Captura de Requerimientos

Versión

ID	*Descripción*
1.0	Versión inicial del documento
2.0	Agregado de supuestos y dependencias
3.0	Glosario actualizado. Referencias a otros documentos

Tabla de Contenidos

Introducción

Propósito

Este documento describe los requerimientos funcionales y no funcionales de la herramienta de anonimización de datos a ser desarrollada para Onapsis, sirviendo como base contractual entre los desarrolladores y el cliente.

Audiencia del documento

Este documento se encuentra destinado a la siguiente audiencia:

- *Los desarrolladores que implementarán el sistema*: este documento debe brindar la información necesaria para el diseño, la implementación y el testeo del sistema solicitado y los correspondientes criterios de aprobación a cumplir.
- *Cliente*: este documento permite definir al cliente en qué consiste su necesidad y cómo desea que la misma satisfecha, especificando los criterios de aprobación que deberá cumplir el sistema.
- *Tutor/Director del desarrollo:* debido a que este sistema está siendo implementando como trabajo profesional, este documento permite reflejar al tutor y director del trabajo el desarrollo a realizar.

Alcance del proyecto

El cliente se encuentra desarrollando una plataforma de seguridad informática orientada a controlar, monitorear y auditar sistemas SAP. Los sistemas SAP permiten a las empresas ejecutar y optimizar distintos aspectos como los sistemas de ventas, finanzas, operaciones bancarias, compras, fabricación, inventarios y relaciones con los clientes.

Esta plataforma permite hacer diversos tipos de escaneos con el fin de encontrar vulnerabilidades existentes en sistemas de este tipo. A su vez permite controlar la red para detectar y responder ante posibles ataques en tiempo real.

Una limitación actual es que no se poseen datos concretos que permitan llevar a cabo la creación de modelos de riesgos eficaces. Los tipos de vulnerabilidades, los sistemas afectados, la frecuencia de los ataques, etc., resultan datos esenciales para el desarrollo de la plataforma.

Debido a esto, el cliente está implementando un nuevo módulo cuyo objetivo es la recaudación de este tipo de información, con el fin de analizarla, obtener conclusiones, y continuar mejorando su plataforma.

Este nuevo módulo lleva consigo una importante restricción: la información debe encontrarse anonimizada para poder ser utilizada. Esto se debe a que se trata de datos sensibles que contiene atributos que permitirían identificar al propietario de esa información.

Por lo tanto el sistema requerido por el cliente consiste en una aplicación que permita anonimizar información de forma segura, respetando los modelos de privacidad existentes hoy en día y con la menor pérdida de información posible, con el fin de que luego pueda ser utilizada en la plataforma de seguridad del cliente.

Referencias

ID	*Documento*	*Descripción*	*Referencia*
0	Proyecto de trabajo profesional	Repositorio para el versionado del código	https://github.com/s-rodriguez/edat
1	User stories & Wireframes	Documento con las user stories y wireframes definidos	https://github.com/s-rodriguez/edat/blob/develop/docs/project_design/UserstoriesyWireframes.pdf
2	Herramientas seleccionadas	Documento con las herramientas a utilizar en el proyecto	https://github.com/s-rodriguez/edat/blob/develop/docs/project_design/Herramientasseleccionadas.pdf

