NAMFS Technology Committee Reviews NPPI/PII in our Industry

In today's litigious and overregulated mortgage industry, one of the most overlooked liabilities we see today is Personal Identifiable Information (PII). PII violations are much more common in other industries, with losses in the millions due to data breach and/or loss but are moving into our industry. Examples of such violations that occur in our industry at the vendor level include printing work orders with NPPI/PII and leaving it where it may be found by outside parties, as well as data stored and passed between unencrypted systems. These violations are an emerging risk and vulnerability to your business. NAMFS' Believes our industry must be aware of the potential fines/penalties a vendor at any level could encounter if a violation of this new area on a PER RECORD basis and the concern around the evolving myriad of state laws that govern this issue without any Federal Oversight.

What is NPPI/PII?
US Department of Labor's definition of PII:

Personal Identifiable Information (PII) is defined as:
Any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means. Further, PII is defined as information: (i) that directly identifies an individual (e.g., name, address, social security number or other identifying number or code, telephone number, email address, etc.) or (ii) by which an agency intends to identify specific individuals in conjunction with other data elements, i.e., indirect identification. (These data elements may include a combination of gender, race, birth date, geographic indicator, and other descriptors). Additionally, information permitting the physical or online contacting of a specific individual is the same as personally identifiable information. This information can be maintained in either paper, electronic or other media.

Today, personally identifiable information (PII) is one of the most discussed, yet confusing topics in our industry. How does an organization that relies on mortgagor information tied to home addresses and loans perform their work without violating consumer rights?

Data breaches are on the rise. The emergence and acceptance of cloud computing has proven to be as safe, if not more safe than on-site data centers. Convenience and ease of access allow cloud tech to flourish, but also create new challenges in protecting data (especially PII).

How does an industry working within all fifty states, Guam & Puerto Rico address a hotly contested issue with no blanketed federal mandate? There is no federal standard, or standard practice for safeguarding PII, however, considering some of the recommendations below could save your organization from unintentionally violating consumer's rights.

1. Role-based administrative access – unauthorized access
    - Restricting the ability to access PII data by users who should not or do not require access

2. Obfuscation of NPPI/PII data

Can be reversible, though many argue that for our industry's data sets, permanent obfuscation of data may not be possible. In some cases, data will need to maintain its referential integrity.

Consider both *test* and *production* environments in the application of PII policies. Consider leveraging test data that is like the real thing being used in production environments. Consider PII policy for Internal Testing & UAT

Consider what the data privacy regulations, policies, and standards that your organization must comply with. Consider the states that you have contracts in and understand their unique legislation(s) and guidelines.

Obfuscation of data areas of consideration:

- Imports/exports
- SQL queries
- Get calls/requests
- Forms data
- Data logs
- Both current data sets and previous datasets and events

3. Data Encryption – Cryptographic method rendering data as useless until decrypted. Algorithmically transforming data into an unreadable format and reversing it with a needed encryption key.

4. Data Tokenization – Originating in the financial sector for payment related obfuscation. A token would need to leverage a one-time generated key to be created and the same goes for a use-once random private key to decrypt.

5. Data Masking – Obfuscation involving replacing displaying sensitive data with characters representing said data.

6. Data Anonymization – Data sanitation process where information is obfuscated for those requesting to view it. A stakeholder within a given process would want to view analytics about a certain process without needing to know a name, address, etc. Data anonymization would return limited information based on the query, or in some cases, return nothing at all.

7. Nulling – Replacing a data field with a null value.

8. Repeatable masking – Data masked with random values will represent those exact random values wherever that data is represented.

Other considerations

1. Identify what is sensitive
    - Define it (public, sensitive/critical, classified)
    - Define the risk associated with (PII Defined) – how much risk will your org be exposed to should compliance be violated?

    Deployment with rules surrounding administrative access and roles-based permissions to limit exposure.

2. Where the data is stored:
    - IE: Sage stores actual credit data outside of our system. Data is obfuscated so that customer payment info is hidden from our org.

Glossary of Terms:

PII - *Personally Identifiable Information* is any data that can be reduced or allow someone to deduce an individual by a single piece of data or correlating multiple fields of data.

CCPA - Is the *California Consumer Privacy Act* that aims to enhance privacy rights and consumer protections regarding companies' collection and use of PII and other data as of Jan. 1, 2020.

Gramm Leach Bliley Act – Set standards/expectations for the Financial Services Industry to follow.

