The healthcare industry continues to struggle with two important goals that many view as being at odds with each other:

1) ensuring patient privacy
2) using aggregated patient data to improve healthcare through analytics

Industry stakeholders, both Covered Entities (CE’s) and Business Associates (BA’s), are well aware of the Health Insurance Portability and Accountability Act (HIPAA). HIPAA sets forth requirements for the handling and protecting of Protected Health Information (PHI). The Office of Inspector General (OIG) is responsible for assessing organizational HIPAA compliance and handing out corrective actions and/or penalties for non-compliance.

To date, the OIG has investigated non-compliance in purely a reactive mode, i.e., responding to complaints or news reports of data breaches or instances of non-compliance. OIG has indicated\(^1\) that starting in early 2016, they will be implementing a proactive audit program.

\(^1\) [http://oig.hhs.gov/oei/reports/oei-09-10-00510.pdf](http://oig.hhs.gov/oei/reports/oei-09-10-00510.pdf)
How many providers could pass an OIG audit?

OIG’s data shows that 85% of investigations result in findings of non-compliance. The areas of non-compliance include, among others, inadequate training, inadequate security, PHI accessible to employees who don’t need it, PHI breaches or release, PHI sent inadvertently to erroneous party, Business Associate breach or non-compliance.

With over 25,000 privacy cases investigated by OIG since 2009, resulting in over $35 million in fines assessed and an estimated costs of a breach of over $2 million for each healthcare entity, it is no wonder that providers are grappling with how to best ensure internal HIPAA compliance and are very hesitant to fully embrace the use of their clinical and administrative data externally for aggregated analytic use.

Capturing more value from data

Aggregated healthcare data is providing great value to providers, payers and life sciences and has been doing so for some time. However, to date, most of that aggregated data has been administrative (i.e., claims data) and has been limited to relatively basic analytical applications. Leading edge providers have been demonstrating remarkable additional care and operational insights through the use of clinical data warehouses, but these efforts have remained largely localized due mainly to uncertainty regarding existing data handling technologies’ ability to provide a regulatory compliant level of security. Because Protected Health Information (PHI) must be kept secure when aggregating, analyzing, or sharing healthcare transaction data, there are a number of logistical and technical challenges that need to be overcome to truly unlock the great potential of population health analytics on representative clinical and administrative data sets.

- Healthcare is delivered and managed across multiple Health Insurance Portability and Accountability Act (HIPAA) covered entities. Patients receive care from a number of clinicians with diverse specialties and at different facilities over their lifetime. These entities lack the ability to securely analyze the collected healthcare transaction data across the patient population they serve because of concerns about inadvertent release of PHI to internal personnel and consultants who have no clinical or administrative support reason to be exposed to PHI. These limitations make critical business and clinical planning difficult if not impossible.

- While multiple national and local Health Information Exchange (HIE) efforts are underway to help the industry exchange critical patient-specific data in support of clinical tasks, the healthcare industry is hampered when it comes to aggregating healthcare transaction records across the continuum of care in a HIPAA-compliant manner for analytical purposes. The main reason is the need to maintain PHI private even while linking the records from multiple sources into one database. This limitation inhibits the ability to gain real insights into the full healthcare lifecycle,

2 2014 Ponemon Institute breach report
negatively impacting the industry’s ability to learn from current practices, innovate and compete effectively.

• In order to use the source data as an effective tool in managing and improving healthcare relationships between services, time frames of treatments need to be maintained. Should regulations, or interpretations of regulations, move too far, the analytical value of the data could be greatly reduced.

• Business Associate Agreements (BAA’s) are broadly implemented to ‘protect’ the source of data and ensure that safe guards are in place at all levels. There are several issues with this assumption:
  
  a. Regulations are constantly changing so existing BAA’s require constant review to ensure proper implementation
  
  b. Current requirements and definitions are inconsistent and vague which can lead to misuse of the BAA concept
  
  c. The regulatory requirements are not in synch with state regulations and other privacy acts.
  
  d. IN THE EVENT OF A DATA BREACH AT ANY LEVEL, THE EXPOSURE OF THE ORIGINAL DATA SOURCE CREATES BOTH PUBLIC AND LEGAL ISSUES

To meet the challenges outlined above, the healthcare industry needs a means by which patients’ PHI can be secured while still allowing the aggregation of transaction data within and across multiple healthcare stakeholder organizations.

