Uruguay Presidencia | agesic

Guía sobre Anonimización de Datos

Tecnologías de la Información

Versión 0.8

Año 2020

Tabla de Contenido

Historial de Revisiones

Fecha	Versión	Cambios	Autor
27-04-2020	0.1	Creación del documento	Fernando Telechea
20-05-2020	0.5	Definición de Marco Teórico	
10-08-2020	0.8	Herramientas de Anonimización	

Agradecimiento

Un especial agradecimiento a la empresa Tryolabs por su colaboración en temas relacionados a la emergencia sanitaria por COVID-19, particularmente a las personas que cooperaron en la elaboración de este documento: Alan Descoins, Martín Alcalá Rubí y Soledad Rivas.

Objetivo

El objetivo de este documento es generar una guía autocontenida sobre el tema de anonimización de datos, describiendo conceptos, algunas de las técnicas existentes y herramientas que se pueden utilizar cuando se requiera anonimizar datos.

Contexto

El empleo de información personal se encuentra alcanzado por normas que reconocen el derecho a la Protección de Datos Personales como un derecho fundamental.

Más allá de la necesidad que tienen determinados Organismos del Estado en analizar la información personal, ante determinadas situaciones críticas o de emergencia sanitaria, como la derivada de la pandemia del coronavirus COVID-19, para atender todo lo relacionado con la enfermedad y, especialmente mitigar sus efectos adversos sobre las personas y la sociedad, es imperativo realizar un uso responsable de dicha información, procurando, alcanzar equilibrios con otros derechos.

En este sentido, con el fin de que se puedan tomar decisiones basadas en análisis de datos, pero salvaguardando la privacidad y protección de los datos personales, es que se genera esta guía práctica en la temática de anonimización de datos.

¿Qué es la anonimización de datos?

La anonimización de datos debe considerarse como una forma de eliminar las posibilidades de identificación de las personas. El avance de la tecnología y la información disponible hacen difícil garantizar el anonimato absoluto, especialmente a lo largo del tiempo, pero, en cualquier caso, la anonimización va a ofrecer mayores garantías de privacidad a las personas.

La finalidad del proceso de anonimización es por tanto eliminar o reducir al mínimo los riesgos de reidentificación de los datos anonimizados manteniendo la veracidad de los resultados del tratamiento de los mismos. Es decir, además de evitar la identificación de las personas, los datos anonimizados deben garantizar que cualquier operación o tratamiento que pueda ser realizado con posterioridad a la anonimización no conlleve una distorsión de los datos reales. *Un análisis de los datos que pueda derivar de los datos anonimizados, no debería diferir del análisis que pudiera obtenerse si hubiera sido realizado con datos no anonimizados.*

Proceso de anonimización

En un proceso de anonimización es altamente recomendable definir un protocolo de actuación. Los pasos o etapas que se describen a continuación, sirven de orientación al momento de definir este protocolo, pero no debe tomarse como un marco o esquema conceptual cerrado, sino que refleja una propuesta de una estructura que puede ser tenida en cuenta en los procesos de anonimización.

Las etapas básicas propuestas para el proceso de anomización, son:

- ✓ Definición del equipo de trabajo
- ✓ Definición de objetivos y finalidad de la información anonimizada
- ✓ Preanonimización
- ✓ Anonimización
- ✓ Control

Figura 1 - Etapas del Procesos de Anonimización

Definición del equipo de trabajo

En el desarrollo de un proceso de anonimización intervienen distintos perfiles o roles. Algunos de los perfiles que deberían ser tenidos en cuenta en dicho proceso son los siguientes:

- Responsable de la fuente de datos inicial
- Responsables de protección de datos o referente de protección de datos personales
- Destinatario o responsable del tratamiento de la información personal anonimizada
- Equipo de evaluación de riesgos
- Equipo de preanonimización y de anonimización
- Equipo de seguridad de la información y del proceso de anonimización

Es recomendable que se garantice, en la medida de lo posible, que cada uno de los actores actuara en el ámbito de su competencia, con independencia del resto de los actores, con el fin de evitar que un error que se produzca en un determinado nivel sea supervisado y aprobado en otro nivel distinto por la misma persona.

Definición de objetivos y finalidad de la información anonimizada

Además de garantizar la privacidad de los interesados, el responsable de la fuente de datos determinará los objetivos que deberá cumplir la información anonimizada en función de los intereses de su destinatario. El diseño del proceso de anonimización estará condicionado por el objetivo final de la información anonimizada dando lugar a información de uso restringido o a datos abiertos.

Cuando la información anonimizada tiene por finalidad convertirse en información de uso restringido, la privacidad de los datos personales puede ser reforzada mediante acuerdos de confidencialidad que formarán parte del conjunto de las garantías jurídicas del proceso de anonimización.

Preanonimización

En la etapa de preanonimización se diseña el proyecto de anonimización, en el que se deberán identificar, con claridad las variables, los identificadores directos e indirectos, los datos confidenciales, cuál o cuáles serán las técnicas adecuadas de anonimización, según el conjunto de datos de que se trate, el riesgo de reidentificación asociado, finalizando con la ejecución del proyecto. Todo ello deberá permitir la ruptura de los eslabones de la relación de identificación-información.

Al mismo tiempo, no todos los actores del proceso tienen por qué acceder a toda la información en todo momento, por lo que es de suma importancia que el responsable de la tarea proporcione solamente la mínima información necesaria a ser utilizada y en consecuencia anonimizada.

Se deberá:

- Clasificar las variables, por ejemplo variable de identificación geográfica, de identificación directa de personas o empresas, de carácter sensible o confidencial, sin restricción para el acceso público.
- Identificar la información que será incluida en el proceso de anonimización. El objetivo es reducir al mínimo necesario la cantidad de variables que permitan la identificación de las personas, restringiendo el acceso a la información confidencial al equipo de trabajo implicado en el proceso y optimizando el coste computacional de las operaciones con datos anonimizados.

