Proteger los datos Es proteger a la gente

GUÍAS DE PROTECCIÓN DE DATOS Y PRIVACIDAD

MERCY CORPS

Desidentificación de los datos

Esta guía ofrece un ejemplo de eliminación de información de identificación personal (PII) de un conjunto de datos. Hay varias formas de «desidentificar» los datos, lo que hace referencia a las actividades o métodos de tratamiento que funcionan para evitar que se revele la identidad de la persona registrada. Dos tipos comunes de desidentificación son la «anonimización» y la «seudonimización».

La anonimización es el proceso por el que los datos personales se convierten en anónimos, de modo que una persona (o «persona registrada») deja de ser identificable: es un cambio permanente de los datos. Los métodos más comunes incluyen la eliminación de la información de identificación personal o la codificación de los valores de ciertos conjuntos de PII.

> Ejemplo: imagine que una organización tiene datos de encuesta que tienen campos para el nombre, el número de identificación nacional, el nombre de la aldea, la afiliación étnica, la edad, el nivel educativo y los indicadores de salud. En este caso, eliminar el nombre y el número de identificación nacional sería el primer paso para anonimizar los datos, ya que estos «atributos directos» son datos personales que identifican directamente a un individuo. Se mantendrían los «atributos indirectos» del nombre del pueblo, la afiliación étnica, la edad, el nivel educativo y los indicadores de salud.

Sin embargo, aunque algunos atributos parezcan «anónimos» pueden no serlo. Si la encuesta se realizó en una aldea muy pequeña en el que solo dos residentes se identifican como de una determinada filiación étnica, y cada uno de ellos tiene una edad diferente, ¡el uso de esos dos atributos indirectos podría permitir identificar a esas personas! El proceso por el que se examinan todos los atributos para reducir el riesgo de reidentificación de una persona registrada se denomina Control de divulgación estadística. El primer paso en este proceso es una evaluación del riesgo de divulgación y el Centro de Datos Humanitarios tiene un tutorial en línea para realizar una evaluación del riesgo de divulgación.

La seudonimización, por su parte, describe el tratamiento de datos personales de manera que estos ya no puedan atribuirse a una persona registrada específica sin el uso de información adicional, como un código clave.

> Ejemplo: imagine que una encuesta tiene su nombre, dirección de correo electrónico, edad, nacionalidad y lugar de trabajo. La seudonimización toma los datos que son identificables específicamente sobre usted (su nombre, dirección de correo electrónico, edad) y los hace inaccesibles y los separa de los datos no identificables, como su nacionalidad. Los datos seudónimos se pueden volver a reunir en algún momento para que toda la información se pueda vincular a una fuente o persona concreta. Por ello, la seudonimización requiere que la información adicional se conserve por separado y esté sujeta a medidas técnicas y organizativas que garanticen que los datos personales no se atribuyan a la persona registrada.

¿Debe elegir la anonimización o la seudonimización?

La anonimización será generalmente más segura y reducirá el riesgo de exponer la PII. Sin embargo, a veces esto puede hacer que los datos sean demasiado generales, lo que puede hacer que no sean útiles para

programas de asistencia con vales de dinero. En el caso de los programas de salud que implican vacunas u otros tratamientos, contactar a las personas para el seguimiento del tratamiento puede ser importante. En ambos casos, la seudonimización sería la mejor opción, ya que siempre se pueden recomponer los datos para identificar a un individuo cuando sea necesario.

No hay una respuesta única y correcta sobre cuándo elegir un método en lugar de otro, y es importante entender el motivo por el que se recolectaron los datos, los riesgos potenciales asociados a la conservación de esos datos y las necesidades del programa, antes de elegir cómo desidentificar sus datos.

También es importante entender que las técnicas usadas tanto para anonimizar los datos como para hackearlos son cada vez más sofisticadas y que incluso los datos desidentificados no son siempre cien por ciento seguros. En caso de duda, contacte a su equipo de datos o de TI para obtener ayuda.

☆ Importance

Las recientes vulneraciones de datos en el Comité Internacional de la Cruz Roja, los hackeos del correo electrónico en la Agencia de los Estados Unidos para el Desarrollo Internacional, y el intercambio inadecuado de datos por parte del Alto Comisionado para los Refugiados de las Naciones Unidas muestra todas las formas en que los datos humanitarios están en riesgo. Los datos procedentes de las encuestas de hogares, las evaluaciones de necesidades y otras formas de microdatos constituyen un volumen de datos cada vez más importante en el sector humanitario. Este tipo de datos es fundamental para determinar las necesidades y perspectivas de los participantes del programa y de las comunidades en las que trabajamos, pero estos datos también presentan riesgos. La compresión sobre la evaluación y la gestión de la confidencialidad de estos datos es esencial para garantizar que se usen de forma segura, ética y eficaz en diferentes contextos de respuesta.

