every three years—that situation was swiftly inverted. Today, less than two percent of all stored information is nondigital.

Given this massive scale, it is tempting to understand big data solely in terms of size. But that would be misleading. Big data is also characterized by the ability to render into data many aspects of the world that have never been quantified before; call it "datafication." For example, location has been datafied, first with the invention of longitude and latitude, and more recently with gps satellite systems. Words are treated as data when computers mine centuries' worth of books. Even friendships and "likes" are datafied, via Facebook.

This kind of data is being put to incredible new uses with the as sistance of inexpensive computer memory, powerful processors, smart algorithms, clever software, and math that borrows from basic statis tics. Instead of trying to "teach" a computer how to do things, such as drive a car or translate between languages, which artificial-intelligence experts have tried unsuccessfully to do for decades, the new approach is to feed enough data into a computer so that it can infer the proba bility that, say, a traffic light is green and not red or that, in a certain context, lumière is a more appropriate substitute for "light" than léger.

Using great volumes of information in this way requires three profound changes in how we approach data. The first is to collect and use a lot of data rather than settle for small amounts or samples, as statisticians have done for well over a century. The second is to shed our preference for highly curated and pristine data and instead accept messiness: in an increasing number of situations, a bit of inaccuracy can be tolerated, because the benefits of using vastly more data of variable quality outweigh the costs of using smaller amounts of very exact data. Third, in many instances, we will need to give up our quest to discover the cause of things, in return for accepting correlations. With big data, instead of trying to understand precisely why an engine breaks down or why a drug's side effect disappears, researchers can instead collect and analyze massive quantities of information about such events and everything that is associated with them, looking for patterns that might help predict future occurrences. Big data helps answer what, not why, and often that's good enough.

The Internet has reshaped how humanity communicates. Big data is different: it marks a transformation in how society processes information. In time, big data might change our way of thinking about the world. As we tap ever more data to understand events and make