Anonimización corporativa, criptografía más allá de los datos personales

Wilson Delgado, Nicolas Piñeros, Juan Monroy.

Resumen - El trabajo busca presentar el análisis obtenido del proceso de comparación y aplicación de las diferentes metodologías de anonimización criptográfica para el sector farmacéutico, en especial el vinculado a estudios COVID-19. Para determinar cuál es la técnica más viable, escalable y segura en el manejo y protección de datos no personales como lo son archivos de investigación, fórmulas u otros datos de carácter sensible en el área de investigación y desarrollo.

Abstract– The work seeks to present the analysis obtained from the process of comparison and application of the different cryptographic anonymization methodologies for the pharmaceutical sector, especially that linked to COVID-19 studies. To determine which is the most viable, scalable and secure technique in the management and protection of non-personal data such as research files, formulas or other sensitive data in the area of research and development.

Palabras clave: Información, Privacidad, Criptografía, Anonimización.

I. INTRODUCCIÓN

La protección de la información no ha sido un asunto que nos preocupe desde hace unas décadas, ha sido un elemento clave que nos concierne desde tiempos inmemoriales. De los primeros sistemas de cifrado que se tiene registro es el de Julio Cesar, sin embargo, estos sistemas se ven vulnerados frente a los avances tecnológicos los cuales crecen de manera exponencial pasando de la computadora de Turing a máquinas de cómputo potentes capaces de procesar más información.

Tal avance tecnológico también dio paso al surgimiento de nuevas problemáticas respecto a la protección de la información donde cibercriminales desarrollan técnicas para apropiarse de activos de información o vulnerar su seguridad para obtener beneficios económicos.

Para hacer frente a dicha problemática se requiere el uso de métodos y técnicas más sofisticadas para poder tener acceso a dicha información, es decir aplicar mejores técnicas de criptografía. El presente caso de uso se centra en la anonimización criptográfica la cual busca eliminar o reducir al máximo los riesgos de reidentificación de los datos ofreciendo garantías de privacidad y seguridad de la información sin distorsionar su usabilidad.

Se procederá a analizar tres técnicas de anonimización: aleatorización, generalización y seudonimización para que, con el avance de la investigación, encontrar la mejor metodología en cifrado de la información sensible, que tratan las farmacéuticas en las vacunas y medicamentos para el COVID-19.

II. ANONIMIZACIÓN DE DATOS PERSONALES

A. Principios de anonimización

Principio proactivo: La protección de la privacidad debe ser siempre el objetivo principal de la anonimización. Y la clasificación de la sensibilidad en los datos también debe hacerse.

Principio de privacidad por defecto: Se debe siempre garantizar la confidencialidad de los interesados.

Principio de privacidad objetiva: En todo proyecto que se anonimicen los datos, siempre se debe tener en cuenta que existe un umbral de riesgo, el cual se asume como riesgo aceptable.

Principio de plena funcionalidad: Desde el inicio se tendrá en cuenta la utilidad final de los datos anonimizados.

Principio de privacidad en el ciclo de vida de la información: Las medidas que garantizan la privacidad deben ser aplicables durante el ciclo completo.

Principio de información y formación: Informar y formar al personal involucrado en el proceso de anonimización.

B. Actores implicados en la anonimización

El proceso de anonimización se vuelve posible gracias a distintos roles implicados como el responsable del tratamiento quien decide sobre la finalidad de la información, el responsable de protección de datos quien atiende solicitudes y promueve auditorías sobre la privacidad de los datos. Equipo de evaluación de riesgo el cual evalúa el procedimiento y los resultados de la anonimización. Equipo de anonimización quien se encarga de determinar las técnicas según elementos y variables clave a anonimizar los cuales posteriormente serán evaluadas por el equipo de evaluación de riesgo. Finalmente, el equipo de seguridad de la información quien se encarga por velar la implementación de las medidas de seguridad necesarias durante la anonimización.[2]

III. TÉCNICAS DE ANONIMIZACIÓN

A. Aleatorización

Es una agrupación de técnicas que alteran o modifican la autenticidad y veracidad de los datos de esta manera se busca reducir la probabilidad de volver a vincular los datos entre ellos y/o con un individuo por deducción. Sus principales modalidades son:

- Adición de ruido: se transforman las propiedades de los datos haciéndolos menos precisos pero creíbles hasta cierto punto.
- Permutación: Intercambio de valores para vincular datos de diferentes registros, buscando que estos no sean altamente correlacionados.