Algunas de las ventajas de usar datos anonimizados frente a los datos personales son:

- protección contra la divulgación indebida de datos personales;
- se aplican menos restricciones legales a los datos anonimizados; y
- permite a las organizaciones crear datos abiertos o de acceso público sin dejar de cumplir sus obligaciones de protección de datos.

Principios

La desidentificación de los datos forma parte del tratamiento de datos, y el tratamiento de datos personales realizado por las organizaciones humanitarias debe cumplir los siguientes principios.

- **Imparcialidad y legalidad del tratamiento:** los métodos deben cumplir con la legislación o las políticas regionales, nacionales o locales que pueden limitar los datos que se pueden desidentificar y el uso de determinadas tecnologías. Todo tratamiento de datos personales debe ser transparente para las personas registradas.
- **Limitación de la finalidad:** las organizaciones humanitarias deben determinar y establecer los fines específicos para los que se tratan los datos. Estos fines deben ser explícitos y legítimos.
- **Proporcionalidad:** garantiza que cada actividad concreta relacionada con el tratamiento de datos personales sea la adecuada para el objetivo declarado. Por ejemplo: ¿solo se recopila la cantidad mínima de datos necesaria? ¿Existen medidas técnicas y organizativas adecuadas para reducir los riesgos asociados al tratamiento de datos?

- **La tecnología cambia:** los nuevos conjuntos de datos y las nuevas herramientas para analizarlos cambian y avanzan rápidamente, al igual que los medios por los que se hackean o roban los datos. Es importante comprender los riesgos nuevos y emergentes para sus datos y seguir ajustando sus métodos y prácticas en consecuencia.

Seudonimización

Este es un ejemplo de una forma de desidentificar datos en una hoja de cálculo. Hay una gran variedad de métodos de desidentificación y este ejemplo usa un «código clave» para eliminar la información de identificación personal que se encuentra en los identificadores directos y guardarla en un archivo separado. La información de identificación personal (PII) es información que se puede usar para identificar a un individuo. Los ejemplos más comunes son el nombre, dirección, número de teléfono, fecha de nacimiento y número de seguro social o de identificación nacional.

Instrucciones

Puede seguir estas instrucciones de seudonimización para ver un ejemplo básico de seudonimización de un conjunto de datos. El ejercicio usa un conjunto de datos de muestra que se encuentra en la carpeta de datos de la guía en línea.

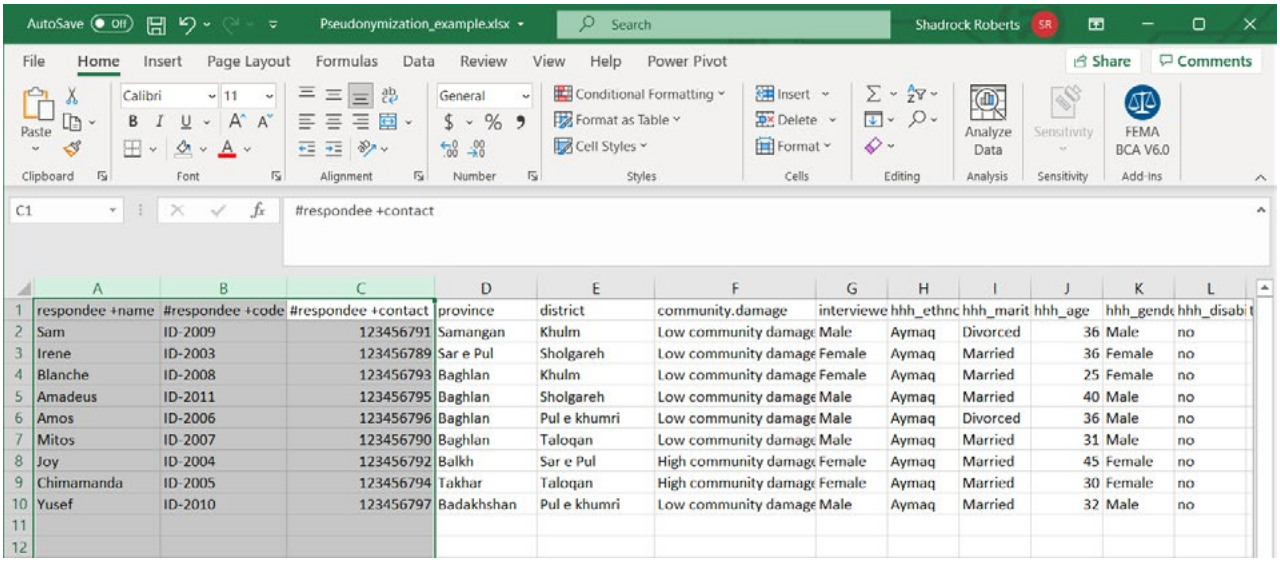
Una vez que terminado la seudonimización de los datos de muestra, se puede continuar con el tutorial del Centro de Datos Humanitarios para realizar una evaluación del riesgo de divulgación.

