For the first few decades of its existence, the National Security Agency was a quiet department with one primary job: keeping an eye on the Soviet Union. Its enemy was well defined and monolithic. Its principal tools were phone taps, spy planes and hidden microphones.

After the attacks of September 11, all of that changed. The NSA’s chief enemy became a diffuse network of individual terrorists. Anyone in the world could be a legitimate target for spying. The nature of spying itself changed as new digital communication channels proliferated. The exponential growth of Internet-connected mobile devices was just beginning. The NSA’s old tools apparently no longer seemed sufficient.

In response, the agency adopted a new strategy: collect everything. As former NSA director Keith Alexander once put it, when you are looking for a needle in a haystack, you need the whole haystack. The NSA began collecting bulk phone call records from virtually every person in the U.S.; soon it was gathering data on bulk Internet traffic from virtually everyone outside of the U.S. Before long, the NSA was collecting an amount of data every two hours equivalent to the U.S. Census.

The natural place for the NSA to store this immense new haystack was the same place it had always stored intelligence assets: in the agency’s own secure facilities. Yet such concentration of data had consequences. The private, personal information of nearly all people worldwide was suddenly a keystroke away from any NSA analyst who cared to look. Data hoarding also made the NSA more vulnerable than ever to leaks.

A three-step plan for using data right in an age of government overreach

By Alex “Sandy” Pentland

Data about human behavior have always been essential for both government and industry to function. But how do we enable institutions to collect and analyze data without abusing that information? We can start by embracing some basic principles. The NSA and other government organizations should leave big data resources spread across functionally separate databases overseen by separate organizations. Everyone who holds or shares personal data, including citizens, must safeguard transmission and storage through encryption.

In the digital era, we must also realize that existing policy and tradition will not suffice. Constant, transparent experimentation with big data procedures is the only way to find out what works.
Information and privacy are two sides of the same coin. The Snowden leaks have demonstrated that the balance is tipped heavily in favor of the former. The NSA, for example, has been collecting data on basically everyone, including phone calls, emails, and internet activity. This data is stored and analyzed for various purposes, from national security to market research.

Despite the concerns raised by the leaks, the government has been slow to address them. The USA Freedom Act, passed in 2015, was a response to the controversy, but it failed to adequately address the privacy concerns. The new law requires the NSA to limit its collection of data, but it still allows for warrantless surveillance of Americans.

The debate over privacy vs. security continues, with advocates on both sides making strong arguments. Some argue that the government needs access to this data to prevent terrorist attacks, while others believe that this kind of mass surveillance is a violation of individual rights.

In the end, it seems that we are going to have to make a choice. We can either prioritize security and allow the government to collect and analyze as much data as possible, or we can prioritize privacy and limit what is collected and how it is used. It is a tough decision, and one that we must make carefully.