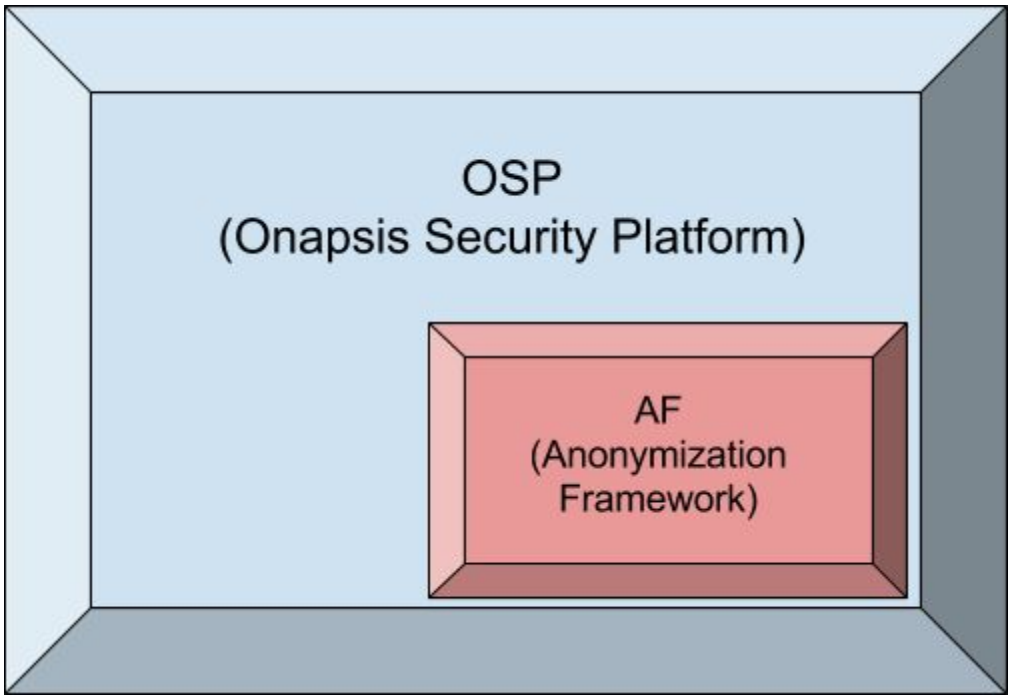
Descripción general

Perspectiva del producto

El desarrollo de esta plataforma de anonimización nace como una necesidad para ampliar un producto ya existente. Este producto, OSP[1] (Onapsis Security Platform), es la primera plataforma orientada hacia la seguridad informática de sistemas SAP. A través del monitoreo continuo, combina capacidades de análisis de vulnerabilidades, cumplimientos de políticas de seguridad, y de detección y respuesta en tiempo real que permiten detectar fallas y corregirlas para asegurar sistemas y aplicaciones SAP.

El objetivo detrás de esta implementación es el poder obtener datos de diversos tópicos relacionados con sistemas y aplicaciones SAP (como por ejemplo parámetros de configuración), con el fin de poder realizar un análisis profundo y generar un correcto modelo de riesgo asociado.

Dado que la información que se desea obtener es de carácter confidencial y sensible, el nuevo módulo de anonimización sería el objeto clave a integrarse dentro de OSP, que permitiría hacer una transformación correcta y segura de los datos, para luego poder ser enviados y analizados.



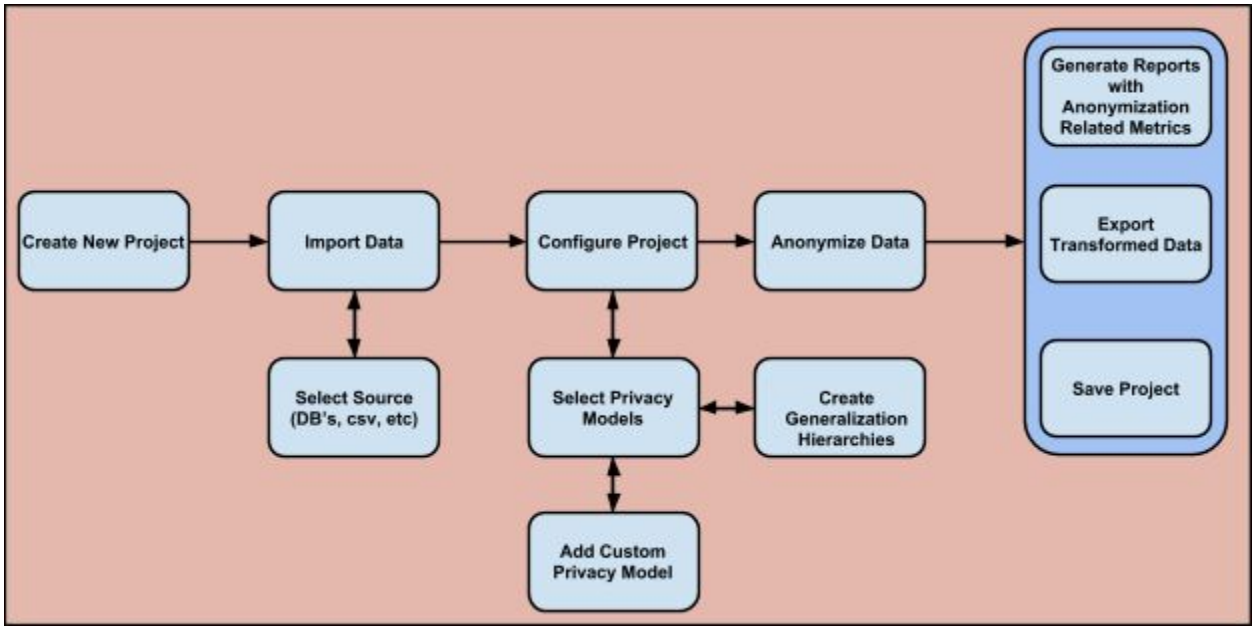
[1] "Onapsis Security Platform | Cyber Security Solution for SAP ..." 2015. 15 Sep. 2015 <https://www.onapsis.com/products/onapsis-security-platform>

Características principales del producto

Se desea que el usuario pueda adaptar la plataforma EDAT a sus necesidades, y no a la inversa; de forma tal que los datos sean anonimizados de manera segura, pero al mismo tiempo se pueda sacar el máximo provecho de ellos basados en los distintos contextos donde se quiera usar dicha herramienta.