B. Generalización

Modificación de datos a través de escalas u órdenes, generando esquemas de datos de acuerdo con características comunes para descartar la singularización. Métodos más utilizados:

- Agregación y anonimato: rangos que agrupan características o intervalos haciendo los datos más compartidos y por ende menos singulares.
- Diversidad y proximidad: agrupación de datos en la que ningún individuo pertenece a una categoría o dato individual.

C. Seudonimización

Es un proceso donde se sustituyen datos o atributos únicos por otros. Es una herramienta que permite reducir los riesgos de identificación. Hace uso de 5 técnicas [1]

1) Cifrado con clave secreta: quien posea la clave puede descifrar el contenido.
2) Función Hash: asigna un código único a un dato que lo identifica de los demás.
3) Función con clave almacenada: añade con la entrada del dato una clave secreta para generar un código con la función hash.
4) Función hash con borrado de clave: asigna números de forma aleatoria al conjunto de datos y borra la tabla de correspondencia.
5) Descomposición en tokens: mediante el uso de llaves de cifrado se combina con el dato original generando un dato con características similares a la inicial.

IV. PATENTES

A. *¿Qué es una patente?*

La patente tiene varias definiciones dependiendo de donde se realice la búsqueda, según la superintendencia de industria y comercio (SIC), una patente es un privilegio que el estado (En este caso el gobierno colombiano) el cual es tipo de reconocimiento por el trabajo realizado por el o los científicos implicados en la misma.

B. *¿Beneficios de una patente?*

El patentar un invento u otro tipo de **cosa** otorga bastantes beneficios al dueño de esta. Durante 20 años puede hacer uso del invento, explotarlo, comercializarlo y aprovecharse de su propiedad para ceder ciertos derechos a los que quieran usar la patente y así obtener beneficios económicos, todo esto aun teniendo la propiedad sobre la misma y teniendo derechos a comisiones y demás beneficios que dicte la ley.

C. *¿Qué no se puede patentar?*

El Articulo 20 de la decisión 486, el cual estableció la Comunidad Andina de Naciones, y que contiene ciertas normas que son de carácter obligatorio, que hablan sobre las patentes, en este caso el artículo se refiere más específicamente al Régimen Común de la propiedad industrial y nos dictamina que no se considera como patente. Allí establecen que los descubrimientos que se encuentren en la naturaleza, las teorías científicas, los métodos matemáticos, terapéuticos, quirúrgicos, financieros o de negocios, las obras de tipo artístico, literario, científico o el software, esto último se entiende protegido por el derecho de autor, no se pueden patentar.

Uno de los ejemplos de esto, es la relatividad general propuesta por Albert Einstein, la cual permitió dar a la física clásica un salto enorme por las brechas que abrió. Esta teoría no se puede patentar debido a que es algo que está presente en la naturaleza, más específicamente el universo, por lo cual no puede ser otorgado una licencia o un dominio sobre esta.

D. *¿Qué requisitos debe tener mi invento para poder patentarlo?*

Para que se pueda patentar un invento debe cumplir ciertos requisitos, según la SIC esta toma en cuenta tres requisitos para poder crear la patente. Primero: El invento tiene que ser algo que no exista previamente en ninguna parte del mundo. Segundo: Debe ser algo innovador y creativo. Tercero: El invento debe poder ser fabricado y utilizado a nivel industrial.

