

*Data  
Protection is  
**People**  
Protection*



**ДОВІДНИК ІЗ ЗАХИСТУ ДАНИХ І КОНФІДЕНЦІЙНОСТІ:**

# **Знеособлення даних**

У цьому посібнику наведено приклад видалення з набору даних інформації, що ідентифікує особу (ІІО). Існує кілька способів «знеособлення» даних, під якими розуміють заходи або методи обробки, спрямовані на запобігання розкриттю особистості суб'єкта даних. Два поширених типи знеособлення — "анонімізація" і "псевдонімізація".

**Анонімізація** — це процес, за допомогою якого персональні дані стають анонімними, щоб особу (або "суб'єкт даних") більше не можна було ідентифікувати. Це постійна зміна даних. Загальні методи включають в себе видалення інформації, що ідентифікує особу, або шифрування значень у певних наборах ПО.

Приклад: уявіть, що організація має дані опитування, які містять поля для імені, державного ідентифікаційного коду, назви села, етнічної належності, віку, рівня освіти та показників здоров'я. У цьому випадку видалення імені та державного ідентифікаційного коду буде першим кроком до того, щоби зробити дані анонімними, оскільки ці «прямі характеристики» є персональними даними, які безпосередньо ідентифікують особу. «Непрямі характеристики» — назва села, етнічна належність, вік, рівень освіти та показники здоров'я залишаються.

Однак, попри те, що деякі характеристики здаються "анонімними", вони можуть такими не бути. Якщо опитування проводилося в дуже маленькому селі, де лише двоє жителів ідентифікують себе із певною етнічною належністю, і вони різного віку, то використання цих двох непрямих характеристик може дозволити ідентифікувати цих людей! Процес, у ході якого досліджуються всі атрибути для зниження ризику повторної ідентифікації суб'єкта даних, називається статистичним контролем розкриття даних. Першим кроком у цьому процесі є оцінка ризиків розкриття даних. У Центрі гуманітарних даних є [навчальний онлайн-посібник для проведення оцінки ризику розкриття даних](#).

**Псевдонімізація**, з іншого боку, полягає в обробці персональних даних таким чином, що персональні дані більше не може бути віднесено до конкретного суб'єкта даних без використання додаткової інформації, такої як ключовий код.

Приклад: уявіть, що опитування містить ваше ім'я, адресу електронної пошти, вік, національність і робоче місце. Псевдонімізація бере дані, за якими можна ідентифікувати конкретно вас (ваше ім'я, адресу електронної пошти, вік), і робить їх недоступними та відокремленими від неспецифічних даних, таких як ваша національність. Псевдонімні дані може бути знову зібрано разом в якийсь момент, так що всю інформацію може бути знову пов'язано з конкретним джерелом або особою. Ось чому псевдонімізація вимагає, щоби додаткова інформація зберігалася окремо і підлягала технічним та організаційним заходам для забезпечення неможливості віднесення персональних даних до суб'єкта даних.

## Що обрати — анонімізацію або псевдонімізацію?

Анонімізація, як правило, буде безпечнішою та зменшить ризик розкриття ПО. Однак іноді це може зробити дані занадто загальними, через що вони можуть втратити користь для таких програм, як робота з грошовими ваучерами. У випадку програм охорони здоров'я, які включають вакцинацію або інші методи лікування, може бути важливо зв'язатися з людьми для подальшого лікування. В обох цих випадках псевдонімізація буде найкращим вибором, оскільки ви завжди можете знову зібрати дані разом, щоби ідентифікувати особу, коли це необхідно.

Немає єдиної правильної відповіді на те, коли віддавати перевагу одному методу перед іншим, і перш ніж обрати, як знеособити свої дані, важливо розуміти, чому дані було зібрано, потенційні ризики, пов'язані зі зберіганням цих даних, і потреби програми.

Важливо також розуміти, що методи, які використовуються як для анонімізації, так і для злову даних, стають все складнішими, і що [навіть знеособлені дані не завжди на сто відсотків в безпеці](#). Якщо сумніваєтеся, зверніться по допомогу до своєї команди з даних або ІТ.

## ☆ Важливість

Останні [витоки даних у Міжнародному Комітеті Червоного Хреста](#), [зломи електронної пошти в Агентстві США з міжнародного розвитку](#) та [неналежне розповсюдження даних Верховним Комісаром ООН у справах біженців](#) — все це показує, яким чином гуманітарні дані наражаються на ризик. Дані опитувань домогосподарств, оцінки потреб та інші форми мікроданих складають все більш значний обсяг даних у гуманітарному секторі. Ці типи даних мають вирішальне значення для визначення потреб і перспектив учасників програми та спільнот, у яких ми працюємо, але ці дані також пов'язані з певними ризиками. Розуміння того, як оцінювати та керувати чутливістю цих даних, має дуже важливе значення для забезпечення їх безпечного, етичного та ефективного використання в різних контекстах реагування.

Деякі переваги використання анонімізованих даних над персональними даними:

- › - захист від неналежного розкриття персональних даних;
- › - до анонімізованих даних застосовується менше юридичних обмежень;
- › - надання організаціям можливості створювати відкриті або загальнодоступні дані, дотримуючись при цьому своїх зобов'язань щодо захисту даних.

## Принципи

Знеособлення даних є частиною обробки даних, а обробка персональних даних, здійснювана гуманітарними організаціями, повинна відповідати наступним принципам.