Entonces, se podrían enumerar las siguientes como las características principales:

- Transformación de datos utilizando técnicas y modelos de privacidad conocidos.
- Posibilidad de agregar plugins de modelos de privacidad personalizados para el usuario.
- Personalización de jerarquías de generalización y de supresión por parte del usuario.
- Creación, Exportación/Importación de proyectos y sus configuraciones para reutilizarlos.
- Lectura de datos originales de distintas fuentes (Distintos tipos de bases de datos, etc.)
- Generación de reportes para evaluar la transformación realizada y la correspondiente pérdida de información
- Aplicación orientada al usuario con una interfaz amigable e intuitiva.



Clases de usuarios y características

User 1: Usa la aplicación de manera enterprise para anonimizar datos que le parecen relevantes.
User 2: Usa la aplicación para realizar investigaciones en el tema de anonimización, probando y comparando los distintos tipos de modelos de privacidad y algoritmos, haciendo uso de los reportes para obtener métricas relacionadas.
User 3: Usa la librería de anonimización para extenderla, agregando nuevos modelos de privacidad métricas y/o algoritmos.
User 4: Usa la librería de anonimización para su integración con otro módulo o aplicación existente.

Entorno operativo

El desarrollo inicial de la plataforma está orientado hacia el siguiente entorno operativo:

Hardware:
- Procesador: Intel Core i5 (o símil)
- RAM: 6Gb (Mínimo)
- Espacio en disco: 700Gb

Software:
- Sistema Operativo: Ubuntu 14.04 (Posibilidad de extenderlo a independencia de plataforma)
- Lenguaje de desarrollo principal: Python 2.7.9

Diseño y restricciones de implementación

Limitaciones de Hardware: Como restricción de tamaño de base de datos, se establecerá un límite de registros máximo dentro del cual se asegura el correcto rendimiento del framework. En principio, se considera una base de datos de tamaño grande, aquella que contenga entre 10^7 y 10^9 registros; en base a evaluaciones a realizar, este rango puede ser modificado para asegurar el correcto rendimiento del software.

Requerimientos de Lenguaje: Toda la documentación debe ser realizada en inglés, así también todo el código a desarrollarse.

Tecnologías, herramientas y bases de datos específicas: Tanto la aplicación como la librería de anonimización deben ser desarrolladas en el lenguaje de programación Python. Esto se debe a que la aplicación luego será mantenida por el cliente. La aplicación deberá trabajar con sistemas de gestión de base de datos relaciones del tipo SQL y archivos de extensión CSV.

Estándares de Programación: El código de la aplicación debe respetar las **Python PEP8 policies**[2], debido a que es el estándar que utiliza Onapsis, y el cliente será quien se ocupe de mantener el framework en un futuro.

Desarrollo de interfaces con otras aplicaciones: El framework que contiene el engine de anonimización, será integrado dentro del producto del cliente. Sin embargo, no es necesario establecer ningún tipo de comunicación con otras aplicaciones, ya que el framework será utilizado como una librería tradicional de Python dentro del producto, y por ende, todo tipo de integración quedará como responsabilidad del cliente.

Documentación de usuario

Junto con la plataforma de anonimización implementada, serán entregados los siguientes documentos:

- Manual de usuario: Informe detallando las distintas funcionalidades del producto, junto con los casos de uso existentes.
- Documentación técnica: Informe técnico detallando todo lo desarrollado (desde un punto de vista open-source). Libre para ser accedido a todo aquel que esté interesado en la parte técnica de la implementación o quien quiera colaborar con el desarrollo del mismo.

Toda la documentación estará disponible de manera on-line.

Supuestos y dependencias

- Los usuarios de la aplicación/framework tienen privilegios suficientes para manejar los datos originales.
- Las librerías de bases de datos ya existentes que serán utilizadas, funcionan de manera correcta para la conexión, realizar consultas y extraer datos de las mismas.
- El cliente proveerá bases de datos para ser utilizadas durante el desarrollo y el testeo de la aplicación.
- Los modelos de privacidad existentes son eficientes y pueden ser plasmados en código para su reúso.
- La aplicación ofrecerá inicialmente dos modelos de privacidad distintos, a ser definidos por el cliente.
- Las reuniones de avance y entregables se cumplen acorde al calendario establecido.

[2] Python PEP8 policies
<https://www.python.org/dev/peps/pep-0008/>

Requerimientos funcionales del sistema

1. Creación o importación del proyecto

Descripción: El usuario deberá poder gestionar un proyecto asociado a un proceso de anonimización de datos, con las correspondientes operaciones de alta, baja y modificación.

Prioridad: Media.

Estímulo/Respuesta y Secuencias:

1. El usuario entra la aplicación.
2. El sistema muestra la opción de crear un nuevo proyecto, o importar un proyecto ya existente.
3. El usuario selecciona si crear un nuevo proyecto o importar un proyecto.
    a. En el caso de que seleccione un nuevo proyecto, se debe crear un nuevo proyecto vacío con un archivo de configuración asociado.
    b. En el caso de que seleccione abrir un proyecto ya existente, se debe importar a partir de un archivo de configuración, toda la información y datos asociados al proyecto tales como: base de datos de input, base de datos anonimizadas, parámetros de configuración de transformación, etc.