Anonimización

- La anonimización posee diversas características que deben ser tomadas en cuenta cada vez que se va a materializar, para que los datos personales puedan ser utilizados en forma abierta:
- No puede establecerse vínculo alguno entre el dato y su titular sin un esfuerzo desproporcionado
- No puede ser reversible, es decir que es el resultado de un tratamiento de datos personales realizado para impedir que se vuelva atrás con lo efectuado y se identifique al interesado
- Que en la práctica sea equivalente al de un borrado permanente
- Que lleve implícito un factor de riesgo que se debe tener en cuenta al valorar las técnicas de anonimización, además de considerarse la gravedad y probabilidad del riesgo en sí mismo
- Siguiendo el plan elaborado por el responsable del proceso de anonimización durante la etapa de preanonimización, los técnicos deberán aplicar las técnicas seleccionadas, los algoritmos necesarios, realizar pruebas de calidad y entregar el resultado al responsable para su aprobación.
- El objetivo final de la anonimización es proveer los datos desagregados para ser utilizados, sin generar conflictos con los titulares de los datos.

Control

Esta etapa implica la realización de controles periódicos por parte de los técnicos en virtud de la aparición de las nuevas tecnologías y métodos para prevenir y evitar los posibles riesgos de reidentificación. Dependiendo del resultado de esta etapa, es posible volver a replantearse distintos escenarios para la etapa de preanonimización y anonimización, para luego volver a realizar un nuevo control.

Marco Teórico sobre Anonimización

Para poder explicar las distintas técnicas de anonimización, es necesario previamente dar una serie de definiciones y conceptos:

Liberación de datos

Al momento de liberar un conjunto de datos se debe determinar el formato que se desea utilizar. Dentro de los formatos más utilizados se encuentran los Microdatos, los datos tabulares y las bases de datos interactivas. ▪⌕

Microdatos: cada registro perteneciente al conjunto de datos contiene información relacionada a un individuo especifico. ▪⌕

Datos tabulares: el conjunto de datos muestra valores agregados para distintos grupos de individuos. Este formato es utilizado comúnmente para presentar el resultado de estadísticas oficiales. ▪⌕

Bases de datos interactivas: los datos no son liberados directamente, sino que se brinda una interfaz que los usuarios pueden utilizar para enviar consultas estadísticas como sumas y promedios.

Las liberaciones del tipo Microdatos son las que aportan una mayor cantidad de información y, por lo tanto, las más complejas de anonimizar. De aquí en más se considera un conjunto de datos en el formato de Microdatos, el cual contiene registros, donde cada uno contiene la información de un individuo concreto. A su vez, cada registro está formado por un conjunto de atributos, donde cada atributo brinda una característica del individuo, por ejemplo, C.I., nombre, apellido, estado civil, etc.

Cada atributo tiene asociado un tipo de dato, donde estos tipos pueden ser:

- **Continuo:** Valores numéricos continuos.
- **Categórico:** Valores pertenecientes a un dominio finito.

Los atributos se clasifican en las siguientes categorías no exclusivas. ▪⌕

1. **Identificadores**: Un atributo se considera un Identificador si permite la identificación de un individuo de forma inequívoca. Por ejemplo, la Cédula de Identidad de una persona es un identificador. Este tipo de atributo debe ser removido durante el proceso de anonimización. ▪⌕
2. **Quasi-identificadores (QI):** Los QI son atributos que por si solos no permiten reidentificar a un individuo, pero que la combinación de varios QI podría llevar a la identificación inequívoca de algunos individuos. Como posibles ejemplos de este tipo de atributo podemos considerar la fecha de nacimiento, el sexo y el estado civil de un individuo.
3. **Confidenciales:** Contienen información sensible de los individuos del conjunto de datos. Posibles ejemplos son la información salarial o el estado de salud. El objetivo principal de las técnicas de protección de

Microdatos es evitar que un atacante obtenga información confidencial de un individuo específico y posibles inferencias sobre el atributo.

4. **No confidenciales:** Es un atributo que no pertenece a ninguna de las categorías anteriores. En un proceso de anonimización se consideran dos conjuntos de datos. Por un lado se tiene el conjunto de datos original, el cual refiere al conjunto de datos inicial que contiene la información completa de los individuos. Por otro lado, se tiene el conjunto de datos anonimizado, que es el conjunto resultado de aplicar una o varias técnicas de anonimización al conjunto de datos original. El conjunto de datos que es liberado o compartido es el conjunto anonimizado.

Riesgos de divulgación

Cuando un conjunto de datos se comparte se somete principalmente a dos tipos de riesgos: el de divulgación de identidad y el de divulgación de atributos.

La **divulgación de identidad** refiere a que un atacante es capaz de asociar un registro perteneciente al conjunto anonimizado con uno del conjunto original. Permitiendo así reidentificar un individuo del conjunto.

Por otro lado, la **divulgación de atributos** alude a que un atacante es capaz de determinar el valor de un atributo confidencial de un individuo con un gran nivel de seguridad.

Preservar la privacidad de un conjunto de datos refiere a mitigar los riesgos mencionados, aplicando distintas técnicas de anonimización de datos.

Técnicas de Anonimización

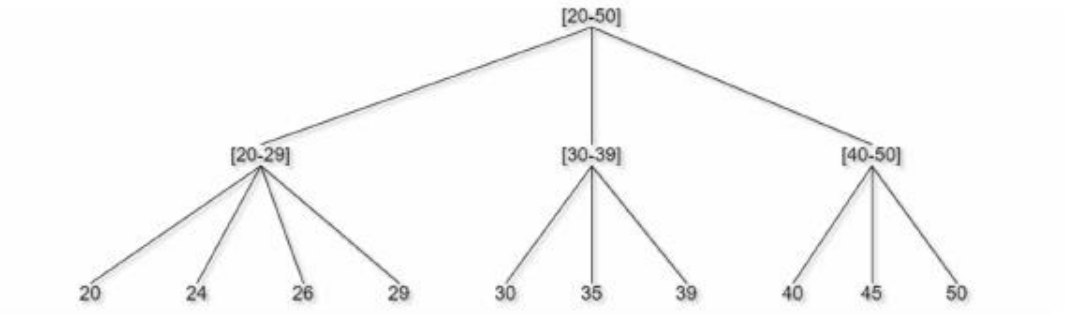
Las **técnicas de anonimización** de datos se dividen en dos grandes familias: las técnicas sin perturbación y las técnicas con perturbación.