GDPR - Stands for *General Data Protection Regulation* and is a European Union data protection privacy regulation in EU law. While not the first of its kind, it is one of the most expansive, and is required to be followed by anyone doing business in EU countries as of May 25, 2018. It has become a basis that a lot of state specific privacy laws have been based upon.

Data encryption - Is generally reversible encryption and can be applied in extremely granular cases such as encrypting a simple field, to encrypting an entire file, database, disk volume, or hard disk. The reversible nature of encryption means that we can retrieve what we encrypt later, often utilizing a key and algorithm to encrypt and decrypt the data.

Data tokenization - Can differ from data encryption in that it does not necessarily utilize an algorithm but is reversible. An example of this would be a list of characters from AAAAAA to ZZZZZZ. If we define ABCDEF as NAMFS, when people list data fields containing NAMFS they will see its assigned token, ABCDEF unless we have access to look up its original value in the index or mapping table.

All the above approaches create some technical constraints for developing applications. Based on data altered with them, functions like searching will be slower or impossible as the search string will need to be run through a similar function to do an exact match search, or the obfuscated data will need to be reversed prior. Wildcard searching is not viable with obfuscated data as only exact matches can be found. To use the NAMFS example above, if we had a database of every professional organization in America and searched for NAMFS, our code could pre-tokenize the search term to ABCDEF, but this will only match that entry, as similar entries like NAMFSUSA would only match a different token. We mention this to point out that adding these data protection approaches requires adding additional steps to common processes and usually there is extended development time as well as slower querying.

Data masking - A form of obfuscation that can at times be reversed, but if performed correctly should be irreversible. The simplest example of this is masking John Smith as J*** S****. As it is possible to guess this, data masking has evolved to mask data multiple times and add characters to words to make guessing less possible. A modern masked version of John Smith could look something like 7d2h323hj 913hasdhDF.

Data anonymization - Looking within data for anything that could identify an individual or group of individuals and requiring a *band* so that someone searching the data set cannot identify someone. A healthcare example of this would be looking at the healthcare data for a company that has one female employee and searching the dataset for employees that are pregnant. Anonymizing tools would return no result as one is too identifiable. Bands work similarly, in the same company if we search for employees that are sixty-five or older, if we only get one result back, we will return no result as returning anything could identify that person. This is more common in healthcare and research where we are generating statistics based on unique characteristics. Anonymization can also just be obfuscating vital fields such as home addresses.

Nulling - Replacing a field with null characters, or in the case of something like a credit report, leaving the score but not showing the outstanding balance of each account.

Obfuscation - Any altering of data to make it unreadable or unidentifiable. This is typically applied on a per field basis, and can utilize encryption, hashing, tokenization, scrambling or anonymization.

Repeatable masking - Repeatable masking we have lightly touched on above, if we are masking John Smith as 7d2h323hj 913hasdhDF, repeatable masking would mean anytime John Smith existed in the original data, 7d2h323hj 913hasdhDF will be in its place. Non-repeatable would mean this is not the case. By having it be repeatable, we can still do some calculations such as count how many times John Smith is in the database by counting how many times, we find the masked version 7d2h323hj 913hasdhDF.
CIS (Center for Internet Security) - Offers hardening templates and a cyber security framework as well but is more consolidated and an easier framework to implement for companies with more technical teams. CIS has under one hundred controls for its lowest level.

National Institute of Standards and Technology (NIST) – NIST exists to establish and recommend standards for several things to both federal agencies and the US. This includes how to establish an information security program, what FIPS level encryption is and how to comply with it, NIST 800-53 is security controls for private companies, 800-171 is for public entities such as municipalities. There are thousands of recommendations, and they also go a step further and offer services such as the National Vulnerability Database and security hardening baselines for operating systems and software. 800-53 Rev 5 is the latest (as of 2020) version of NIST 800-53 and has been updated with new controls guidance for things like supply chain management and data protection and privacy (particularly PII). NIST also has the NIST Cyber Security Framework that serves as the basis for other Frameworks in this space. NIST 800-53 and 800-171 have thousands of controls, with at least four hundred required for its lowest level.

Contributors

- Paul Palmer, Pruvan, Inc. paul.palmer@pruvan.com,
- Bob Whelan, Verisk robert.whelan@verisk.com,
- Eric Miller NAMFS eric.miller@namfs.org