

CONTROL DE LA DIVULGACIÓN ESTADÍSTICA

CONCLUSIONES FUNDAMENTALES:

- En el sector humanitario, los datos obtenidos en las encuestas de hogares, las evaluaciones de necesidades y otras modalidades de microdatos generan un gran volumen de información que no para de crecer. Este tipo de datos son fundamentales para determinar las necesidades y perspectivas de las personas afectadas por las crisis, pero también plantea riesgos específicos.
- Incluso si una organización elimina de los microdatos los identificadores directos, como el nombre o el número de teléfono de una persona, mediante la combinación de variables clave como la ubicación o la etnia, sigue siendo posible reidentificar a las personas y los grupos vulnerables.
- El control de la divulgación estadística (SDC, por sus siglas en inglés) es una técnica empleada para evaluar y reducir el riesgo de reidentificación de una persona o un grupo en el análisis de los microdatos. La aplicación del proceso SDC a los microdatos permite a las organizaciones compartir la información más ampliamente sin causar perjuicio alguno a las personas afectadas.
- El proceso SDC puede utilizarse para reducir el riesgo de reidentificación hasta un umbral acordado, que puede variar dependiendo del contexto en el que tiene lugar la respuesta humanitaria. Dado que el uso de la técnica SDC afecta al valor informativo o a la utilidad de un conjunto de datos, para garantizar un uso seguro, ético y eficaz de los datos es fundamental buscar un equilibrio adecuado entre el riesgo de reidentificación y la pérdida de información.
- Para poder empezar a utilizar la técnica SDC, las organizaciones deben invertir recursos en los siguientes procesos: 1) elección de la herramienta adecuada, 2) integración del proceso SDC en los flujos de trabajo de gestión de datos existentes, y 3) mejora de las prácticas mediante el aprendizaje permanente.

¿EN QUÉ CONSISTEN LOS MICRODATOS HUMANITARIOS?

Los datos sobre las características de unidades de población (por ejemplo, personas, hogares o establecimientos) obtenidos mediante un censo, una encuesta o un experimento se denominan en estadística “microdatos”.¹ En la respuesta humanitaria, este tipo de datos se reúne generalmente en las encuestas de hogares y las evaluaciones de necesidades. En el sector humanitario, los microdatos generan un gran volumen de información que no para de crecer y es fundamental para determinar las necesidades y perspectivas de las personas afectadas por las crisis.² Las organizaciones humanitarias tienen que saber cómo evaluar y gestionar la sensibilidad de este tipo de datos para poder garantizar que se utilicen de manera responsable en los diferentes contextos de respuesta.

¹ Survey Design and Statistical Methodology Metadata, Software and Standards Management Branch, Systems Support Division, United States Bureau of the Census, Washington DC., agosto de 1998, Sección 3.4.4, página 39.

² En el momento de redactar la presente nota, una búsqueda de la palabra “survey” (encuesta) en la plataforma *Humanitarian Data Exchange* dio 1.198 resultados de los 9.805 conjuntos de datos incluidos en la plataforma; una búsqueda de la palabra “assessment” (evaluación) dio 1.399 resultados.

Los microdatos brutos pueden contener datos personales y no personales sobre una serie de temas, incluidos temas sensibles, como la violencia por razón de género, las enfermedades infecciosas y otra información que puede estar registrada en campos de texto libre. La mayoría de las organizaciones humanitarias reconocen la sensibilidad de la información personal, como los nombres, los datos biométricos o los números de documento de identidad, y, en consecuencia, es práctica habitual anonimizar los conjuntos de datos. Sin embargo, combinando diferentes puntos de datos puede ser posible reidentificar a personas o divulgar información confidencial incluso después de aplicar la anonimización mencionada.

RIESGO DE REIDENTIFICACIÓN Y DIVULGACIÓN

Una cadena de puntos de datos puede permitir la reidentificación, ya sea de forma aislada o acompañada de una comprensión contextual básica. Las técnicas avanzadas de análisis de datos también pueden extraer información más sensible de la que se puede obtener en los análisis básicos, aumentando así la sensibilidad de los microdatos en el sector humanitario.