Las técnicas sin perturbación mitigan los riesgos de divulgación suprimiendo ciertos valores y/o reduciendo el nivel de detalle de los datos del conjunto. Algunas técnicas pertenecientes a esta familia son la Generalización, la Codificación superior e inferior y la Supresión Local.

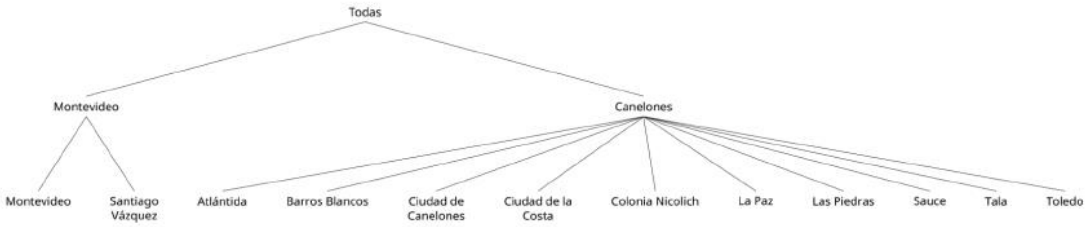
Por otro lado, las **técnicas con perturbación** en lugar de suprimir valores del conjunto de datos, alteran la distribución estadística de los mismos. Algunas técnicas de esta familia son la Adición de ruido y la Microagregación. A continuación de describen algunas de las técnicas mencionadas.

1. Generalización

La técnica de Generalización logra la anonimización de un conjunto de datos disminuyendo el nivel de detalle de los valores de los atributos. La técnica contempla tanto atributos continuos como categóricos. Para los atributos continuos se generan intervalos de valores y para los atributos categóricos se crean nuevas categorías que contienen a otras más específicas. Para disminuir el nivel de detalle, está técnica hace uso de una estructura denominada Jerarquía, la cual contiene los valores que puede presentar un atributo. Debe haber una jerarquía para cada atributo de tipo quasi-identificador. Ejemplos de jerarquías se describen a continuación. Para el quasi-identificador Edad:



Para el quasi-identificador Localidad:



2. Codificación superior e inferior

Es un caso particular de la técnica anterior. Esta técnica determina un valor de cada quasi-identificador y expresa el resto de los valores del conjunto de datos cómo valores mayores o menores al valor seleccionado.

Por ejemplo, si el valor seleccionado para el atributo Edad es 35, los valores del conjunto de datos anonimizado para este atributo podrán ser >=35 o <35.

Esta técnica solo puede ser aplicada en sobre atributos que presenten un orden total, por ejemplo, atributos continuos o categóricos que cumplan esta restricción.

3. Supresión local

Es una técnica basada en eliminar algunos valores del conjunto de datos. Los valores suprimidos son aquellos que tienen pocas apariciones dentro del conjunto original dado que representan un gran riesgo de divulgación para los registros que lo contienen.

4. Microagregación

Es una técnica aplicable exclusivamente a atributos numéricos. La idea detrás de esta técnica es que la privacidad de un conjunto de datos se preserva si los datos publicados contienen grupos de registros que contengan al menos cierta cardinalidad, definida por el usuario. Los valores de los atributos que pertenecen a un grupo son sustituidos por otro valor, por ejemplo, el valor medio del atributo dentro del grupo.

Modelos de Privacidad

Al momento de seleccionar uno o varios modelos de privacidad se debe seleccionar, por lo menos, un modelo que mitigue el riesgo de divulgación de identidad y otro que mitigue la divulgación de atributos. A continuación se describen algunos de los modelos existentes.

1. K-anonymity

anonymity es un modelo de privacidad que ataca el riesgo de divulgación de identidad. La idea detrás de este modelo se basa en generar grupos dentro del conjunto de datos, donde cada grupo contiene al menos k registros. Los registros de un mismo grupo poseen los mismos valores en todos los quasi-identificadores, generando que los registros de un grupo sean indistinguibles unos de otros y, que la probabilidad de re-identificación de un registro en el conjunto anonimizado, sea menor o igual a 1/k.

El desafío más grande que presenta este modelo es la elección del parámetro k ya que no se tienen heurísticas para determinar su valor en un contexto dado.

Este modelo puede ser implementado mediante varias técnicas, por ejemplo, generalización y microagregación.

2. K-Map

Map es un modelo de privacidad muy relacionado con k-anonymity donde la diferencia principal radica en que k-Map es capaz de considerar otro conjunto de datos, además del conjunto original, para satisfacer la condición de privacidad.

El desafío planteado para la aplicación de este modelo no solo radica en la elección del parámetro k sino también en proveer el conjunto de datos extendido. Además, asume que un atacante externo no tiene más información sobre el conjunto de datos que la provista por el conjunto anonimizado, condición que no puede ser asumida en la mayoría de los contextos.

Un ejemplo de aplicación se puede encontrar en https://desfontain.es/privacy/k-map.html.

3. ℓ-Diversity

Modelo de privacidad que ataca el riesgo de divulgación de atributos asegurándose que cada atributo confidencial tenga, al menos, ℓ representantes en cada clase de equivalencia del conjunto anonimizado. El desafío del modelo radica en la elección del parámetro ℓ.