The Universal Patient Key (UPK) Solution

UPK provides software and services for healthcare industry stakeholders requiring a certified HIPAA-compliant system that can de-identify a patient to secure PHI even while maintaining the organization’s ability to aggregate the healthcare transaction data for analytical purposes without compromising patient confidentiality.

• The UPK offering enables the building and use of HIPAA-compliant healthcare data warehouses from disparate sources for internal analytics (limiting PHI to only those who “need to know”).

• UPK protects sensitive PHI housed within existing and new “Big Data” analytics systems, enabling the external linking and use of data. This unlocks the value of healthcare data for downstream analytics and monetization.

UPK’s, turn-key solution is the most cost-effective solution in the market and it has been designed specifically with the flexibility required to meet the needs of organizations across the healthcare continuum.
The UPK solution protects and honors the distinct Intellectual Property (IP) and ownership rights for each organization’s data records.

- Each organization is able to tailor the level of access allowed to healthcare transaction data inside their organization.
- Should the organization decide to share its data with (or sell it to) another organization, the organization can easily manage the parameters of that relationship, setting the time and access limits for the data sharing.
- UPK never has access to any organization’s healthcare transaction data. All UPK software is installed at either the customer’s facility or at a customer-authorized facility.

How the UPK offering works

UPK has developed a unique and proprietary double-encryption, site-specific method to protect patient information (patent pending).

1. UPK creates a unique encryption key for each participating organization.
2. The organization-specific encryption keys are used to generate encrypted patient-specific hash keys that allow for aggregation across diverse data sets within an organization or when shared across multiple organizations.
3. All hash seeds and keys are encrypted and secured inside compiled libraries that are accessed during system processing as required. Optionally the seeds and keys are stored in a secrets management system that ensures that only UPK can ever access the keys.
4. The generated patient-specific keys generated are dynamic values that are not decipherable by end-users, either internal or external to an organization.

All UPK software and supporting processes are designed to provide the highest level of protection for our healthcare customers’ information.
UPK facilitates data exchange within and across all healthcare stakeholders

Healthcare Providers – Clinicians and facilities providing direct patient care.

- UPK offers healthcare providers the greatest degree of HIPAA compliance when managing their healthcare data beyond direct patient care using employees or Business Associates who do not have a legitimate or limited need-to-know PHI.

Healthcare Payers – Organizations that pay for healthcare services.

- UPK offers healthcare payers the simplest way to merge and analyze healthcare data while achieving the highest degree of HIPAA compliance.

Healthcare Service Providers – HIPAA-covered entities or their Business Associates (entities that have signed a Business Associate Agreement – BAA) who support healthcare delivery (patient care, payment, and related services).

- UPK offers healthcare service providers the safest way to merge and share (or sell) healthcare data to multiple sources while maintaining the highest degree of HIPAA compliance.
- UPK’s unique template management system allows for identical source reports to deliver data sets with varying degrees of content based on the actual needs of the Business Associate.

Pharmaceutical and Medical Device Manufacturers, Data Aggregators and Analytics Firms – Firms needing to analyze healthcare transaction data but are not equipped to manage the responsibility required of a HIPAA Covered Entity or a Business Associate.

- UPK offers the broader healthcare ecosystem the ability to discover greater insights from healthcare transaction data aggregated from multiple sources in a HIPAA-compliant manner.
Universal Patient Key Technical Specifications

Data Management & De-Identification

The UPK process begins by working with the data source provider to define the format of the input data. This is accomplished by either the use of standard formatting instructions (e.g., 837) or by joint design efforts with the data source’s technical resources (e.g., pipe delimited database extracts).

Once the input mapping is complete, the next step is to define the modifications and scrubbing rules. At this point UPK walks through following steps:

1. Identify the PHI data elements and their positions
2. Identify any obtuse PHI elements (e.g., the location of a service address if that service is defined as “home service”)
3. Define business and contractual rules that are to be enforced by the process
4. Define the output formats

UPK has proven and proprietary scrubbing rules in place for names, dates, gender, patient, and subscriber zips as well as other values that are implemented. These rules ensure the tokens generated are consistent across all data sources. Furthermore, UPK is Safe Harbor compliant. Table 1 describes how the UPK process handles each Safe Harbor designated PHI element.