Son tres las modalidades de riesgo de divulgación comúnmente reconocidas³ y cada una de ellas puede manifestarse en relación con los microdatos en el sector humanitario:

- **Divulgación de identidad:** cuando es posible asociar a una persona conocida con un registro de datos publicado
- **Divulgación de atributos:** cuando es posible determinar alguna característica nueva de una persona a partir de la información contenida en los datos publicados
- **Divulgación inferencial:** cuando es posible determinar el valor de alguna característica de una persona a partir de los datos publicados con mayor precisión de lo que hubiera sido posible de cualquier otro modo

CONTROL DE LA DIVULGACIÓN ESTADÍSTICA

El control de la divulgación estadística (SDC) es una técnica utilizada en estadística para evaluar y reducir el riesgo⁴ de reidentificación de una persona o un grupo a partir de los resultados del análisis de una encuesta o de datos administrativos, o tras la publicación de microdatos. Esta técnica ha sido utilizada principalmente por las oficinas nacionales de estadística y otras organizaciones en este ámbito en relación con los datos de los censos.

Dado que la aplicación de la técnica SDC afecta al valor o la utilidad generales de un conjunto de datos, es fundamental buscar un equilibrio adecuado entre el riesgo de reidentificación y la pérdida de información. Para buscar un equilibrio adecuado entre el riesgo y la utilidad, las organizaciones deben considerar los diversos usos posibles de un conjunto de datos y el contexto en el que se recopilaron los datos.

La aplicación del proceso SDC para limitar el riesgo de divulgación en los microdatos consta de tres etapas:

1. Evaluar el riesgo de reidentificación

El primer paso consiste en llevar a cabo una evaluación del riesgo de divulgación para determinar la probabilidad de que puedan divulgarse los datos de las personas encuestadas contenidos en un determinado conjunto de datos.⁵ Si esta probabilidad (también conocida como porcentaje de riesgo) es o no aceptable para un conjunto de datos dependerá del contexto en el que se hayan generado dichos datos. Por ejemplo, en un entorno de conflicto, el porcentaje de riesgo permisible será generalmente más bajo que en un entorno de respuesta ante un desastre natural.

³ Para obtener más información sobre el riesgo de divulgación y las consideraciones técnicas conexas para la evaluación y la gestión de dicho riesgo mediante el proceso SDC, véase [Statistical Disclosure Control for Microdata: A Practice Guide](#).

⁴ Nota: la técnica SDC tiene por objeto prevenir la divulgación de identidad y atributos pero no está pensado específicamente para prevenir la divulgación inferencial.

⁵ Aprenda a realizar una evaluación del riesgo de divulgación siguiendo los programas de estudio del Centro sobre este tema: <https://centre.humdata.org/learning-path/>.

2. Reducir el riesgo de reidentificación

El siguiente paso consiste en someter al conjunto de datos al proceso SDC real, el cual reduce el riesgo de reidentificación aplicando diferentes métodos de anonimización. Estos métodos pueden clasificarse en dos categorías: métodos perturbativos, que no eliminan los valores contenidos en el conjunto de datos pero perturban (es decir, alteran) dichos valores para limitar el riesgo de divulgación, generando incertidumbre en torno a los valores auténticos, y métodos no perturbativos, que reducen el nivel de detalle de la información mediante la generalización o supresión de determinados valores (es decir, enmascaramiento) sin distorsionar la estructura de los datos.

3. Cuantificar la pérdida de información

El último paso consiste en medir la pérdida de información derivada de la aplicación del proceso SDC al conjunto de datos. El objetivo que se persigue en este paso es comparar el valor de la información que figura en el conjunto de datos original con el valor de la información final.

Mediante la utilización del proceso SDC para evaluar y reducir el riesgo de divulgación en los microdatos, las organizaciones humanitarias pueden compartir de manera más responsable los datos obtenidos en las encuestas y en las evaluaciones de necesidades para que sirvan de base a la labor de respuesta global.

Aplicación del control de la divulgación estadística a los datos compartidos en la plataforma HDX⁶

Desde el comienzo de 2018, el Centro de Datos Humanitarios (“el Centro”) ha llevado a cabo una evaluación de riesgos de 59 conjuntos de datos cargados en la plataforma Humanitarian Data Exchange (HDX). El riesgo de divulgación de las identidades de las personas encuestadas en 38 de los 59 conjuntos de datos se consideró demasiado alto para su publicación en la plataforma HDX. El equipo HDX aplicó el proceso SDC a 14 de dichos conjuntos de datos, según lo acordado por sus contribuyentes, con lo que el nivel de riesgo se redujo a un porcentaje aceptable (es decir, inferior a un 5%⁷). Gracias a esta reducción, 5 de los 14 conjuntos de datos pudieron publicarse después de la anonimización, mientras que los 9 restantes se eliminaron o se compartieron en privado en la plataforma HDX, al igual que los 24 de alto riesgo, a los que sus contribuyentes se negaron a aplicar la técnica SDC. Con respecto a estos 24 conjuntos de datos, muchas organizaciones adoptaron sus propias medidas para reducir el riesgo de reidentificación, que en algunos casos consistieron en la eliminación total de variables sensibles no esenciales.