Paso 1: identificar la PII

Comience por identificar la PII en los datos. Lo ideal es que disponga de metadatos (datos o un documento que defina sus datos) para ayudarle a comprender los campos que tienen PII. En los datos de la muestra, hay tres columnas que tienen posible PII:

- **# de respondedor +nombre** parece incluir un nombre.
- **# de respondedor + código** probablemente tiene un número de identificación de algún tipo.
- **# de respondedor + contacto** posiblemente tiene un número de teléfono móvil

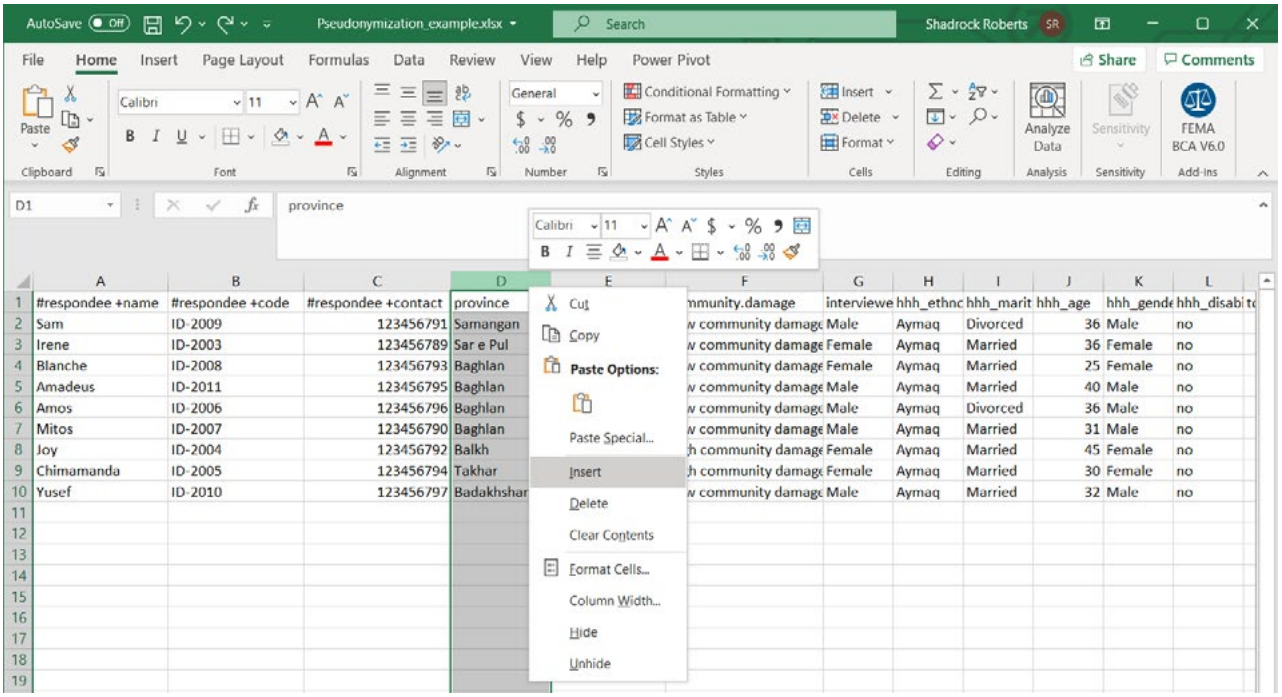
Cada uno de estos identificadores directos usa el Lenguaje de Intercambio Humanitario para etiquetar los datos.