Once all of the compliance values are defined, both for regulatory and business rules, a template is developed that is used to manage the processing stream. The information in the template is used to process the input file as well as to define the processing variables used in the generation of concatenated values that are fed to the core hashing and encryption process that create the site-level token representation of the values. The template then defines the format of the output and the placement of the token values on the data.

Every template uses the same base combination of name, date of birth, and gender to create a pair of core token values indicated as “key1” and “key2” in the template layouts. Additional key values are generated when needed using unique values from the data stream that assist in linkage of data where full name, date of birth, or gender are not present consistently within the input file.

All keys are encrypted using the site keys defined in the company definition in the UPK Vault, which are compiled into the libraries when the production code is generated.

Once the token creation is completed, the output data is written to the file in the format defined by the template. The tokens written at the end of the de-identification process are the site-encrypted versions of the core universal tokens created during the process. The core universal tokens/keys are never presented in any output or log stream.
Table 1: How UPK handles Protected Health Information (PHI)

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Names</td>
<td>While elements of patient and/or subscriber names are used as values to create parts of concatenated strings of values that are converted to tokens, there is never a direct usage of a name value to generate a key</td>
</tr>
<tr>
<td>Zip Code</td>
<td>All patient and subscriber zip codes are reduced to the initial three characters. Based on data source rules all three digit zip codes with a combined population of 20,000 or less are either replaced with “)))” or are combined again with additional zip zones to ensure that populations exceed this minimum</td>
</tr>
<tr>
<td>Dates</td>
<td>Dates of birth, admission dates, discharge dates, and dates of death are converted using the standard template. The standard template converts all dates to January 1st of the birth year (the data source compliance officer may alter this to month versus year). This is template managed in the data relationship. All dates of birth where the individual would be 89 years of age or greater as of the date of the de-identification would be modified to reflect an age of 89. (Note: Actual dates of birth are used in token generation, but this only reflects the output values in the data set)</td>
</tr>
<tr>
<td>Medical Records Numbers</td>
<td>Removed where present. However, if accepted by source data compliance, medical record numbers, in part or in whole, may be modified and combined with other values as part of a concatenated value to create a record matching key</td>
</tr>
<tr>
<td>Telephone Numbers</td>
<td>Removed where present</td>
</tr>
<tr>
<td>Email</td>
<td>Removed where present</td>
</tr>
<tr>
<td>Address</td>
<td>Removed where present</td>
</tr>
<tr>
<td>Social Security Numbers</td>
<td>Removed where present</td>
</tr>
<tr>
<td>Beneficiary Numbers</td>
<td>Treated the same as Medical Records Numbers</td>
</tr>
<tr>
<td>Vehicle Information</td>
<td>Removed where present</td>
</tr>
<tr>
<td>Device Identifiers and Serial Numbers</td>
<td>Removed where present</td>
</tr>
<tr>
<td>URL’s</td>
<td>Removed where present</td>
</tr>
<tr>
<td>IP’s</td>
<td>Removed where present</td>
</tr>
<tr>
<td>Biometric Values</td>
<td>Removed where present</td>
</tr>
<tr>
<td>Image Fields</td>
<td>Removed as defined by the data source</td>
</tr>
</tbody>
</table>
Token Management & Secure Data Exchange

UPK has developed a proprietary methodology that protects against the intentional or unintentional release of key values and their PHI. If a public mapping of the tokens to PHI were to ever become available, then the tokens could become PHI. Therefore, the UPK process is designed to protect the token values it creates.

Each data source and data target organization is assigned a site level translation key which is managed by the UPK Vault. When production code is released to a data source, the site level key values are obfuscated within the libraries delivered. These values are assembled during the process and used to create the two levels of keys generated. The Universal Key is created only in process and is immediately translated by encryption to a site level representation of the value. This site level key is the value placed in the output file.