4. t-Closeness

Al igual que el modelo anterior, este modelo mitiga el riesgo de divulgación de atributos. Requiere que las distribuciones de valores de un atributo confidencial, dentro de cada clase de equivalencia, disten menos de t respecto a la distribución de los valores del atributo en el conjunto original.

El desafío de aplicar el modelo es definir la noción de distancia que se desea utilizar y seleccionar el valor del parámetro t.

Otros modelos de privacidad son Average risk, Population uniqueness, Sample uniqueness, δ-Disclosure privacy, β-Likeness, δ-Presence, Profitability y Differential privacy. Este último modelo es comúnmente utilizado cuando se cuenta con una base de datos interactiva. Vale la pena mencionar que la mayoría de los modelos de privacidad presentan variantes para aplicaciones específicas a contextos puntuales o a determinadas características del conjunto de datos original.

Herramientas

A continuación se describen algunas herramientas que permiten proteger datos privados, no necesariamente son herramientas de anonimización, pueden ser herramientas de pseudoanonimización o enmascaramiento de datos (data masking). El criterio para definir esta lista de herramientas, fue tener en cuenta las que son provistas por los fabricantes de software de bases de datos, por lo que seguramente ya puede contar con acceso a alguna de ellas. Además se establecen algunas herramientas del tipo open source, ya que cualquier organización puede instalarlas y utilizarlas libremente.

Herramientas asociadas a bases de datos

En algunas ocasiones estas herramientas se proveen como funcionalidad del software de base de datos y en otras son componentes que se licencian de forma independiente. En ambos casos, el proceso es relativamente sencillo y los principales casos de uso son:

Generar ambientes de testeo a partir de los de producción salvaguardando la información privada de las personas.

Generar ambientes o exportación de datos con el objetivo de generar análisis de datos, sin datos privados pero asegurándonos de obtener los mismos resultados estadísticos que se lograrían con los datos originales.

Algunas de estas herramientas son:

Microsoft SQL Server Dynamic Data Masking: es una funcionalidad que se provee con el software de base de datos, (disponible desde la versión 2016 de Microsoft SQL Server)

Oracle Data Masking and Subsetting: funcionalidad provista para la versión de Oracle Enterprise Edition

IBM InfoSphere Optim Data Masking

Herramienta de anonimización - ARX

Esta herramienta open source (del tipo desktop) de la cual daremos un detalle a continuación, es particularmente interesante porque además de incorporar las características del marco teórico descripto en esta guía, provee un SDK para Java (“Software Development Kit” - Kit de desarrollo de software), que puede ser utilizado para invocar de forma programática, toda la funcionalidad del desktop. De esta forma se pueden generar procesos personalizados de anonimización, invocando a métodos de las clases expuestas por dicho SDK.

Algunas herramientas que permiten generar workflows de datos de forma gráfica (como por ejemplo KNIME), se valen de este SDK para generar nodos que ejecutan tareas asociadas al proceso de anonimización (ver siguiente figura).

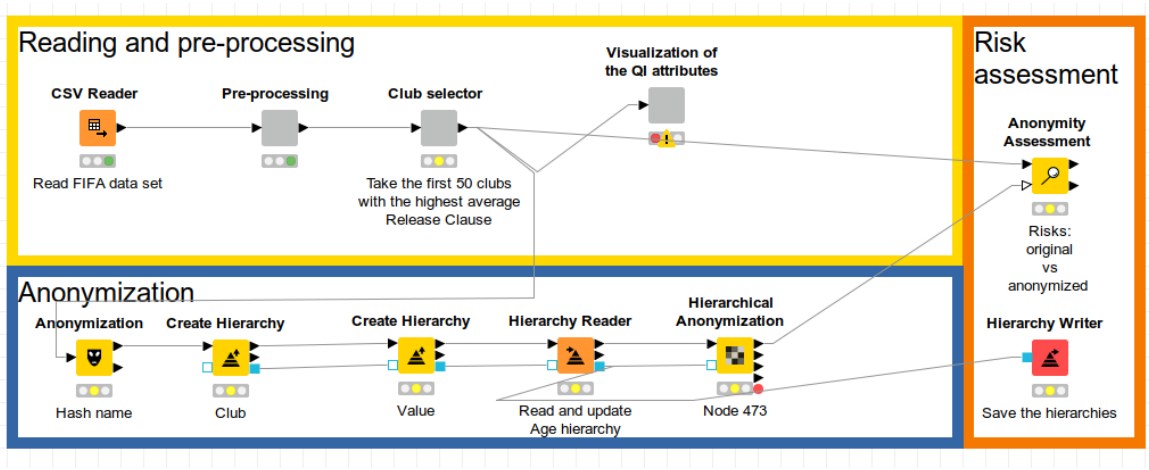
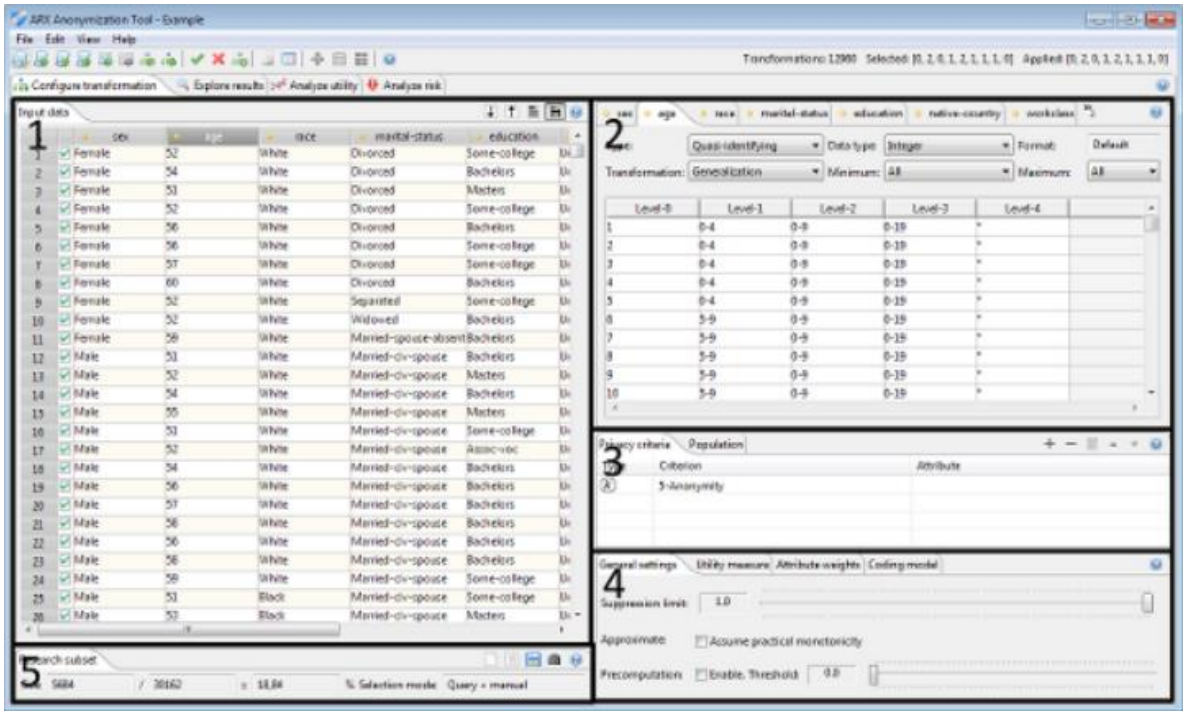