	A	B	C	D	E	F	G	H	I	J	K	L
1	respondee +name	#respondee +code	#respondee +contact	province	district	community.damage	interviewe	hhh_ethnc	hhh_marit	hhh_age	hhh_gende	hhh_disabi
2	Sam	ID-2009	123456791	Samangan	Khulm	Low community damage	Male	Aymaq	Divorced	36	Male	no
3	Irene	ID-2003	123456789	Sar e Pul	Sholgareh	Low community damage	Female	Aymaq	Married	36	Female	no
4	Blanche	ID-2008	123456793	Baghlan	Khulm	Low community damage	Female	Aymaq	Married	25	Female	no
5	Amadeus	ID-2011	123456795	Baghlan	Sholgareh	Low community damage	Male	Aymaq	Married	40	Male	no
6	Amos	ID-2006	123456796	Baghlan	Pul e khumri	Low community damage	Male	Aymaq	Divorced	36	Male	no
7	Mitos	ID-2007	123456790	Baghlan	Taloqan	Low community damage	Male	Aymaq	Married	31	Male	no
8	Joy	ID-2004	123456792	Balkh	Sar e Pul	High community damage	Female	Aymaq	Married	45	Female	no
9	Chimamanda	ID-2005	123456794	Takhar	Taloqan	High community damage	Female	Aymaq	Married	30	Female	no
10	Yusef	ID-2010	123456797	Badakhshan	Pul e khumri	Low community damage	Male	Aymaq	Married	32	Male	no
11												
12												

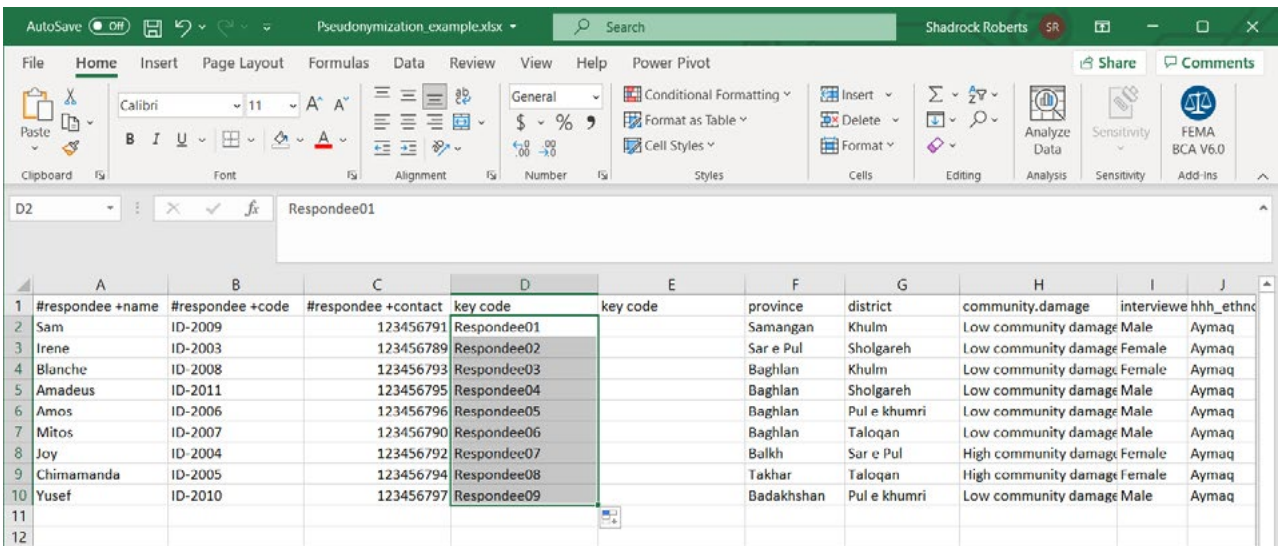
Paso 2: crear nuevas columnas para el código clave

Usaremos un código clave, un valor que generamos, para separar la PII. Como los identificadores directos están todos agrupados, crearemos dos nuevas columnas entre las columnas C, **# de respondedor +contacto** y la columna D, **provincia**. En Excel, lo hacemos al resaltar una columna a la derecha de donde queremos insertar nuevas columnas, hacemos clic con el botón derecho en la columna y seleccionamos **Insertar**. Vuelva a repetir este proceso para crear otra columna vacía.



