**How Trustworthy Is Big Data?**

February 2, 2018

[Timo Elliott](https://www.brinknews.com/author/timo-elliott/)

**Vice President, Global Innovation Evangelist at SAP**

big data is mainstream part of modern business. But there’s less control and governance compared to traditional IT architectures

big data systems typically involve raw, unprocessed data in some form of a [data lake](https://en.wikipedia.org/wiki/Data_lake),

It brings some big benefits, notably greater flexibility and more scalable processing than traditional approaches. But there are also disadvantages, such as little or no initial controls over the quality of the information stored in the data lake.

So businesses for their decision-making can actually trust the data coming from these new systems?

key practical dimensions that organizations should consider when embarking on big data projects

1. **Data completeness and accuracy.**

The recorded values may be approximations rather than exact real values—uncertainty and imprecision might be an inherent part of the process, example – Sentiment Analysis

1. **Data credibility.**

Big data often comes from systems not directly controlled by an organization and may contain inherent biases or outright false values (for example because of [bots in social media](https://en.wikipedia.org/wiki/Social_bot)). And the schema-on-read approach means that you might not be able to ascertain the credibility of the data until after you’ve tried to use it for analysis.

1. **Data consistency.**
2. **Data processing and algorithms.**
3. **Data validity.**
4. **Define the ROI for big data quality.**
5. **Robust governance.**
6. **Transparency.**
7. **Training.**
8. **Ethics.**