APLICACIÓN DEL CONTROL DE LA DIVULGACIÓN ESTADÍSTICA EN EL MANEJO DE DATOS HUMANITARIOS

A principios de 2019, para conocer las actuales prácticas de gestión de microdatos, el Centro entrevistó a miembros de siete organizaciones humanitarias que realizan encuestas y evaluaciones de necesidades. Algunas de estas organizaciones, como el ACNUR (véase el estudio de caso que figura a continuación) aplican enfoques relativamente avanzados y cuenta con personal especializado suficiente para llevar a cabo un control de la divulgación estadística en diferentes modalidades de microdatos. Sin embargo, la mayoría de las organizaciones entrevistadas necesitan apoyo adicional para realizar este trabajo.

⁶ Podrá encontrar más información sobre el uso, por el Centro, de la técnica SDC y sobre el proceso global de aseguramiento de la calidad (AC) de los datos compartidos en la plataforma HDX en el siguiente enlace: <https://data.humdata.org/about/hdx-qa-process>.

⁷ Recientemente, el Centro modificó su umbral predeterminado del riesgo de reidentificación aceptable, reduciéndolo de un 5% a un 3%. El umbral exacto para un conjunto de datos concreto es siempre contextual y se determina conjuntamente con la organización que contribuye a dicho conjunto de datos.

Custodia y gestión responsables de microdatos sobre refugiados

Experiencia del ACNUR

El ACNUR recopila de manera rutinaria datos sobre los refugiados y otras poblaciones bajo su mandato. Estos datos se utilizan luego para evaluar las necesidades y vulnerabilidades de dichas poblaciones, para que sirvan a base a los programas y para mejorar la asignación de la asistencia. Aunque estos datos no se han retenido históricamente en formatos y ubicaciones que facilitarían su posterior recuperación para ser utilizados en el futuro, el ACNUR se encuentra actualmente inmerso en el proceso de crear una biblioteca de microdatos interna y externa. Con la creación de estos depósitos de datos en línea que proporcionarán a los usuarios internos y externos acceso público a los microdatos, el ACNUR pretende posibilitar un uso más amplio de la información a las partes interesadas y prevenir la duplicación de las actividades de acopio de datos.

La difusión pública de microdatos proporciona numerosos beneficios pero también conlleva posibles riesgos. Cuando se difunde información sin adoptar medidas adecuadas de control, existe el riesgo de que posibles intrusos puedan identificar las entidades (personas u hogares) cuyos datos se comparten, incluso cuando se han eliminado los identificadores directos, como nombres y direcciones. De conformidad con la [política de protección de datos del ACNUR](#), debe protegerse la identidad de las personas de interés para la organización y, por consiguiente, los conjuntos de datos deben anonimizarse adecuadamente antes de compartirlos. Los datos del ACNUR son especialmente sensibles, dado que se refieren a grupos de personas especialmente vulnerables, cuya protección reviste la máxima importancia.

Para garantizar la protección y difusión responsables de los microdatos, el ACNUR utiliza la aplicación *sdcMicro* en el entorno R para calcular los riesgos de reidentificación antes de su publicación. El proceso es gestionado por el equipo de custodia del ACNUR, que trabaja conjuntamente con los propietarios de los datos para identificar las variables clave, evaluar la sensibilidad de la información y definir un nivel de riesgo aceptable para cada conjunto de datos. Una vez anonimizados, los datos modificados se comparan con los originales y se examinan para evaluar la pérdida de información. Si el propietario de los datos estima que determinadas variables modificadas son esenciales para los usuarios, los métodos de control de la divulgación pueden ajustarse según proceda. Por ejemplo, en el caso de las encuestas normalizadas ampliadas de nutrición (SENS, por sus siglas en inglés), el equipo de custodia decidió no aplicar la agregación a grupos de edad a los que normalmente se aplicaría, porque dichos grupos eran fundamentales para caracterizar la malnutrición de los niños por años y meses de edad. El equipo mantuvo la variable de edad pero excluyó la fecha de nacimiento y la fecha de la encuesta. Este método dio lugar a un escenario de riesgo aceptable y, al mismo tiempo, mantuvo la utilidad de los datos para los nutricionistas.