Figura 2 - Flujo de KNIME con nodos de anonimización

Consola de ARX

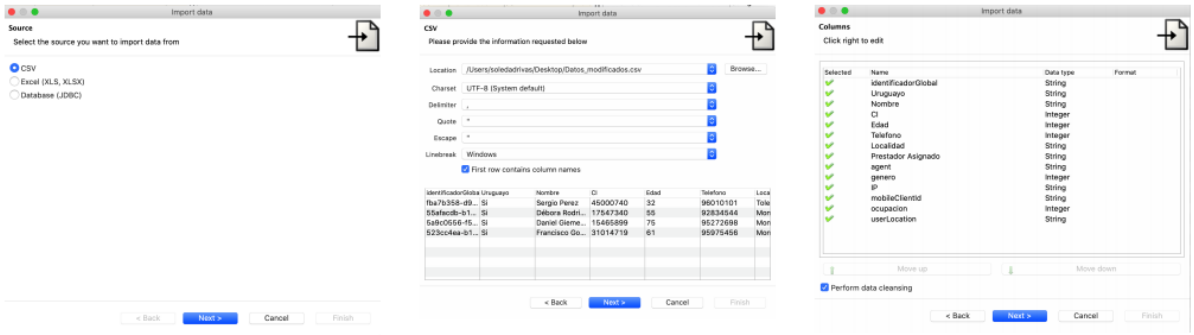


La consola de ARX se divide en 5 áreas:

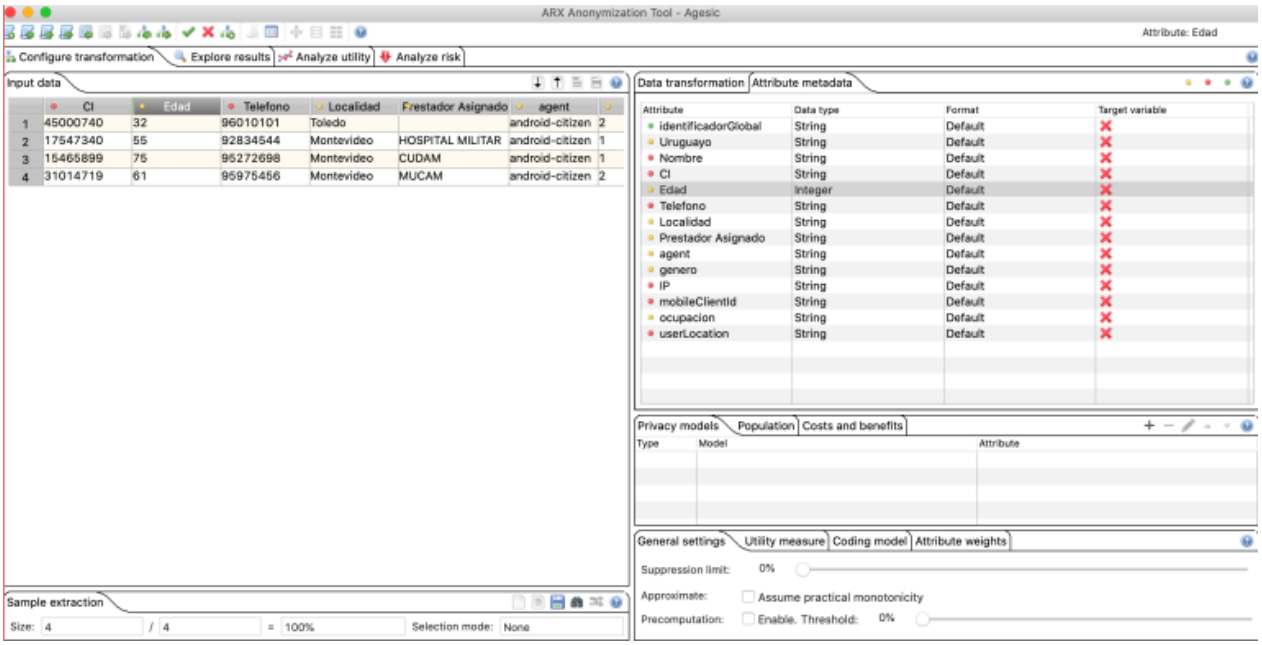
1. Permite visualizar el conjunto de datos original. Los formatos de datos aceptados por ARX para subir un conjunto de datos son: CSV, TSV, Excel ✓XLS, XLSX★, base de datos JDBC o deid (en caso de ser un proyecto que previamente fue guardado con la herramienta).
2. Permite especificar metadatos de los atributos (categorización de cada atributo) y visualizar las jerarquías definidas.
3. Configuración de los modelos de privacidad (se puede especificar más de uno) y de sus hiperparámetros.
4. Configuración de las medidas de utilidad.
5. Métodos de extracción disponibles. Por ejemplo, se puede cargar un conjunto de datos que sea un ejemplo de cómo debería quedar el conjunto de datos original luego de la anonimización

