts.





Understanding the Value of Integrated Data Wrangling

Integrated data wrangling offers benefits for both the analyst and the business.

- **Benefits to the Analyst.** Integrated data wrangling makes your job easier and improves the quality of insight by enabling creative focus. By eliminating context switching between tools, you get less distracted and can work faster. You don't run out of time, get fed up, or lose your place and have to start over again.
- **Benefits to the Business.** Integrated data wrangling means finding and acting on insights more quickly. That can lead to improved profitability, higher productivity, and happier customers, because you more quickly understand your business.

By incorporating analytics and data wrangling in a single package, Spotfire offers a cost-effective environment. It streamlines the analytics process and enables creativity to flow by minimizing the effort required for data preparation and making analytics discovery productive and exciting.

So, now I'd like to ask you about your experience with data wrangling. What other techniques, besides the ones I mentioned, do you use? Can you picture yourself using a more integrated approach to data wrangling and analytics? What would that change for you?

For more information, see spotfire.tibco.com/data-wrangling ►

This paper was created by CITO Research and sponsored by TIBCO

Dan Woods

Chief Analyst and Founder, CITO Research

I do research to understand and explain how technology makes people more effective in achieving their goals. My strong belief is that we are at the threshold of a golden age of IT, in which the promise of gaining value from technology will be fulfilled. I have written or coauthored more than 20 books about business and technology, including *APIs: A Strategy Guide*. I write about data science, cloud computing, and IT management in articles, books, and on CITO Research, as well as in my column on Forbes.com.