E. *Datos sensibles de las patentes*

Para fines de nuestro proyecto como ya se menciono en la introducción, vamos a tener en cuenta el sector farmacéutico. Vamos a centrarnos en las vacunas del COVID-19, las cuales hay varias y cada una debería tener una formula diferente para su creación. Dentro de esto existen varios datos de tipo sensible como lo son, la formula química, los ingredientes y las proporciones a usar de cada ingrediente para crearla, así como también el proceso de creación de esta es un dato sensible.

V. TOKENIZACIÓN

A. *¿Qué es?*

La tokenización permite convertir datos sensibles o confidenciales en datos no sensibles, los cuales se llaman "tokens". Con estos tokens se intercambia los datos originales (que van a ser guardados en la nube) por datos completamente adulterados que no tienen ninguna relación, es decir que no se pueden revertir para llegar a los datos originales y que por tanto usar tokens permite proteger información confidencial.[5]

B. *¿Cómo funciona?*

A través del uso de llaves de cifrado se combina el dato original y la llave asignada para adulterar el dato original y de esta manera la información que será almacenada en los sistemas de información no corresponde al dato original. Como resultado se obtiene un dato que conserva las características iniciales del dato original (longitud y tipo de carácter) pero que no corresponde al original. [6]

C. *Aplicación al proyecto*

Se va a hacer uso de un tipo especial de tokens, los tokens de utilidad los cuales dan acceso directo a un producto o una plataforma de forma temporal. En este caso, los tokens van a permitir el acceso a los datos sensibles en un archivo de los laboratorios de investigación de las farmacéuticas. En donde se va a generar un token que varía con la hora, metadata del archivo con los datos sensibles y los datos del investigador que ingresa con el objetivo de facilitar el acceso a información confidencial y monitorear el acceso de una forma segura y sencilla.

VI. GESTIÓN DE RIESGOS

Partiendo de la premisa "Para saber que debemos proteger, primero debemos conocer que activos tenemos", realizaremos un proceso de identificación de activos relevantes para el proyecto.

A. *Identificación de activos*

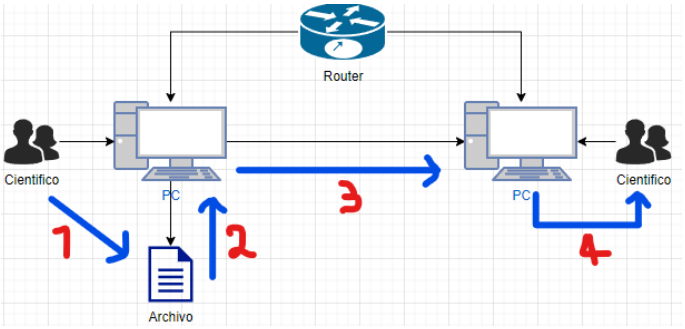
Tomamos en cuenta la triada de la seguridad CID (Confidencialidad, integridad y disponibilidad) para el siguiente cuadro donde realizamos la identificación de los activos que consideramos relevantes para el proceso.

Tomamos en cuenta cada parte de la CID para darle una clasificación a los activos, la cual es de bajo, medio y alto, donde bajo implica que para esa parte la triada no es importante, mientras que medio y alto eleva su relevancia y nos indica en donde se debe tener más cuidado.

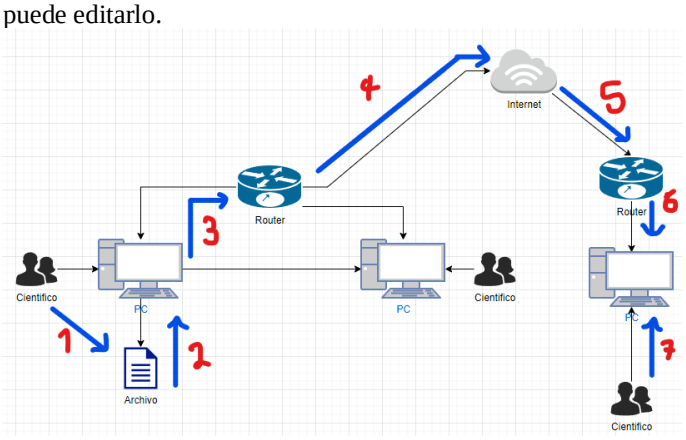
	C	I	D
Computador	Alto	Alto	Alto
Científicos	Alto	Bajo	Medio
Archivos web/ (formulas, documentación mensajes)	Alto	Alto	Alto
Router	Media	Alto	Alto