Requerimiento funcional: La gestión del proyecto le brindará al usuario la posibilidad de persistir las transformaciones de datos realizadas con sus correspondientes reportes y las configuraciones utilizadas para llevar a cabo dicho proceso. A su vez permitirá poder agrupar de forma lógica toda esta información.
De esta forma el usuario podrá en cualquier momento guardar el proyecto, permitiendo poder recuperarlo posteriormente en el mismo estado en el cual fue guardado por última vez.
La gestión del proyecto estará regulada por un archivo de configuración. La corrupción o eliminación del mismo ocasionará la imposibilidad de la importación del proyecto.
Cualquier proyecto existente podrá ser eliminado, permitiendo seleccionar si borrar únicamente el proyecto o también todos los datos asociados al mismo (base de datos, reportes, etc.).
No se podrá realizar una anonimización de datos sin que se haya creado un proyecto que contenga esta transformación.

2. Importación de datos

Descripción: El usuario podrá anonimizar datos que sean resultantes de la importación de una base de datos relacional y/o archivos de extensión CSV.

Prioridad: Alta.

Estímulo/Respuesta y Secuencias:

1. El usuario se encuentra dentro de un proyecto existente y selecciona la opción de importar datos.
2. El sistema muestra un explorador de archivos, permitiendo seleccionar un archivo que respete las extensiones soportadas.
3. El usuario selecciona un archivo a importar al proyecto.
4. El sistema importa los datos:
    a. En el caso de que sea un archivo de extensión invalida, el sistema notifica el error y retorna a 2).
    b. En el caso de que sea un archivo de extensión válida pero corrupto o mal formado, el sistema muestra el error correspondiente y retorna la pantalla principal del proyecto.
    c. En el caso de que se trate de un archivo bien formado y de extensión soportada, se importa el proyecto vinculándolo al proyecto actual. Se retorna a la pantalla principal del proyecto, pudiendo realizar operaciones con la base de datos importada.

Requerimiento funcional: La importación de datos permitirá al usuario poder seleccionar los datos que desee transformar vinculando al mismo tiempo estos datos a un determinado proyecto.
Se podrá importar datos pertenecientes a una base de datos relacional. Inicialmente se soportará únicamente bases de datos del tipo SQLite y hojas de cálculo CSV. Se deberá implementar este módulo de forma tal de que sea fácilmente extensible ya que el cliente desea agregar en el futuro soporte para nuevas extensiones (MySQL y PostgreSQL).No se brindará soporte para base de datos no relacionales.
Al importar una base de datos, se podrá visualizar y trabajar con cualquiera de las tablas existentes en la misma. En el caso de archivos de extensión CSV, se podrá transformar cualquiera de sus hojas.
Los datos importados serán vinculados al proyecto mediante su archivo de configuración. En el caso de que los archivos sean removidos de forma externa a la aplicación, se perderá toda la información vinculada.

3. Configuración de atributos

Descripción: El usuario podrá definir configuraciones particulares de los atributos a ser utilizadas durante el proceso de anonimización de los datos seleccionados.

Prioridad: Alta

Estímulo/Respuesta y Secuencias:

1. El usuario se encuentra dentro de un proyecto existente y ha seleccionado los datos a importar.
2. Dentro de la ventana de gestión del proyecto, se muestra un panel de configuración. El usuario se mueve por los distintos campos, definiendo las distintas configuraciones deseadas:
    a. *Tipo de Atributo*: El usuario selecciona para el atributo al que está relacionado el dato, de que tipo es.
    b. *Categoría de Anonimización:* El usuario selecciona a qué categoría pertenece el campo, siendo las opciones posibles:
        i. Identificable
        ii. Cuasi-Identificable
        iii. No Identificable
        iv. Sensible
        v. No sensible
    c. *Peso:* El usuario podrá definir un peso a aquellos atributos que haya marcado como cuasi-identificables.

Requerimiento funcional: La configuración de los atributos tiene como objetivo que el usuario pueda determinar distintos aspectos y características relativas a la información que se desea anonimizar, con el fin de que el proceso de transformación se ajuste de manera adecuada al contexto en el cual está siendo utilizada.

En principio, para el tipo de atributo, las opciones pedidas son las siguientes:

- Número entero (*Int*)
- Cadena de caracteres (*String*)
- Fecha (*Date*)

El peso será para atributos que hayan sido seleccionados como cuasi-identificables, con el fin de darles mayor relevancia a la hora del proceso de anonimización y tratar de evitar que dichos atributos sean suprimidos o generalizados a grandes niveles.

4. Transformación de datos mediante supresión

Descripción: El usuario podrá anonimizar la información mediante la técnica de supresión parcial o total de los datos, permitiendo configurar cuáles atributos y de qué forma serán suprimidos.