Ejemplo de uso con ARX

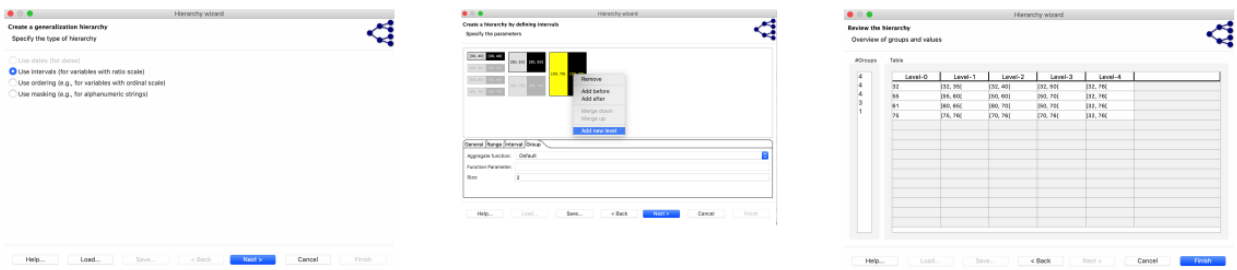
1. Crear un nuevo proyecto en ARX con la opción New Project.
2. Cargar el conjunto de datos con la opción Import Data.



3. Categorizar los atributos del conjunto en Identificadores en Identificadores, Quasi-identificadores, Confidenciales o Sensibles y No confidenciales.



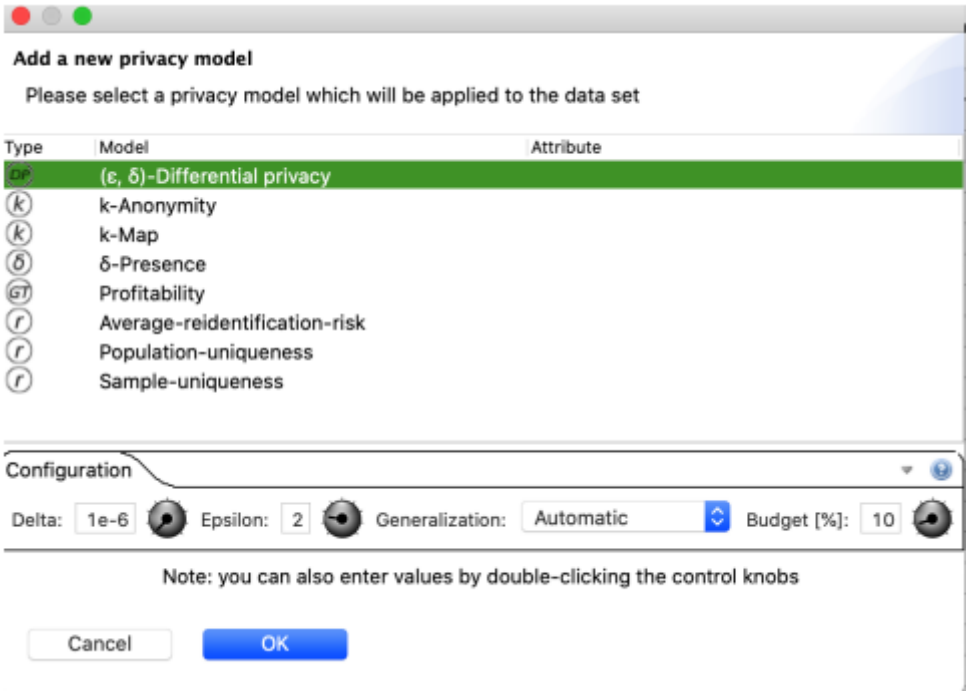
4. Una vez que los atributos fueron categorizados es necesario crear una jerarquía para cada quasi-identificador con la opción Create hierarchy. En el ejemplo, se crea la jerarquía para el atributo Edad.



Para los atributos numéricos se usan comúnmente jerarquías basadas en intervalos y para los atributos categóricos se usan jerarquías basadas en orden.

Nota: Es recomendable guardar cada jerarquía una vez definida la misma. Las jerarquías pueden ser reutilizadas en otros proyectos o ser cargadas al mismo en caso de ser necesario.

5. Luego de creadas las jerarquías se debe seleccionar uno o varios modelos de privacidad con la opción Add privacy model. Esto abrirá un dialogo donde se despliegan las opciones de modelos disponibles. Por un lado se tienen los modelos que se focalizan en el riesgo de divulgación de identidad (haciendo hincapié en los quasi-identificadores), algunos de estos modelos son: k-Anonymity, k-Map, δ-presence, differential privacy y gametheoretic. Mientras que por otro lado se tienen los modelos que atacan la divulgación de atributos (haciendo, si los hay, hincapié en los atributos confidenciales), estos son: l-diversity, t-closeness, β-likeness y δ-disclosure privacy.



Se debe tener presente que no todos los modelos están disponibles para todos los conjuntos de datos ya que la aplicación de algunos de ellos se habilita, por ejemplo, si existen atributos sensibles o si una jerarquía fue creada para cada quasi-identificador. Además, el resultado de aplicar un modelo de anonimización depende directamente de la calidad de los datos del conjunto original.