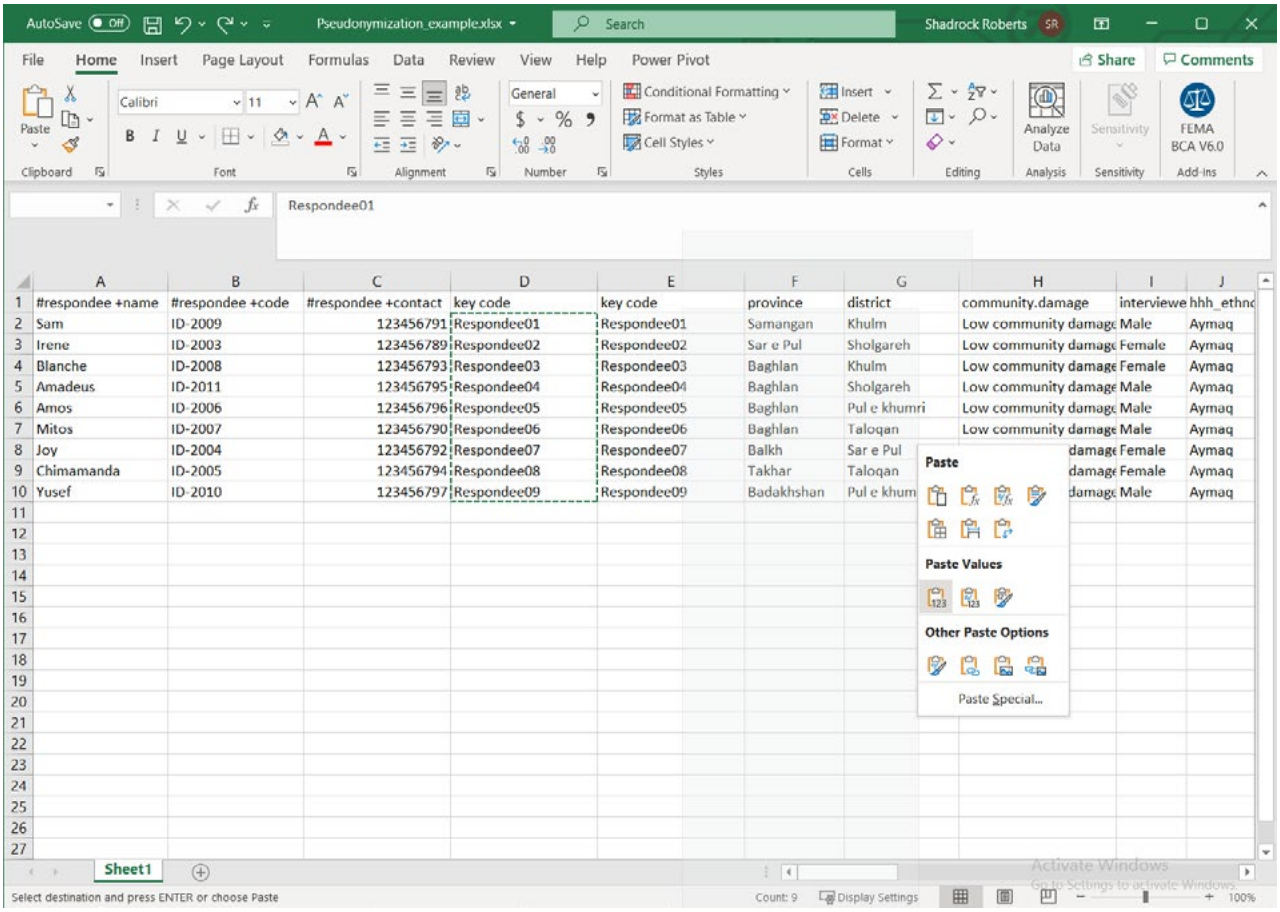
Paso 3: crear el código clave

Comience por nombrar sus nuevas columnas. Usaremos el «código clave» en cada una de ellas: cada columna contendrá los mismos valores. Este sería un buen momento para actualizar los metadatos de este conjunto de datos y explicar el significado del **código clave**. A continuación, usaremos la función de Llenado Automático de Excel para crear un código sencillo. Escriba **Respondee01** en la primera celda. A continuación, resalte esa celda, haga clic en el manejador de arrastre de la esquina inferior derecha de la celda y arrastre hasta el final del conjunto de datos. Esto llenará automáticamente el número final de cada registro para que cada respondedor tenga ahora un nuevo código.