Prioridad: Alta

Estímulo/Respuesta y Secuencias:

1. El usuario se encuentra dentro de un proyecto existente, y con un set de datos importados. El usuario selecciona dentro de la ventana de gestión del proyecto, la opción de supresión de datos.
2. El sistema muestra un menú con todos los atributos asociados a esa tabla u hoja de cálculo, permitiendo seleccionar cuáles se deben suprimir y de qué manera aplicar esta técnica.
3. El usuario selecciona cada uno de los atributos que desea suprimir en el proceso de anonimización de datos y cuando termina, selecciona la opción aceptar.
4. El sistema actualiza la configuración para realizar el proceso de anonimización según la selección del usuario.

Requerimiento funcional: La supresión de atributos permitirá al usuario eliminar atributos de forma parcial o total. Los atributos seleccionados como identificables previamente por el usuario, serán suprimidos de forma automática en el proceso de anonimización, sin requerir configuración alguna. Los atributos que podrán ser sometidos a la técnica de supresión (parcial o total) serán los catalogados por el usuario como cuasi identificables.
Se podrá seleccionar si la supresión se realizará de atrás hacia adelante o viceversa, y si se desea mantener una cantidad mínima de valores para un determinado atributo al momento de realizar el proceso de anonimización.
Los atributos cuasi identificables que han sido seleccionados para ser generalizados previamente, no podrán ser seleccionados para ser suprimidos.
Se deberá persistir toda esta información en el archivo de configuración del proyecto.

5. Transformación de datos mediante generalización

Descripción: El usuario podrá anonimizar la información mediante la técnica de generalización de datos, permitiendo crear jerarquías para aquellos atributos que considere necesarios.

Prioridad: Alta

Estímulo/Respuesta y Secuencias:

1. El usuario se encuentra dentro de un proyecto existente, y con un set de datos importados. El usuario selecciona dentro de la ventana de gestión del proyecto, la opción de generalización de datos.
2. El sistema muestra un menú con todos los atributos asociados a esa tabla u hoja de cálculo. El usuario podrá seleccionar los distintos atributos a los cuales les quiera crear una jerarquía de generalización.
3. Cuando el usuario selecciona un atributo, el sistema muestra una nueva ventana. La misma sirve para crear y editar jerarquías de generalización.
4. El usuario guarda la nueva configuración seleccionado el botón correspondiente.
5. El sistema actualiza la configuración para realizar el proceso de anonimización según la selección del usuario

Requerimiento funcional: La generalización de atributos permitirá al usuario ocultar datos que pueden resultar sensibles o cuasi identificables, y al mismo tiempo brindar información de dichos atributos.
Sólo aquellos atributos seleccionados como cuasi identificables previamente por el usuario, podrán ser elegidos para armar las reglas de generalización (o también conocidas como jerarquías de generalización)
Para el armado de jerarquías, el usuario podrá generar grupos entre los distintos valores posibles de un atributo, y con dichos grupos armar niveles de jerarquías. (Por ejemplo, aquellos números menores a 5 entrarían en un grupo, y los mayores a 5 pero menores a 10 en otro; resultando en un primer nivel de jerarquías: "<=5" & "5<x<=10")
Se deberá persistir toda esta información en el archivo de configuración del proyecto

6. Selección de criterio de privacidad

Descripción: El usuario podrá seleccionar bajo qué criterio de privacidad la transformación de datos se llevará a cabo; y de ser posible, elegir el algoritmo a utilizar de dicho criterio.

Prioridad: Alta

Estímulo/Respuesta y Secuencias:

1. El usuario se encuentra dentro de un proyecto existente, y con un set de datos importados. El usuario selecciona dentro de la ventana de gestión del proyecto, la opción de criterios de privacidad
2. El sistema muestra una nueva ventana, donde listará los distintos criterios de privacidad disponibles al momento.
3. El usuario podrá seleccionar un criterio del listado, lo cual generará que se muestre la configuración para dicho criterio. (La configuración estará con valores por defecto, pero el usuario podrá modificarlos en el caso de desearlo)

4. Una vez seleccionada la configuración del criterio elegido, el sistema actualiza toda la información necesaria para realizar el proceso de anonimización y se persiste en la configuración del proyecto.

Requerimiento funcional: La selección de criterio de privacidad, le permitirá al usuario especificar bajo qué modelo teórico se llevará a cabo el proceso de transformación; siendo este modelo el marco de referencia que determinará si la transformación realizada cumple o no con las premisas requeridas por dicho criterio.
Dado que los criterios de privacidad son modelos, y en la práctica se emulan con algoritmos, el usuario podrá elegir qué algoritmo utilizará para realizar el proceso de transformación de datos.
Anonimizar datos cumpliendo con criterios de privacidad, permitirá que el usuario logre tener una mayor confianza en que la información ha sido transformada de manera segura, y que el riesgo por de-anonimización sea bajo, o mínimamente cuantificable relativo al modelo elegido.