Ilustración 1TABLE 1

B. *Diagrama de activos y procesos*



Aquí podemos visualizar un diagrama con los activos descritos en la TABLE I, pero también podemos encontrar el proceso de envío de un archivo de un Pc a otro de forma local, donde esta divido por 5 pasos. Paso 1: El científico crea o edita un archivo en un computador con datos sobre la investigación de la vacuna. Paso 2: Se prepara el archivo para su envío. Paso 3: Se envía el archivo hacia otro computador donde otro científico espera el mismo. Paso 4: El científico recibe el archivo y ahora puede editarlo.



Aquí podemos visualizar un diagrama con los activos descritos en la TABLE I, pero con un proceso diferente donde se envía un archivo de un PC a otro que no se encuentra en la misma red, donde esta divido por 5 pasos. Paso 1: El científico crea o edita un archivo en un computador con datos sobre la investigación de la vacuna. Paso 2: Se prepara el archivo para su envío. Paso 3: Se envía el archivo hacia otro computador por medio de un correo donde otro científico espera el mismo.

Paso 4: El archivo entrará al internet donde deberá navegar hasta el otro PC. Paso 5: El archivo llega a la otra red desde el internet. Paso 6 y 7: el archivo finalmente llega al PC para ser edita por el otro científico.

C. *Identificación de riesgos*

1. El científico envíe el archivo sin consentimiento hacia otra persona que no deba conocer el archivo.

- Alto impacto
- Probable
- Alto riesgo
- Mitigar

2. El computador se dañe y se pierdan los datos.

- Alto impacto
- Poco Probable
- Riesgo medio
- Mitigar

3. Si el archivo solo se maneja de local, y sucede un desastre natural y se pierda toda la investigación.

- Alto impacto
- Poco Probable
- Riesgo medio
- Mitigar

4. Un ciberataque.
- Alto impacto
- Muy Probable
- Alto riesgo
- Mitigar

D. *Controles*

1. Registro de envíos que se realizan en los computadores

2. Backups

3. Backups en un servidor o base de datos que tenga el laboratorio

4. Login falso, donde se muestra una página de autenticación que solo retorna al mismo login sino se tiene un token para acceder al archivo

5. Tokenizacion del archivo para protegerlo de ciberataques

VII. ENCUESTA I

Se realizo una encuesta para conocer la opinión y percepción de la gente acerca del tema de las vacunas y las patentes, además de temas de seguridad en el cual se fundamenta este proyecto.
Esta encuesta fue de tipo virtual, y se realizo mediante la herramienta de formularios de Google. Se encuestaron a 32 personas en total.

