How Data Diets Can Improve Cyber Health

This article was originally published by the *Wall Street Journal's* CIO Journal on March 3, 2015.

by Brian E. Finch

Brian E. Finch
Public Policy
+1.202.663.8062
brian.finch@pillsburylaw.com

Brian E. Finch is a partner in Pillsbury's Public Policy practice in Washington, DC. He is a recognized authority on global security matters and is co-leader of the firm's Global Security practice.

The vast majority of data held by most companies has little or no business value—including drafts, duplicates, personal communications, even data or documents held over from spun-off companies or inherited from buildings they've purchased. At best, this type of data lacks productive value. At worst, this data can embarrass a company or lead to financial loss if it is ever exposed. In either case, retaining this data glut leaves companies maintaining an increased "cyber perimeter" and spending more and more time monitoring their larger infrastructure, as well as trying to use increasingly sophisticated tools to detect patterns that are indicative of the loss of information.

So, instead of allowing vast amounts of data to accumulate indiscriminately in the hope that there will be some slight business advantage gained from analyzing it, companies should think about the potential value of reducing its electronic storage profile. With lesser amounts of data to protect, the company will better be able to manage the security profile of truly valuable data, and monitor who has accessed it and why. For any companies who have gone through the exercise of performing traditional data classification efforts, especially of unstructured data, they have learned the hard way how trying to perform that work without getting rid of the enormous volume of "junk" data makes an already-challenging project feel impossible.

Similarly, reducing the amount of data stored decreases the likelihood of the exposure of information that is not necessarily high-value intellectual property or protected personal or health information, but merely embarrassing communications or carelessly-worded personal documents. As we have seen over the past few months, exposure of this kind of content can have impact equal to or even greater than the traditionally sensitive categories of data we're all used to protecting.

Put more simply, keeping several years' worth of email records should not be the default position of a company. Instead it should be looking to cull records (in compliance with applicable preservation obligations, regulations and statutes, of course) so that they don't wind up with the problem known as "zombie data." Here, zombie data refers to information that a company did not even realize it still had, but could be harmful if discovered and used by an outside party. Obviously the less harmful information that exists, the better off a company is.

Reducing the amount of data retained also has the benefit of stretching the cyber defense

dollar. With fewer critical bytes of information to protect, companies can concentrate their defenses and response capabilities around truly valuable information. They can also afford to utilize techniques to make it more difficult for attackers to gain value from any information seized—namely through encryption, false/misleading data, and other techniques designed to make it more likely that hackers will gain access to wrong or useless information.

Eliminating "junk" data also has the side effect, by the way, of benefitting Big Data programs. Regardless of the power of advanced analytics, it is undeniable that one can get to better analytics faster if one knows that the data under review is high-value and high-integrity, and when one knows something about the content in advance. Cutting away at known junk data therefore serves to increase the fidelity of all your data and position you better for reaping the benefits of advanced analytics.

Ultimately, it is up to companies to realize that they can only do so much to protect themselves if they have an ever increasing vat of data. Like an ancient empire overstretching its reach, collecting data and storing it based upon its possible use will likely only lead to trouble. Going on a data diet then, so as to focus on protecting what truly matters, can and should be an important layer of a company's cyber defense.