7. Extensión de criterios de privacidad (Modelos y Algoritmos)

Descripción: El usuario podrá agregar nuevos modelos y algoritmos de criterios de privacidad cómo módulos nuevos para la plataforma

Prioridad: Alta

Estímulo/Respuesta y Secuencias:

1. El usuario se encuentra dentro de un proyecto existente, y con un set de datos importados. El usuario selecciona dentro de la ventana de gestión del proyecto, la opción de criterios de privacidad
2. El sistema muestra una nueva ventana, donde se listará los distintos criterios de privacidad disponibles al momento.
3. El usuario selecciona un botón que hace referencia al importar nuevos módulos de criterios de privacidad
4. El sistema muestra un pop up de tipo ventana, para buscar dentro de los archivos existentes de la computadora
5. El usuario busca el archivo en el directorio correspondiente, y lo selecciona para su importación
6. El sistema importa los datos:
    a. En el caso de que sea un archivo de extensión inválida, el sistema notifica el error y retorna a 2).
    b. En el caso de que sea un archivo de extensión válida pero corrupto o mal formado, el sistema muestra el error correspondiente y retorna la pantalla principal del proyecto.
    c. En el caso de que se trate de un archivo bien formado y de extensión soportada, se importa el nuevo módulo a la plataforma, guardando dicho

módulo para futuros usos en otros proyectos. Se retorna a la ventana donde se listan los criterios de privacidad, y entre ellos se debe mostrar el nuevo módulo importado. El usuario puede seleccionar un criterio, si desea hacerlo, o volver para atrás a la pantalla principal del proyecto.

Requerimiento funcional

La posibilidad de agregar nuevos algoritmos y modelos de privacidad, permitirá al usuario extender la plataforma con nuevos módulos, e incluso adaptarla de forma tal que la transformación de datos tenga en cuenta el contexto bajo el cual se está anonimizando.

Los módulos deberán ser en principio archivos de tipo Python, que respete una determinada interfaz a especificar en el futuro, de forma tal que la estructura de los módulos sea la misma y pueda ser reutilizable polimórficamente.

Cada vez que se importa un nuevo módulo, el sistema debe validar que sea de extensión correcta, y que contenga las firmas de la interfaz. Si el módulo tiene errores de sintaxis o de otro tipo, no serán validados, quedando estos como responsabilidad de quien los desarrolle.

La plataforma deberá, una vez validado el módulo, integrarlo a un subdirectorio propio donde tendrá todos los modelos y algoritmos disponibles para ser seleccionados en la configuración de los proyectos.

8. Anonimización y Análisis de Resultados

Descripción: El usuario podrá realizar la transformación de datos desde un set original a un set anonimizado, utilizando las configuraciones previamente establecidas en las etapas. Al mismo tiempo, será capaz de hacer un análisis rápido de los resultados mediante distintas métricas que serán mostradas al finalizar la anonimización de los datos.

Prioridad: Alta

Estímulo/Respuesta y Secuencias:

1. El usuario se encuentra dentro de un proyecto existente, y con un set de datos importados.
2. El usuario ya ha configurado el proyecto acorde a sus preferencias para el proceso de anonimización (Relacionado con requerimientos 3, 4, 5, 6 y 7)
3. El usuario selecciona el botón “**Anonimizar**”
4. El sistema muestra una ventana nueva. Dicha ventana contiene la configuración que el usuario ha seleccionado previamente.
5. El usuario
    a. De estar conforme con la configuración, selecciona la opción de *Anonimizar*.

i. El sistema muestra una ventana de progreso, y realiza la transformación de datos adecuada.
ii. Al finalizar el proceso de anonimización de los datos, el sistema:
    1. Genera y guarda métricas (basadas en la transformación realizada), y las muestra en una nueva ventana.
iii. El usuario puede analizar dichas métricas, y luego cerrar la ventana, retornando a la pantalla principal de gestión del proyecto

b. De no estar conforme, puede seleccionar la opción sobre el botón *Cancelar* y la ventana se cierra volviendo a la pantalla principal de gestión del proyecto.

Requerimiento funcional:

El poder realizar anonimización de datos basándose en configuraciones previamente establecidas (personalizadas para el contexto en el cual se encuentra el usuario) es el requerimiento funcional clave para librería a implementar.

A su vez, con las métricas generadas al finalizar la anonimización de datos, el usuario será capaz de analizar si dicha configuración es segura y se adapta a los modelos de privacidad elegidos; y de no serlo, modificar la configuración de manera iterativa.

Mediante esto, el mismo será capaz de transformar sus datos originales tantas veces quiera, hasta estar convencido que el set resultante con determinada configuración es el óptimo.

Las métricas serán a definir en el futuro, cuando el proyecto se encuentre en un estadío más maduro

9. Reportes con métricas

Descripción: El usuario podrá generar reportes con métricas basadas en el resultado de la anonimización de datos

Prioridad: Media

Estímulo/Respuesta y Secuencias:

1. El usuario se encuentra dentro de un proyecto existente, y con un set de datos importados. Ya ha realizado la transformación de los datos, y está nuevamente en la pantalla principal
2. El usuario hace click en el botón ***Reportes***.
3. El sistema muestra una nueva ventana conteniendo todas las métricas generadas en el proceso de anonimización.
4. El usuario selecciona el botón ***Exportar***.
5. El sistema muestra una ventana, donde el usuario seleccionará donde será exportado el reporte.
6. El usuario elige una ubicación y hace click en ***Aceptar***. El sistema genera el reporte y vuelve a la pantalla principal.