The second program delivered by the UPK Vault is a translation process which will process the data source output file and convert the keys into the key values used by the data target. This process can be executed at either the source or target site. By processing the translation at the source, UPK ensures that the source keys are never shared at any other partner site.

Using the translation process, accomplishes two goals:

1. The target data organization has its own unique key structure which can come from multiple sources using the UPK process. Additionally, if using the UPK process internally, it will match the keys generated by the local process
2. The data source can provide the same data set to multiple data targets, each having their own key value sets, all of which can be related to the data source and other UPK data sources. This is accomplished by processing the file through the translation process designating the data target organization.

The primary feature of this process is that the universal key is never exposed. The site keys are unrelated to the data at the site and therefore, if site key token values were to become exposed, UPK can provide new site keys to any of its clients along with new translation programs. These translation changes can even be applied to existing data sets, completing eliminating the exposure created by the loss of the original site token values.

In addition to increasing the level of protection against PHI violations, the site level token system allows the data source to control the usage of its data by other organizations. Data sources control the access to the site keys through interface management in the UPK Vault. In order to merge multiple data sources the organization must have authorization in place to translate the keys to a common site level in their data operations. This authorization is directly managed by the data source.
The UPK Vault

The UPK Vault is a web-based service hosted in an Amazon AWS HIPAA-compliant cloud. It manages all implementation of the data relationships between data sources and targets.

All licensed companies are entered into the vault as they enter into a contract agreements with UPK. A single-user ID is initially created which allows for the creation of users who can then manage the interfaces for each company and only data source administrators can create interfaces.

Prior to creating an interface, the data target must be entered into the UPK Vault and the template to be used must be moved into a production state.

The interface is created by logging in, selecting a target company, selecting a template for the interface and indicating the start and end dates of the initial agreement. UPK developers will receive an automated notification of the new interface and generate the appropriate software for de-identification and token translation. The translation and de-identification code includes the effective date of the agreement.
## APPENDIX: DEFINITIONS

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer/Company</td>
<td>An organization that has a master or site level license agreement with UPK to implement the de-identification and translation software</td>
</tr>
<tr>
<td>Data Provider</td>
<td>An organization that has PHI populated data it wishes to share internal or externally</td>
</tr>
<tr>
<td>Data Target</td>
<td>An organization, internal or external, that will receive de-identified data from a data provider or multiple data providers</td>
</tr>
<tr>
<td>Data Provider &amp; Target</td>
<td>An organization that both provides data to external sources and also generates de-identified data for internal usage</td>
</tr>
<tr>
<td>Company Seed</td>
<td>The value that is appended to source to data prior to being processed by the UPK engine</td>
</tr>
<tr>
<td>Company Site Key(s)</td>
<td>The values used to convert the Universal token generated by hashing the seeded values to the site specific token value</td>
</tr>
<tr>
<td>Template</td>
<td>The set of directions which manage the de-identification process. The Template contains input and output file descriptions as well as key generation rules</td>
</tr>
<tr>
<td>Interface</td>
<td>The entire process definition of defining the source of data, the target, the template and the effective dates of the agreements to process the interface</td>
</tr>
</tbody>
</table>
APPENDIX: SAFE HARBOR REGULATIONS

(2) (i) The following identifiers of the individual or of relatives, employers, or household members of the individual, are removed:

(A) Names

(B) All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census:

(1) The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and

(2) The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000

(C) All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older

(D) Telephone numbers

(E) Fax numbers

(F) Email addresses

(G) Social security numbers

(H) Medical record numbers

(I) Health plan beneficiary numbers

(J) Account numbers

(K) Certificate/license numbers

(L) Vehicle identifiers and serial numbers, including license plate numbers

(M) Device identifiers and serial numbers

(N) Web Universal Resource Locators (URLs)

(O) Internet Protocol (IP) addresses

(P) Biometric identifiers, including finger and voice prints

(Q) Full-face photographs and any comparable images

(R) Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section [Paragraph (c) is presented below in the section “Re-identification”]; and

(ii) The covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.
Contact us to learn more about how UPK can help you mitigate your HIPAA compliance risk and unlock the value of your data.

www.universalpatientkey.com | info@universalpatientkey.com | joe.austin@universalpatientkey.com | (978) 563-1798