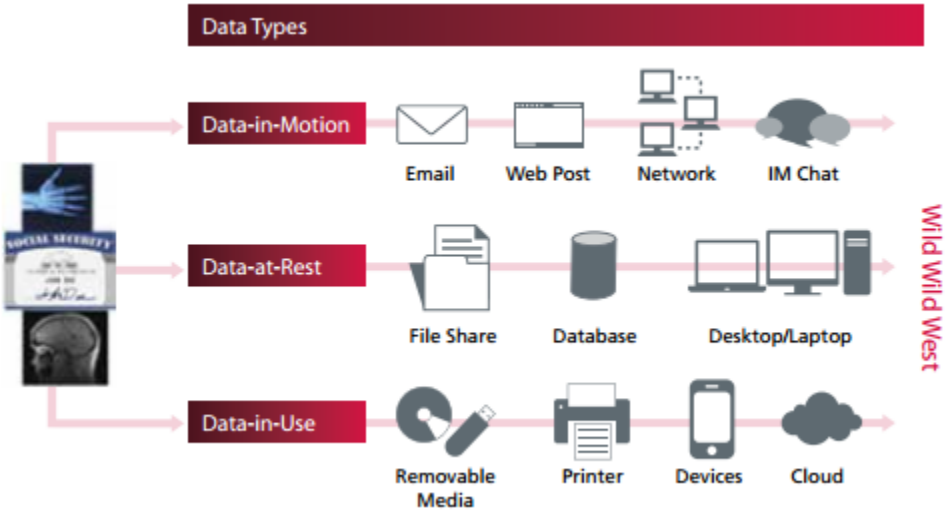
Data Protection: The center of everything we do.

Nate Dell, Manager, IT Infrastructure and Security 6/14/16

Defining Data Protection strategies is critical in today's digital and data driven business. Data protection strategy can be broken down into three distinct groups or functions. The first is tagging or labeling data based on specific data elements. Every organization may place different amounts of value on different data elements. For example, a finance institute may have a high value on bank account numbers or credit cards whereas a healthcare organization may care about member ID numbers or patient information. What is important is that data elements need to be simple and discoverable in both the unstructured and structured world.

Tagging data is not a trivial task especially in the unstructured world. For example, let's take a look at everyone's favorite nine digit number, the Social Security Number (SSN). Simply creating a regular expression that searches all ASCII file types for a 9 digit number will result in a large amount of false positives. This rule must be coupled by identifying multiple key common words, (SSN, Social, Social Security, Security Number, Membership ID number, etc.) within proximity, 4 excel columns or the first 1/3 of the page of a word document of that 9 digit number. Tagging this information in the structured format is inherently less complicated because the data is structured and typically labeled within a database. By first identifying a key word list for each data element one can simply use that list to search database column names. Note data tagging is a "set it and forget it" type activity; the technology used to identify this information will need constant minor adjustments as people develop new business document templates or bring new databases online.

Once the data has been tagged and labeled across the organization, a huge feat within itself, it must be classified into buckets. Note that some data elements like SSN can be classified in multiple categories, regulated and confidential, and by default should follow the control with the most rigors. This classification scheme will drive how technology is deployed. Data is protected in following three distinct states.



(McAfee DLP for Health Care Public Document)

Once tagged and classified, the data must now create processes and technical controls on how the data will be protected. For example, let's follow a document that can be found in most health care organizations. A word document tagged with data elements SSN, Member ID, First/Last Name, Address, and medical claims information. Now think about the life cycle of this document and how it must be protected in all three states.

- Data at Rest – Any document that contains these data elements must be housed on an encrypted storage medium. This means regardless of where this document is stored (Network Fire Share, laptops/Desktops, Database, USB Thumb Drives, etc.) all drives must be encrypted.
- Data In Motion – Any time this document is transmitted it must be encrypted and can only be done by authorized personnel. This would mean before the document is transmitted to a new location authorization is needed. Once the first hurdle is passed, restrictions to approve encrypted communication channels to transfer this file (Encrypted Emails, IM, Web Uploads, Secure File Transport Protocols, etc) are required.
- Data in Use – Only authorized personnel can read this document based on business reason and only on authorized systems. This means that rigorous access controls and monitoring needs to be set up so at any point the question of who/when/why can be answered.

This data protection strategy is a powerful framework and although it sounds simple, is one of the most complex strategies to institutionalize. Data protection impacts everyone and everything.