Requerimiento funcional:
El poder generar reportes con métricas, le permitirá al usuario hacer un análisis más profundo respecto a la configuración que está utilizando, a los datos que está procesando, y si son correctos o no para el fin que busca.
Así mismo, dicho documento puede ser utilizado para mostrarse como prueba de que el proceso de transformación se está realizando de manera segura y correcta (En ambientes enterprise por ejemplo, a clientes cuya información está siendo anonimizada para ser luego analizada)
En primera instancia, solo se exportará a formato ***PDF***.

10. Exportación de datos

Descripción: El usuario tendrá la opción de exportar los datos anonimizados.

Prioridad: Alta

Estímulo/Respuesta y Secuencias:

1. El usuario se encuentra dentro de un proyecto existente, y con un set de datos importados. Ya ha realizado la transformación de los datos, y está nuevamente en la pantalla principal
2. El usuario selecciona el botón ***Exportar***, y el sistema muestra una nueva ventana.
3. El usuario deberá especificar la ubicación a donde se quiere exportar los datos anonimizados, así como también el formato del archivo que los contendrá.
    a. Por defecto, el sistema mostrará el formato en el cual se encuentran contenidos los datos originales (Por ejemplo, una nueva hoja de cálculo si los datos provenían de ahí)
4. Una vez especificada la ubicación y el formato, el usuario selecciona la opción *Siguiente*. El sistema exporta los datos procesados usando el input previo del usuario.
5. Al finalizar la exportación de los datos, se vuelve a la pantalla principal.

Requerimiento funcional
Exportar datos que han sido anonimizados es otra pieza clave dentro de la plataforma a realizar; dado que el objetivo detrás de este desarrollo es lograr anonimizar datos para luego utilizarlos (con diversos fines, como puede ser el análisis de los mismos)
Para la primer versión de la plataforma, se requiere que se pueda exportar a los mismos formatos de los cuales se puede importar.

Requerimientos no funcionales del sistema

Hardware

Debido a las características del framework, se requerirá realizar un gran procesamiento de datos. Es por esto que las características de hardware donde se utilizará el sistema tienen vital importancia para un correcto desempeño. Se requiere tener al menos 6gb de memoria RAM y un procesador Intel Core i5 o de características similares (en este momento, es un valor teórico basado en el análisis e investigación de consumos, durante el transcurso del proyecto el mismo será puesto a prueba para validarlo)

Rendimiento

Debido a que se deberán procesar grandes volúmenes de datos, este apartado es crítico para el desarrollo del framework.
Es por este motivo que se analizarán distintas opciones para poder mejorar la latencia del proceso de transformación, como por ejemplo utilizando el meta-modelo de programación MapReduce[3] (el cual está asociado con el procesamiento y generación de grandes sets de datos sobre un clúster con un algoritmo de tipo paralelo y distribuido)
Así también, se analizarán alternativas para reducir el uso de memoria, como por ejemplo la utilización de generadores o cursores a bases de datos.
Así mismo, se realizará un benchmarking contra distintos tipos y tamaños de bases de datos para establecer el tiempo promedio de anonimización de datos, con el fin de tener un límite teórico para determinar si es necesario cortar con el proceso o no. Este valor puede ser guardado como defecto, siendo el usuario final el que determina si quiere cortar a determinado tiempo de ser muy larga la operación, o si desea determinar un tiempo personalizado acorde al contexto en el que está trabajando.

Extensibilidad

Una de las premisas del sistema es que el mismo brinde la capacidad de permitir en el futuro el desarrollo de nuevas funcionalidades y la extensión de funcionalidades ya existentes.

Importación y exportación de datos
Originalmente se podrá importar base de datos SQLite y archivos CSV, sin embargo se implementará una interfaz que permita, de manera transparente, agregar el soporte de otros tipos de bases de datos relacionales.

Modelos de privacidad

[3] 2011. MapReduce - Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/MapReduce.

Este es otro aspecto que será diseñado con el fin de que sea fácilmente extensible, contemplado la aparición de nuevos modelos de privacidad.
Se tendrá en cuenta además los casos en los cuales un modelo se basa en otro ya existente, reutilizando para esto los que ya se encuentran integrados al framework. Por ejemplo, l-diversity es un modelo k-anonimity con ciertos criterios de privacidad adicionales. Si se desea agregar un nuevo método basado en l-diversity, se podrá partir del modelo existente agregando solamente las nuevas funcionalidades requeridas.
De esta manera, el framework será capaz de extender los modelos de privacidad disponibles de una manera sencilla, dado que al ser los modelos desarrollados como módulos extra, se manejarán como plugins para ser integrados al framework.