El ACNUR sigue invirtiendo recursos en este proceso y ha ampliado su equipo de custodia y ha aumentado los conocimientos especializados en técnicas de anonimización que existen en la organización. En el marco del actual plan, para finales de 2019, la herramienta [UNHCR Microdata Library](#) estará totalmente operativa e incluirá microdatos sobre desplazamientos forzados.

⁶ Podrá encontrar más información sobre el uso, por el Centro, de la técnica SDC y sobre el proceso global de aseguramiento de la calidad (AC) de los datos compartidos en la plataforma HDX en el siguiente enlace: <https://data.humdata.org/about/hdx-qa-process>.

⁷ Recientemente, el Centro modificó su umbral predeterminado del riesgo de reidentificación aceptable, reduciéndolo de un 5% a un 3%. El umbral exacto para un conjunto de datos concreto es siempre contextual y se determina conjuntamente con la organización que contribuye a dicho conjunto de datos.

Mediante la colaboración con contribuyentes de datos, como la iniciativa REACH (véase el estudio de caso que figura a continuación) para desarrollar un modelo de servicio fiable y seguro para el proceso SDC, el Centro pretende promover el intercambio responsable de esta información y demostrar el valor de técnicas más sólidas de evaluación del riesgo de divulgación y anonimización de datos. La experiencia con estas técnicas también ayuda a las organizaciones humanitarias a determinar las herramientas y los métodos que pueden incorporar en sus propios procesos de gestión de datos, al tiempo que contribuyen a ampliar el caudal de conocimientos dentro del sector sobre la forma de gestionar y compartir microdatos de manera más responsable en los entornos humanitarios.

Oportunidades y desafíos en relación con la incorporación del control de la divulgación estadística en el flujo de trabajo de una organización

Experiencia de la iniciativa REACH

La iniciativa **REACH** empezó a explorar el potencial del control de la divulgación estadística en junio de 2018, cuando el equipo de la plataforma HDX aplicó por primera vez el paquete `sdcMicro` en el entorno R a un conjunto de datos cargado en la plataforma. Los tipos de datos a los que el equipo HDX aplicó el proceso SDC para REACH incluyen encuestas de hogares y entrevistas a informantes clave (y los metadatos conexos). REACH todavía no ha aplicado directamente la técnica SDC pero sus administradores están examinando los requisitos necesarios para hacerlo.

Basándose en la experiencia de REACH a día de hoy, las organizaciones que estén interesadas en incorporar el proceso SDC en su flujo de trabajo deberían plantearse las siguientes preguntas:

- ¿Es esta la metodología adecuada para sus actuales procesos de gestión de microdatos?
- ¿Hasta qué punto la aplicación del proceso SDC reduce la validez y utilidad de los datos?
- ¿En qué medida afecta a la transparencia la aplicación del proceso SDC?
- ¿Cómo se puede evitar que el personal confíe demasiado en los resultados de una evaluación del riesgo de divulgación tras la aplicación del proceso SDC y garantizar que siga pensando de manera crítica en los posibles riesgos que plantean diferentes tipos de datos?

La iniciativa REACH considera que sería operacionalmente viable implementar con relativa facilidad los aspectos técnicos del proceso SDC en la sede y sobre el terreno. En la sede, dicha implementación exigiría ejecutar un programa en todos los conjuntos de datos generados o publicados por REACH. Para la implementación sobre el terreno, los equipos en los países deben utilizar `sdcMicro` o una herramienta similar con los conjuntos de datos generados en los países.

Más allá de los aspectos técnicos del proceso SDC, la iniciativa REACH ve un posible problema o un cuello de botella en el componente manual del proceso, cuando el personal deba decidir si una determinada técnica de control de la divulgación es adecuada, qué variables deben eliminarse u ocultarse y cómo interpretar o comunicar los resultados del proceso. Estas decisiones llevan tiempo y exigen la comprensión del contexto con el que están relacionados los datos.

A corto plazo, la iniciativa REACH seguirá colaborando con el equipo HDX en la aplicación del proceso SDC a los datos de las encuestas y evaluaciones antes de su publicación en la plataforma HDX. Esta experiencia permitirá a REACH determinar cuál es la mejor manera de incorporar en el futuro la técnica SDC en sus propios flujos de trabajo a nivel mundial y nacional.