A. *Recolección de datos*

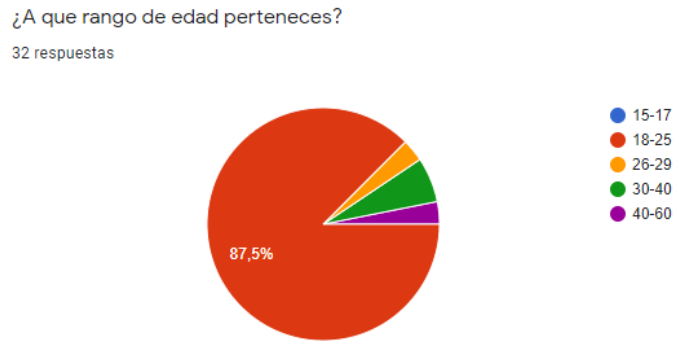


Ilustración 2Rango de edad

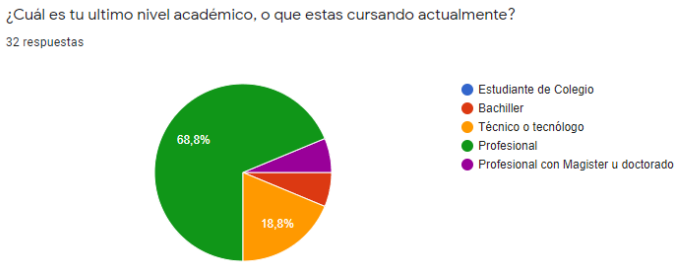


Ilustración 3Nivel educativo

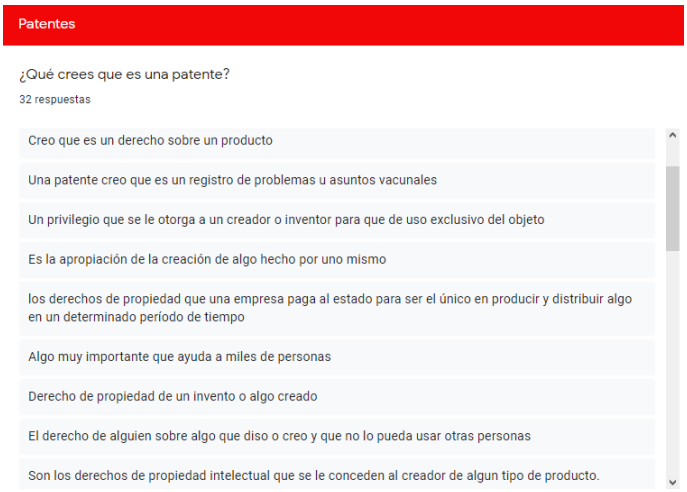


Ilustración 4¿Qué es una patente?

Vacunas

¿Crees que es importante que las vacunas contra el Covid-19 estén patentadas?

32 respuestas

- No, los medicamentos o vacunas de importancia máxima para la salud no deberían ser patentados. La salud se vuelve un negocio y éticamente estaría bastante errado.
- Si, porque es algo creado y eso se debe respetar, no todos los países cuenta con la infraestructura para producir vacunas.
- No, ya que es una solución para la población entera. Y no un negocio
- No, esta patente debería ser libre para poder producir en mayores cantidades las vacunas y lograr cubrir más rápido mayor parte de la población
- si
- No, ya que está relacionado con un problema de salud a nivel global, deberían ser libres de patentes para que aún más personas tengan acceso a ellas
- No, porque implicarían costos de más.

Ilustración 5Importancia patentes

¿Sabias que si las patentes fueran libres las vacunas serian prácticamente gratis y de fácil acceso?

32 respuestas

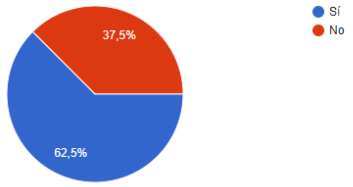


Ilustración 6Patentes gratis

¿Cuáles crees que serian los riesgos de que las patentes de las vacunas sean de tipo libre?

32 respuestas

- La poca regulación
- Las vacunas podrían ser modificadas ya que habría una libre distribución
- Riesgo de que hallan patentes de un mal sitio
- las consecuencias y efectos secundarios que tendrían estas en nosotros
- Que se le de mal uso y resulte una vacuna contraproducente y peligrosa
- Irresponsabilidad
- la creación de vacunas alternas o similares que no cumplan con todo lo necesario para combatir el virus
- No saber los casos contraprucentes que se adquieren al aceptar la vacuna
- No seria regulado

Ilustración 7Riesgos

Supongamos que la patente es libre. Si vendieran la vacuna en tu ciudad, y crees que es de una "fuente confiable", la comprarías?

32 respuestas

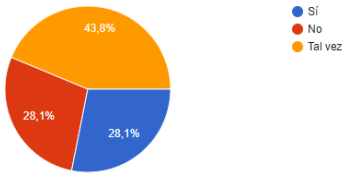


Ilustración 8Venta vacuna

Siguiendo el punto anterior, ¿aplicarías esta vacuna a tus familiares?