Algoritmos de modelos de privacidad
Uno de los apartados más importantes de investigación dentro de la anonimización de datos es la creación de nuevos algoritmos cada vez más óptimos y performantes que satisfagan un determinado modelo. Debido a esto resulta de suma importancia brindar la posibilidad del agregado sencillo de nuevos algoritmos. Por lo tanto se tomarán decisiones de diseño que busquen favorecer este apartado.

Usabilidad

La forma en la cual el usuario final interactuará con el sistema es otro apartado muy relevante a tener en cuenta durante el diseño y el desarrollo del sistema.

Configuración de la herramienta
Los métodos de transformación de datos deberán ser orientados hacia un modelo parametrizable y usable, de manera tal que se permita al usuario elegir qué métodos aplicar y ver fácilmente el resultado obtenido, teniendo la posibilidad de cambiarlos a gusto hasta encontrar un set anonimizado que les satisfaga.
La aplicación, por lo tanto, debe ser intuitiva y de fácil uso, con el fin de que se logre aprovechar al máximo el diseño de la interfaz gráfica

Creación de jerarquías de generalización
Con el fin de que el usuario pueda crear jerarquías a la hora de definir la generalización deseada para los distintos atributos, se deberá plantear un modelo tal que el armado de las mismas pueda realizarse de forma sencilla e intuitiva.
Es por esto que se investigará la usabilidad orientada al armado de grupos (Como por ejemplo: ‘Drag and Drop’, ‘Selección por check-boxs’, ‘Usos de colores intuitivos’, entre otras)

Seguridad

La información que se manipulará dentro del proceso de anonimización puede llegar a ser crítica y muy valiosa para el usuario. Hoy en día existen estrictas leyes que regulan el manejo

de datos personales y velan por la protección de ellos. Entre estas leyes y regulaciones, las más conocidas son:

- Ley de Protección de Datos Personales[4] (***Argentina***)
- HIPPA privacy rule[5] (***Estados Unidos***)
- European data protection regulation[6] (***Unión Europea***)

Cada una de ellas tiene sus particularidades en cuanto a que se consideran datos personales, y a como se debe proteger.

El framework, dado que se encuentra en constante contacto con datos de carácter personal, privado y sensible, debería seguir los lineamientos generales de estas leyes para asegurarse que el manejo de datos es correcto.

Un ejemplo sería el ofrecer cifrado de datos para aquellos proyectos que son guardados. El usuario proveería credenciales, que serían utilizadas como autenticación al iniciar el programa, así como también serían utilizadas para cifrar y descifrar los datos originales que se guardan y cargan, al cerrar o abrir un proyecto.

Claramente, esto podría afectar el rendimiento del framework debido a que el proceso de descifrar/cifrar es algo costoso cuando se realiza con grandes volúmenes de datos; es por esto que se puede dar como posibilidad para que el usuario elija a criterio propio.

Instalabilidad

Se desea que el proceso de instalación sea lo más ameno y cercano al usuario en aspectos de usabilidad, para que no deba perder tiempo con cuestiones menores como es un proceso de instalación.

Debido a que el desarrollo del software será hecho bajo el lenguaje de programación Python, se utilizarán las buenas prácticas de paquetización que provee, de forma tal que cada componente sea un paquete y así lograr que el proceso de instalación sea simplemente seguir una serie de pasos que involucran agregar paquetes; haciendo que la plataforma sea lo más cercano a algo portable.

Mantenibilidad

Cada uno de los componentes de software que forman parte de la solución propuesta deberán estar debidamente documentados tanto en el código fuente como en los manuales de administración y de usuario. Esto facilitará no solo la extensibilidad del sistema sino cualquier modificación que desee realizar el cliente en el futuro sobre las sobre las funcionalidades ya existentes.

---

[4] "Ley de Protección de Datos Personales - Ministerio de Economía." 2005. 17 Oct. 2015 <http://infoleg.mecon.gov.ar/infolegInternet/anexos/60000-64999/64790/norma.htm>

[5] "The Privacy Rule - HHS.gov." 2009. 17 Oct. 2015 <http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/>

[6] "Protection of personal data - European Commission." 2011. 17 Oct. 2015 <http://ec.europa.eu/justice/data-protection/>

Una alta mantenibilidad también es crítica debido a que el sistema será desarrollado utilizando un ciclo de vida incremental o iterativo.

Apéndice A: Glosario

Atributos identificables	Atributos que por sí solos son capaces de identificar a un individuo
Atributos cuasi-identificables	Atributos que combinados con otros, pueden unir información externa para identificar a un individuo
Base de datos relacional	Es un tipo de base de datos que cumple con el modelo relacional. Consiste en un conjunto de datos interrelacionados almacenados en conjunto, sin redundancias innecesarias, de forma independiente de los programas que acceden a ellos.
CSV	Tipo de documento en formato abierto sencillo para representar datos en forma de tabla, en las que las columnas se separan por comas y las filas por saltos de línea
SAP	Es un sistema de información que gestiona de manera integrada, "on-line", todas las áreas funcionales de la empresa.
SQL	Es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en ellas