

# ACCELERATING THE IMPACT OF BIG DATA INITIATIVES WITH DATA WRANGLING

Employing data wrangling techniques can help expedite advanced data analytics.



**ADAM WILSON**  
CEO, TRIFACTA

**T**O BEAT THE CLOCK on time-sensitive big data initiatives—security threat detection, disease suppression, emergency evacuation planning—today’s organizations are taking a hard look at their traditional processes to identify what could be made more efficient. The biggest culprit? Data wrangling, or the process of converting raw data into a format that is usable for analysis, which consumes up to 80 percent of any analysis process.

It’s not just the volume of data that makes data wrangling so time-consuming, but the variety, too. Today’s organizations must contend with a range of complex data—everything from weather data to call detail records to medical files—which, taken at face value, can be indecipherable to business-oriented analysts. In most cases, the work of wrangling this data into a digestible format gets shipped out to their tech-savvy counterparts, who are limited in number (not to mention tasked with other work) and can take weeks, even months, to deliver on the data wrangling requirements.

With the right tools, organizations can eschew traditional processes for a more agile solution, one that automates the bulk of this process and allows analysts to work directly with the raw data and quickly transform it into the required format for analysis. It has tremendous impact on the entire analysis process. Cutting out the (technical) middlemen allows organizations to scale wrangling operations across the entire organization, accomplishing more analysis in less time.

The Center for Disease Control knows the importance of data wrangling all too well. When faced with an HIV/AIDS outbreak in Austin, Indiana last year, officials sought out the source to quickly put a stop to the spread of the disease. Leveraging a combination of internal CDC data, historical, geospatial and clinical data, the team of analysts was ultimately able to determine that the root problem was intravenous drug use and

contaminated needles isolated to a particular strain of HIV.

Not only was the CDC able to more efficiently identify the cause of this outbreak, they were also able to be more effective. The analysts with the most context of the problem at hand were working directly with the data, instead of needing to delegate requirements about a dataset to others. It gave them the opportunity to iterate early on, reducing cycle time of these iterations, and undoubtedly giving them more confidence in their final results.

Beyond the CDC, data wrangling has proven its worth to a multitude of organizations with varying use cases from cybersecurity and fraud detection to natural disaster mitigation to predictive maintenance and repair.

When it comes to detecting fraud, for example, agencies are using advanced data wrangling to combine and standardize a diverse set of user, system, and application data that allow them to more quickly identify suspicious behavior. To address evacuation plans for natural disasters, agencies are leveraging real-time and historical weather data with geospatial data and third-party travel systems data to better predict the impact of an impending disaster. Finally, predictive maintenance use cases allow military agencies to combine sensor/log data from aircrafts and ships with historical maintenance records and flight log data to proactively address maintenance.

While seemingly simple, the time saved from more efficient data wrangling can mean all the difference in suppressing an outbreak, ceasing a security breach, or identifying and correcting a defective airplane part. Data wrangling has never been so readily available as it is today, empowering today’s organizations to unlock the potential of their data.

*Adam Wilson is CEO of Trifacta.*



## The #1 Data Wrangling Solution for Federal Agencies

Our software helps government agencies unlock the potential of their data with a new approach to data preparation and analysis. Whether you want to save time with an existing process or utilize new data sources for an analytics initiative, Trifacta helps your team do more with data of all shapes and sizes.

[Carahsoft.com/innovation/Trifacta-analytics](http://Carahsoft.com/innovation/Trifacta-analytics)