Paso 4: duplicar el código clave y eliminar las fórmulas

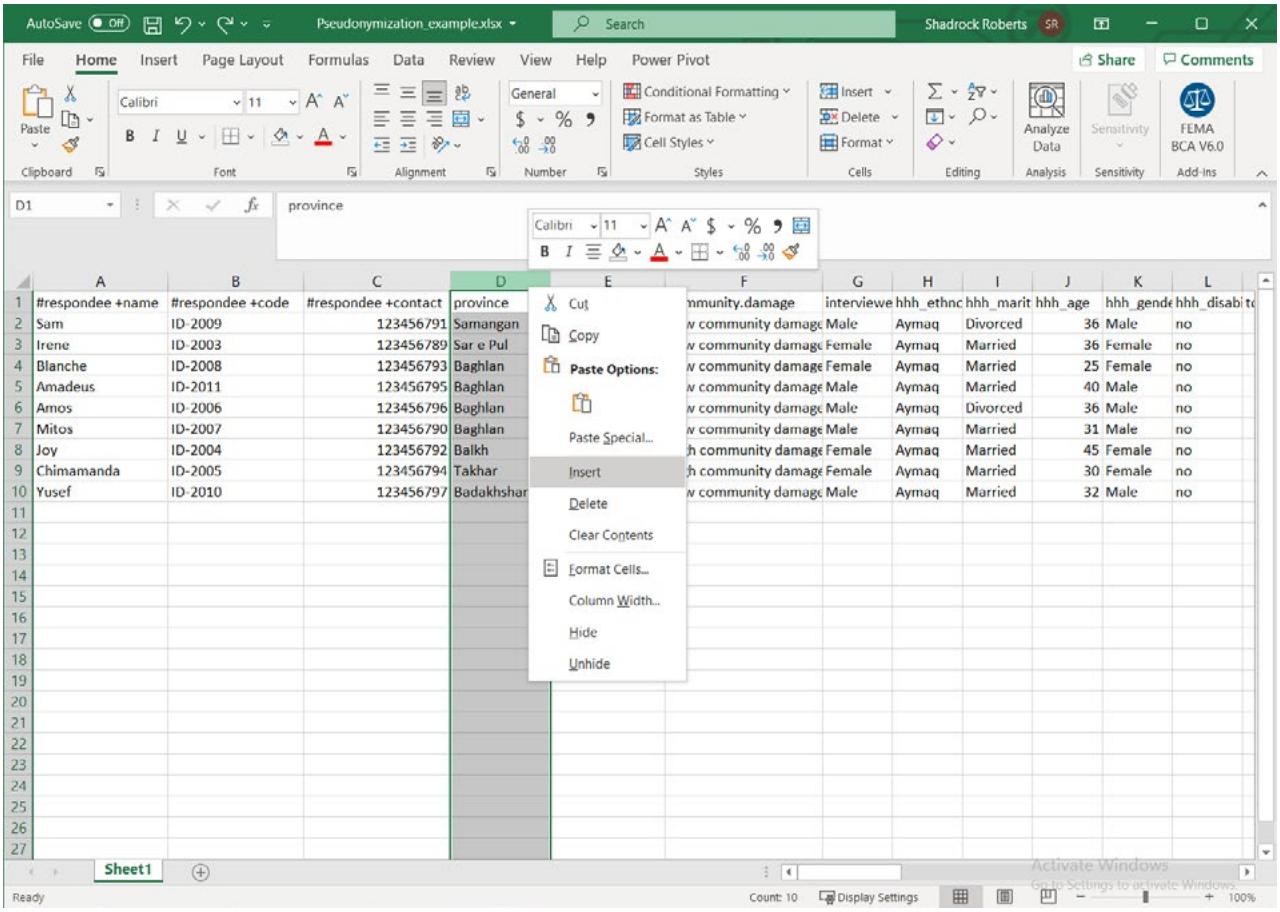
Ahora copiaremos el código clave y lo pegaremos en la columna adyacente. Puede hacerlo con comandos básicos del teclado como **ctrl + C** o al resaltar las celdas que desea copiar, al hacer clic con el botón derecho del mouse y seleccionar **Copiar**. En la columna adyacente, resalte las celdas en las que desea pegar el nuevo código clave, haga clic con el botón derecho y elija **Pegar**. Opté por pegar específicamente solo los valores. Si ha usado una fórmula para crear un nuevo código, ¡será importante conservar *solo los valores* para usarlos como código clave!