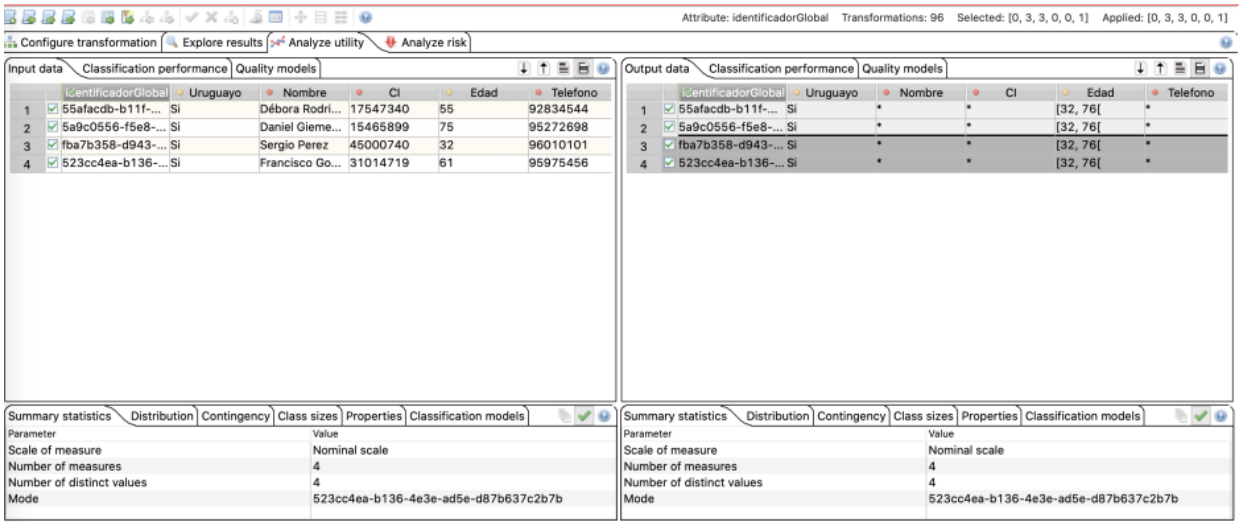
Por ejemplo, consideremos el siguiente **conjunto de datos original**:

identificadorGlobal	Uruguayo	Nombre	CI	Edad	Telefono	Localidad	agent	genero	IP
fba7b358-d943-46c5-9edc-58cad5f25354	Si	Sergio Perez	45000740	32	96010101	Toledo	android-citizen	2	167.58.111.11
55afacdb-b11f-4589-bf71-1cfce59e3134	Si	Débora Rodriguez	17547340	55	92834544	Montevideo	android-citizen	1	152.156.211.80
5a9c0556-f5e8-4eeb-9663-75c231eebadc	Si	Daniel Giemenez	15465899	75	95272698	Montevideo	android-citizen	1	186.52.67.41
523cc4ea-b136-4e3e-ad5e-d87b637c2b7b	Si	Francisco Gonzalez	31014719	61	95975456	Montevideo	android-citizen	2	167.58.249.94

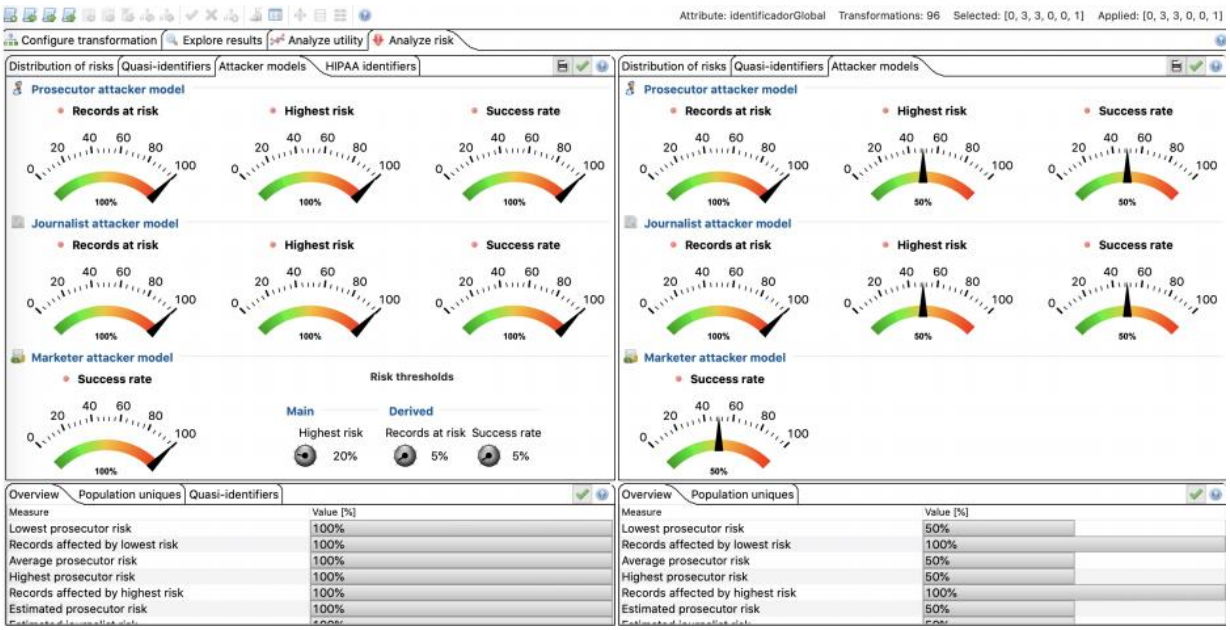
Habiendo definido previamente las jerarquías para todos los quasi-identificadores ✔Edad, Localidad, Género, Agent y Uruguayo), el resultado de aplicar el modelo de k-anonymity con k=2 para el conjunto de la tabla anterior es:

identificadorGlobal	Uruguayo	Nombre	CI	Edad	Telefono	Localidad	agent	genero	IP
fba7b358-d943-46c5-9edc-58cad5f25354	Si	*	*	[32, 76]	*	Todas	android-citizen	2	*
55afacdb-b11f-4589-bf71-1cfce59e3134	Si	*	*	[32, 76]	*	Todas	android-citizen	1	*
5a9c0556-f5e8-4eeb-9663-75c231eebadc	Si	*	*	[32, 76]	*	Todas	android-citizen	1	*
523cc4ea-b136-4e3e-ad5e-d87b637c2b7b	Si	*	*	[32, 76]	*	Todas	android-citizen	2	*