32 respuestas

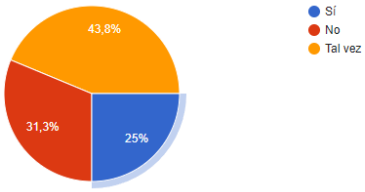


Ilustración 9Vacuna familiares

Seguridad

Al la vacuna estar patentada, se deben proteger los datos acerca de esta, como la formula, composición y demás, ¿Usted cree que es importante proteger estos datos?

32 respuestas

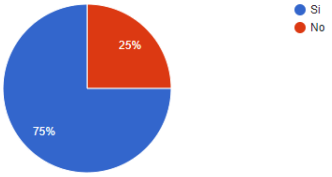


Ilustración 10Seguridad

¿Cuáles cree que serian los riesgos de que estos datos de las vacunas sean robados o filtrados en el internet?

32 respuestas

- La falsificación de estas vacunas
- La vacuna se puede alterar
- Que halla mala información de esta
- Que se puede robar información de gran validez
- Un mal uso. Pero si no se cuenta con los recursos, no se podría hacer nada con esa información
- Irresponsabilidad
- pueden ser utilizados negativamente para la sociedad
- No saber de que está hecha la vacuna
- Comercialización ilegal de las vacunas

Ilustración 11Riesgos

¿Usted intentaría crear su propia vacuna si los datos de esta se encuentran rondando por la red?

32 respuestas

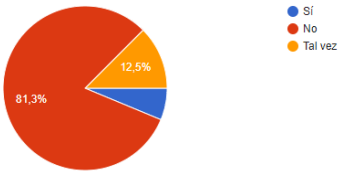


Ilustración 12Vacuna propia

B. Análisis de datos

Con base en la encuesta de estas 32 personas las cuales en su mayoría se encuentran en el rango de 18 a 25 años y tienen una formación (bachiller/técnico/profesional/magister u doctorado) identificamos que:

Los encuestados tienen una buena noción de lo que son las patentes y la importancia de estas, más sin embargo difieren en que las vacunas deban o no patentarse pues algunos cuestionan la seguridad de la información para su creación, pero otros dicen que patentarla hace de la vacuna un negocio.

Posteriormente los encuestados coinciden en el ítem de los riesgos de que las patentes sean libres (Ilustración 7 y 11) pues nos encontramos con temas como "No regulación, falsificación, modificación de la vacuna, casos contraproducentes, mal uso, comercialización ilegal" entre otros efectos negativos que podemos ver traducidos en que un 75% de los encuestados considere importante la protección de estos datos (Ilustración 10).

VIII. CONCLUSIONES

- En el proceso de compartir datos sensibles dentro de un laboratorio, la tokenización es un elemento clave que permite mitigar los riesgos presentes.
- Teniendo en cuenta la criticidad en la confidencialidad de activos de información críticos como los científicos se vuelve imprescindible que los sistemas garanticen el acceso seguro y eficiente a los archivos de investigación.
- Entre las técnicas de anonimización, se encuentra que la descomposición en tokens permite una mayor seguridad y facilidad de implementación en los entornos de investigación de los laboratorios.

REFERENCIAS

1. https://www.archivogeneral.gov.co/sites/default/files/Estructura_Web/5_Consulte/Recursos/Publicacionees/Guia_de_Anonimizacion-min.pdf
2. https://www.aepd.es/sites/default/files/2019-12/guia-orientaciones-procedimientos-anonimizacion.pdf
3. https://ayudaleyprotecciondatos.es/2019/04/29/anonimizacion-datos-personales/
4. https://adefinitivas.com/arbol-del-derecho/analisis-de-la-seguridad-del-tratamiento-en-tecnologia-blockchain-anonimizacion-cifrado-seudonimizacion-a-cargo-de-ulises-david-gonzalez/
5. https://www.ambito.com/opiniones/token/que-es-la-izacion-y-como-funciona-n5196307
6. https://www.b-secure.co/estrategias/datos/tokenizacion-y-enmascaramiento