RECOMENDACIONES PARA PROMOVER LA UTILIZACIÓN DEL CONTROL DE LA DIVULGACIÓN ESTADÍSTICA EN ENTORNOS HUMANITARIOS

Para adoptar con éxito el proceso SDC, el Centro y sus colaboradores en la elaboración de la presente nota de orientación recomiendan a las organizaciones que inviertan recursos en los tres pasos enumerados a continuación:

1. Seleccionar la herramienta adecuada

Para la implementación de la técnica SDC existen diversas herramientas gratuitas en línea. El Centro y otras organizaciones humanitarias consultadas por éste durante su labor de investigación utilizan actualmente **sdcMicro**. El Centro eligió sdcMicro por su escalabilidad y porque es un software gratuito y de fuente abierta. Otras herramientas gratuitas y de fuente abierta son **µArgus** y **ARX**. A la hora de seleccionar la herramienta adecuada para el proceso SDC, las organizaciones deben tener en cuenta la flexibilidad para seleccionar las variables clave, la capacidad para tratar grandes conjuntos de datos, la funcionalidad para la elección entre riesgo y utilidad y la facilidad con la que el personal podría navegar por la interfaz de usuario del software.

2. Integrar el control de la divulgación estadística en los flujos de gestión de datos existentes

Establecer un proceso para la aplicación de esta técnica a los flujos de gestión de datos existentes es clave para adoptar este método de manera sostenible. La técnica SDC exige la participación de personal con diferentes conocimientos y competencias, incluido un especialista técnico para aplicar los métodos estadísticos, y un especialista de programas con conocimiento del contexto de los datos para determinar un equilibrio aceptable entre riesgo y utilidad. Un flujo de trabajo bien organizado ayudará a mejorar la eficacia del proceso y a prevenir errores de interpretación de los resultados del proceso SDC y la excesiva dependencia de dichos resultados.

3. Mejorar la práctica mediante el aprendizaje permanente

A medida que las organizaciones vayan aplicando la técnica SDC, irán conociendo la sensibilidad de las variables clave, el nivel de riesgo que deben intentar establecer, el nivel aceptable de pérdida de información y otras consideraciones que exijan un equilibrio en el proceso. El mantenimiento de un registro de cada una de las aplicaciones del proceso SDC y la documentación de las lecciones aprendidas contribuirá a perfeccionar dicho proceso de cara al futuro. Compartir internamente los conocimientos adquiridos y, en su caso, con la comunidad humanitaria en general puede propiciar una gestión más coherente y responsable de los microdatos en el sector.

Como parte de sus esfuerzos para promover una gestión y un intercambio más responsables de los datos humanitarios, el Centro está mejorando su modelo de servicios⁸ para la aplicación de la técnica SDC. Esta labor consiste en introducir un proceso automatizado de detección de riesgos para todos los datos compartidos a través de la plataforma HDX, lo cual –cuando se realiza manualmente– puede llevar varias horas en el caso de hojas de cálculo grandes. Mediante este proceso, se ejecutará un programa con toda la información cargada en la plataforma para identificar los microdatos y otros tipos de datos potencialmente sensibles. Los datos de alto riesgo se enviarán a un flujo de trabajo exclusivo que se evaluará y, en caso necesario, se modificará por medio del proceso SDC para reducir el riesgo de reidentificación antes de compartir los datos más ampliamente.

Para obtener más información sobre la labor del Centro en lo tocante al proceso SDC diríjase a centrehumdata@un.org.

⁸ Podrá encontrar más información sobre el enfoque del Centro en relación con el proceso SDC en el siguiente enlace: <https://humanitarian.atlassian.net/wiki/spaces/HDXKB/pages/1381498881/>.

COLABORADORES: ACNUR E INICIATIVA REACH.

El **Centro Para Los Datos Humanitarios** ('El Centro'), conjuntamente con sus principales asociados, publica en 2019 y 2020 una serie de ocho notas de orientación sobre la responsabilidad de los datos en la acción humanitaria. Esta serie de notas de orientación sigue a la publicación, en marzo de 2019, del **borrador de trabajo de directrices sobre la responsabilidad de los datos, elaborado por la OCHA**. Con esta serie, el Centro pretende proporcionar directrices adicionales sobre cuestiones, herramientas y procesos específicos relacionados con la responsabilidad de los datos en la práctica. La serie ha sido posible gracias al generoso apoyo de la Dirección General de Protección Civil y Operaciones de Ayuda Humanitaria Europeas (DG ECHO).



El presente documento abarca las actividades de ayuda humanitaria llevadas a cabo con la asistencia financiera de la Unión Europea. Las opiniones aquí expresadas no deben interpretarse en modo alguno como la opinión oficial de la Unión Europea, y la Comisión Europea no es responsable de la utilización que pueda hacerse de la información que contiene.

Este proyecto está
cofinanciado por la Unión
Europea