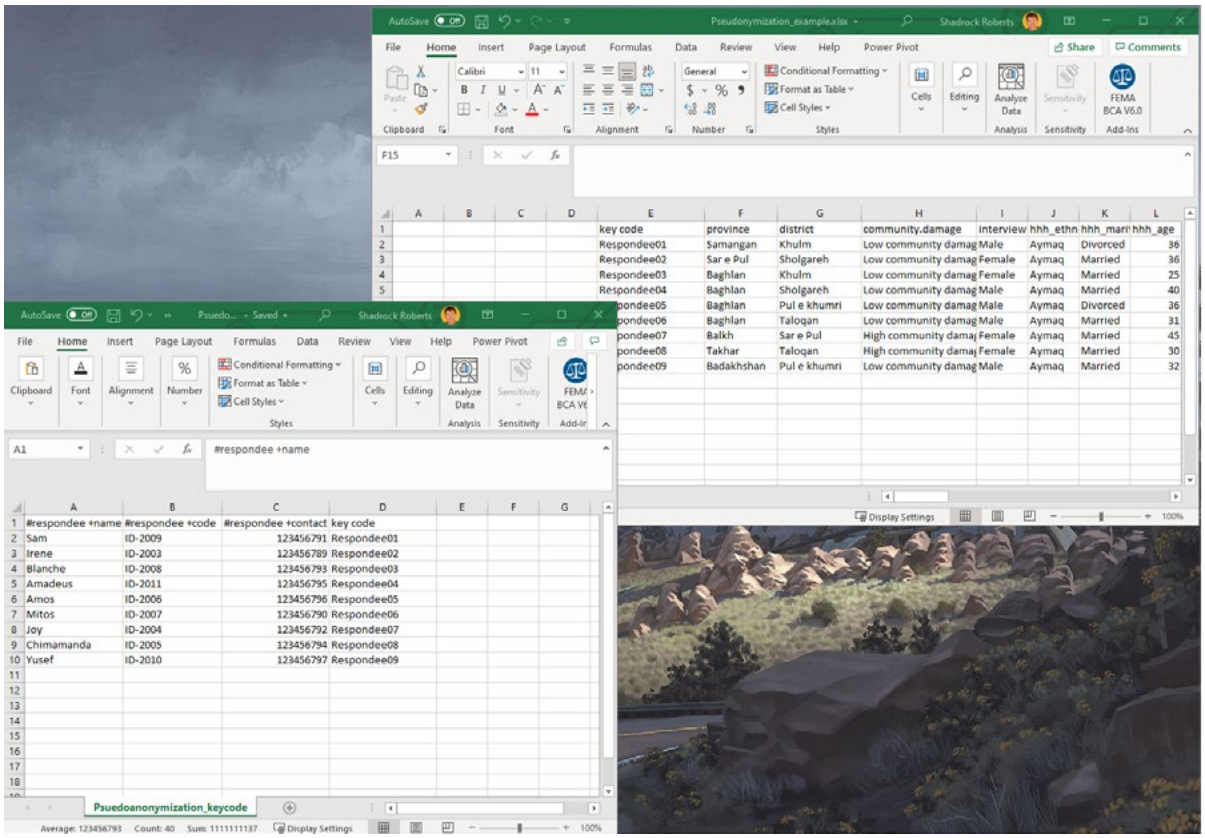
	A	B	C	D	E	F	G	H	I	J
1	#respondee +name	#respondee +code	#respondee +contact	key code	key code	province	district	community.damage	interviewe	hhh_ethnc
2	Sam	ID-2009	123456791	Respondee01	Respondee01	Samangan	Khulm	Low community damage	Male	Aymaq
3	Irene	ID-2003	123456789	Respondee02	Respondee02	Sar e Pul	Sholgareh	Low community damage	Female	Aymaq
4	Blanche	ID-2008	123456793	Respondee03	Respondee03	Baghlan	Khulm	Low community damage	Female	Aymaq
5	Amadeus	ID-2011	123456795	Respondee04	Respondee04	Baghlan	Sholgareh	Low community damage	Male	Aymaq
6	Amos	ID-2006	123456796	Respondee05	Respondee05	Baghlan	Pul e khumri	Low community damage	Male	Aymaq
7	Mitos	ID-2007	123456790	Respondee06	Respondee06	Baghlan	Taloqan	Low community damage	Male	Aymaq
8	Joy	ID-2004	123456792	Respondee07	Respondee07	Balkh	Sar e Pul	damage	Female	Aymaq
9	Chimamanda	ID-2005	123456794	Respondee08	Respondee08	Takhar	Taloqan	damage	Female	Aymaq
10	Yusef	ID-2010	123456797	Respondee09	Respondee09	Badakhshan	Pul e khum	damage	Male	Aymaq

Paso 5: separar los identificadores directos e indirectos

Resalte las columnas que tienen los identificadores directos con PII junto con una de las columnas de códigos clave. En este ejemplo, resaltamos las columnas A-D. Haga clic con el botón derecho sobre ellas y seleccione **Cortar**.