6. Una vez ejecutado el/los modelo/s de privacidad se deben evaluar dos aspectos del conjunto anonimizado: la utilidad de los datos y los riesgos de divulgación. ARX posee un análisis automático de ambos bajo las opciones Analyze Utility y Analyze Risk, en ambos casos, la pantalla desplegada muestra a la izquierda los resultados del conjunto original y a la derecha los del conjunto anonimizado.
   En el análisis de utilidad se consideran aspectos estadísticos del conjunto anonimizado y se comparan entre ambos conjuntos.



En el análisis de riesgo se simulan distintos modelos de ataques conocidos y se comparan los resultados entre los conjuntos.



Herramienta de anonimización - Amnesia

Amnesia es una herramienta open source de anonimización de datos, que permite eliminar la información asociada a los identificadores directos como nombres o números de documentos identificativos, además de que también transforma los atributos cuasi-identificadores como la fecha de nacimiento y el código postal para mitigar los riesgos de reidentificación de los sujetos que figuran en los orígenes de datos, utilizando para esto el método de k-anonimización. Dispone de una versión cliente y de una versión online.

La versión online es para probar las funcionalidades de la herramienta, por lo que no deberá subir datos verdaderos que desea anonimizar, sino datos de prueba. Luego de cargar el dataset, se presentará un wizard en el que cada paso deberá ir brindando la información necesaria para luego ejecutar la anonimización.

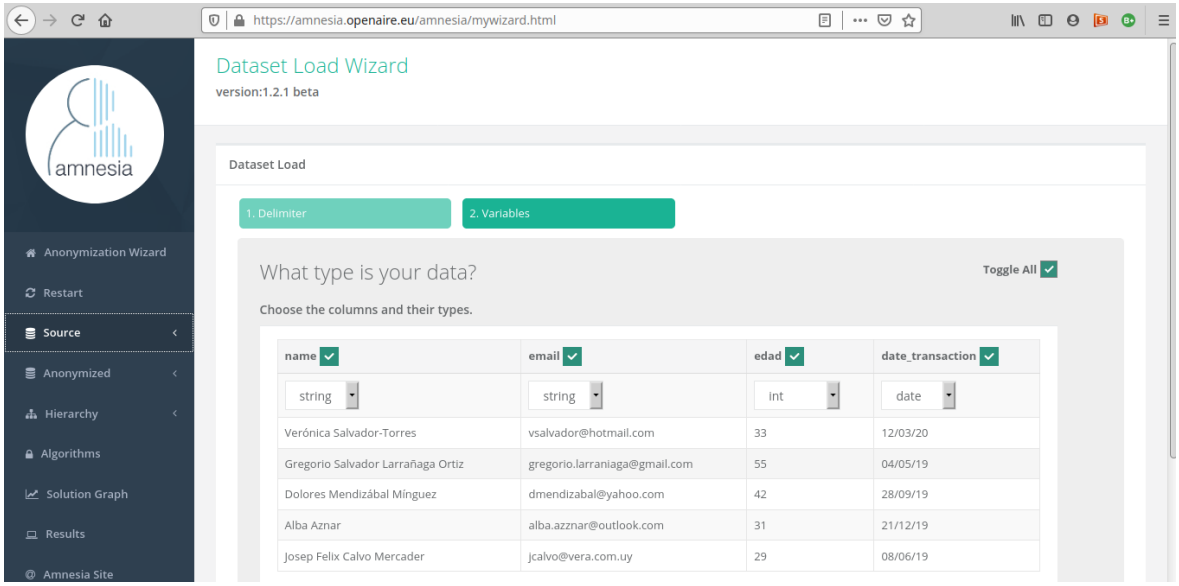


Tabla de referencias sobre material recopilado

Referencia	Link	Organización
Guía con criterios de disociación	https://www.gub.uy/unidad-reguladora-control-datos-personales/comunicacion/publicaciones/guia-criterios-de-disociacion-de-datos-personales	URCDP
Recomendaciones para el tratamiento de datos personales ante la situación de emergencia sanitaria nacional	https://www.gub.uy/unidad-reguladora-control-datos-personales/comunicacion/publicaciones/recomendaciones-para-tratamiento-datos-personales-ante-situacion	URCDP
Guía de disociación y anonimización de datos personales en el ámbito de la salud	https://centrodeconocimiento.agesic.gub.uy/web/salud.uy/gu%C3%ADas/-/document_library/GWvfsJCm0Iij/view_file/600037	Agesic - Salud.uy
Personal data anonymization: key concepts & how it affects machine learning models	https://tryolabs.com/blog/2020/06/11/personal-data-anonymization-key-concepts--how-it-affects-machine-learning-models/	Tryolabs
Josep Domingo-Ferrer; David Sanchez; Jordi Soria-Comas, "Database Anonymization	https://ieeexplore.ieee.org/document/7385400	IEEE
Orientaciones y garantías en los procedimientos de anonimización de datos personales	https://joinup.ec.europa.eu/collection/open-source-observatory-osor/news/spain-publishes-two-guides https://datos.gob.es/sites/default/files/doc/file/orientaciones_y_garantias_anonimizacion.pdf	Comisión Europea Joinup