A continuación, abra una nueva hoja de cálculo y pegue estos valores con el atajo de teclado **ctrl + V**, u otro método. Guarde la nueva hoja de cálculo. Ahora tiene dos hojas de cálculo: una tiene los identificadores indirectos mientras que la nueva tiene los identificadores directos con PII. Ambos conjuntos de datos tienen un código clave para cada registro de los datos, de modo que todos los datos se puedan recombinar cuando sea necesario.



Próximos pasos

Ambos archivos tienen un código clave que permitirá volver a unirlos. Una forma de hacer esto en Excel, es usar la función VLOOKUP para llenar automáticamente las celdas según el valor de otras celdas. En este caso, podría llenar las celdas vacías del archivo original con la PII que falta, según el valor del **código clave**.

Dado que el nuevo archivo tiene los identificadores directos que tienen PII, se debe almacenar de forma segura. Una forma excelente de hacerlo es cifrar el archivo y usar el almacenamiento en la nube para limitar el acceso al archivo (**consulte las guías de buenas prácticas de cifrado y uso compartido de archivos**).

Recuerde: aunque se ha desidentificado la hoja de cálculo original al eliminar los identificadores directos que tienen PII evidentes, los demás identificadores indirectos tienen el potencial de combinarse con otros datos o analizarse de forma que permitan identificar a una persona.

Por esta razón, ambos archivos se deben seguir almacenando de forma segura. Si se desea compartir el archivo original, no-PII, de forma más amplia, sería fundamental realizar *una evaluación del riesgo de divulgación* para garantizar que el riesgo de que los datos puedan ser reidentificados sea mínimo. El Centro de Datos Humanitarios tiene un tutorial en línea para realizar una evaluación del riesgo de divulgación con el

software estadístico de código abierto «R». Además, la página web de Poverty Action Lab's De-identification for data publication (Desidentificación para la publicación de datos) ofrece un excelente debate sobre la desidentificación de datos e incluye un código de muestra para el software estadístico Stata. Para el personal de Mercy Corps, el borrador de la guía de T4D está disponible internamente y proporciona fórmulas adicionales en Excel.

Por último, todos estos pasos juntos ayudan a mitigar el riesgo o la exposición de la PII, por lo que deben estar en la PIA (**consulte la guía de Evaluación de impacto en la privacidad**) para que otros entiendan cómo se protegen estos datos.

Asistencia adicional

La desidentificación de los datos forma parte de las buenas prácticas de gestión de datos y del ciclo de vida más amplio de los datos, que son las actividades generales de recolección de datos individuales como parte de un programa o respuesta. Los siguientes recursos son excelentes para comenzar a comprender mejor la gestión de los datos de forma responsable.

- El conjunto de herramientas para la responsabilidad de los datos de Cash Learning Partnership está diseñado específicamente para los profesionales que reciben su pago en efectivo y vales, pero es una regla de oro en la orientación para los datos responsables. El conjunto de herramientas está disponible en inglés, árabe, francés, y español.
- El *conjunto de iniciación a los datos para el personal humanitario de campo de la Red de Aprendizaje sobre Transferencias Monetarias Electrónicas (Electronic Cash Transfer Learning Action Network)* ofrece una serie de hojas de consejos sobre datos para comprender diversos aspectos de las buenas prácticas de gestión y protección de datos.
- El *Manual de protección de datos en Acciones Humanitarias del Comité Internacional de la Cruz Roja* es una guía detallada de casi todos los aspectos de los datos humanitarios. El capítulo 2 trata específicamente sobre la desidentificación de los datos.
- El Engine Room's *Handbook of the Modern Development Specialist (Manual de la Sala de Máquinas del Especialista Moderno en Desarrollo)* es una buena visión general de los datos en el contexto de las actividades de desarrollo internacional. La sección sobre el *Intercambio de datos* trata específicamente de la desidentificación.

CONTACTO

HEATHER LOVE
Directora, Protección Global de Datos y Privacidad | IT
hlove@mercycorps.org

SHADROCK ROBERTS
Especialista en protección de datos | IT
shroberts@mercycorps.org

Acerca de Mercy Corps
Mercy Corps es una organización global líder impulsada por la creencia de que un mundo mejor es posible. Ante catástrofes y dificultades, en más de 40 países en todo el mundo, nos asociamos para poner en marcha soluciones audaces y ayudar a que las personas triunfen en la adversidad y construyan comunidades más fuertes desde adentro. Ahora y para el futuro.

MERCY CORPS

Sede global
45 SW Ankeny Street
Portland, Oregon 97204
888.842.0842
mercycorps.org

Sede europea
40 Sciences
Edinburgh EH9 1NJ
Scotland, UK
+44.131.662.5160
mercycorps.org.uk