- › **Справедливість і законність обробки:** методи повинні відповідати регіональному, національному чи місцевому законодавству або політиці, які можуть обмежувати те, які дані може бути знеособлено і як використовуються певні технології. Будь-яка обробка персональних даних має бути прозорою для залучених суб'єктів даних.
- › **Обмеження цілей:** гуманітарні організації повинні визначити та викласти конкретні цілі, для яких обробляються дані. Ці цілі повинні бути чіткими та законними.
- › **Пропорційність:** забезпечити, щоб кожен конкретний захід, пов'язаний з обробкою персональних даних, відповідав заявленій меті. Наприклад, чи збирається лише мінімально необхідний обсяг даних? Чи вжито відповідні технічні та організаційні заходи для зменшення ризиків, пов'язаних з обробкою даних?
- › **Технологічні зміни:** нові набори даних та нові інструменти для їх аналізу швидко змінюються та розвиваються, як і засоби, за допомогою яких дані зламують або крадуть. Важливо розуміти нові та потенційні ризики для ваших даних і продовжувати коригувати свої методи та практики відповідним чином.

## Псевдонімізація

Це приклад одного зі способів знеособлення даних в електронній таблиці. Існує широкий спектр способів знеособлення. У цьому прикладі використовується «ключовий код» для видалення інформації, що ідентифікує особу, знайденої в прямих ідентифікаторах, і збереження її в окремому файлі. Інформація, що ідентифікує особу, (PIO) — це інформація, яка може бути використана для визначення особи. Типовими прикладами є повне ім'я, адреса, номер телефону, дата народження та номер соціального забезпечення або національний ідентифікаційний код.



## Інструкції

Ви можете слідувати цим [Інструкціям із псевдонімізації](#), щоб розглянути базовий приклад псевдонімізації набору даних. У цій вправі використовується [зразок набору даних, знайдений у папці з даними онлайн-посібника](#).

Після псевдонімізації зразка даних ви можете продовжити роботу з навчальним посібником від Центру гуманітарних даних [для проведення оцінки ризиків розкриття даних](#).

### Крок 1 — визначте PIO

Почніть із виявлення PIO в даних. В ідеалі у вас будуть метадані — дані або документ, який визначає ваші дані, — щоб допомогти вам зрозуміти, які поля містять PIO. У зразку даних є три стовпці, які містять потенційну PIO:

- » № респондента +ім'я, схоже, містить ім'я.
- » № респондента +код, ймовірно, містить якийсь ідентифікаційний номер.
- » № респондента +контакт, можливо, містить номер мобільного телефону

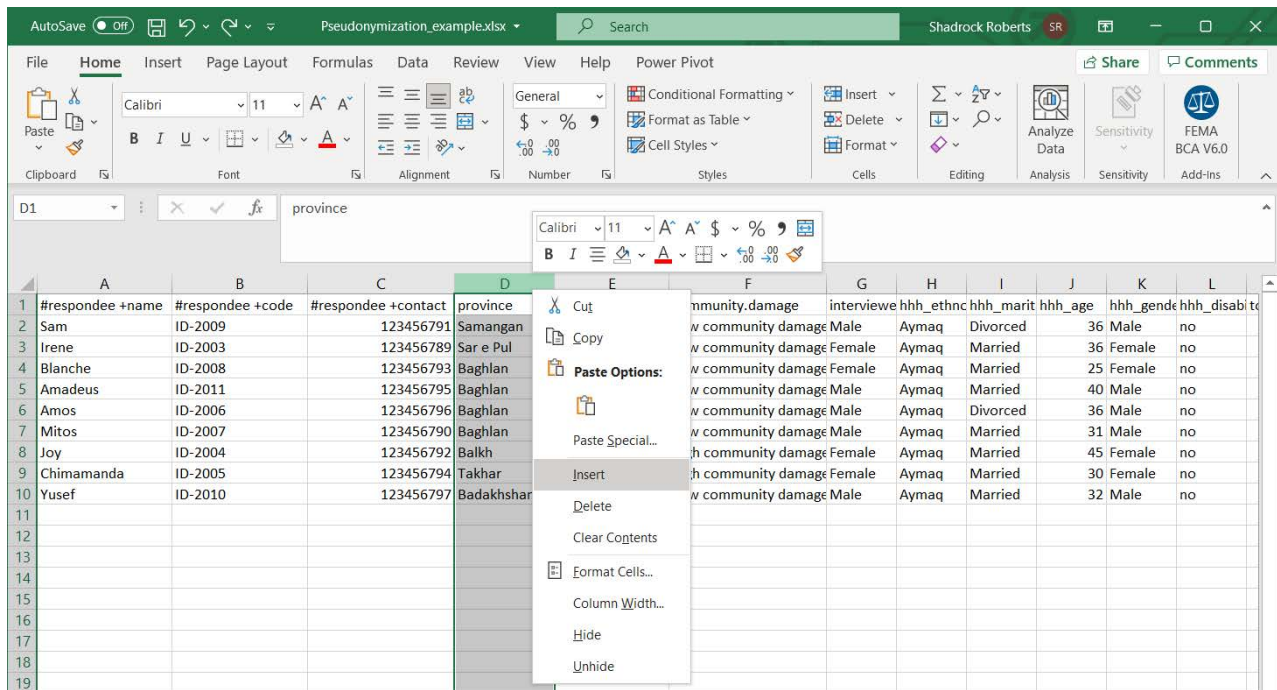
Кожен із цих прямих ідентифікаторів використовує [гуманітарну мову обміну для позначення даних](#).

|    | 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_disab |
| 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 |                 |                  |                     |            |              |                              |            |           |           |         |           |           |



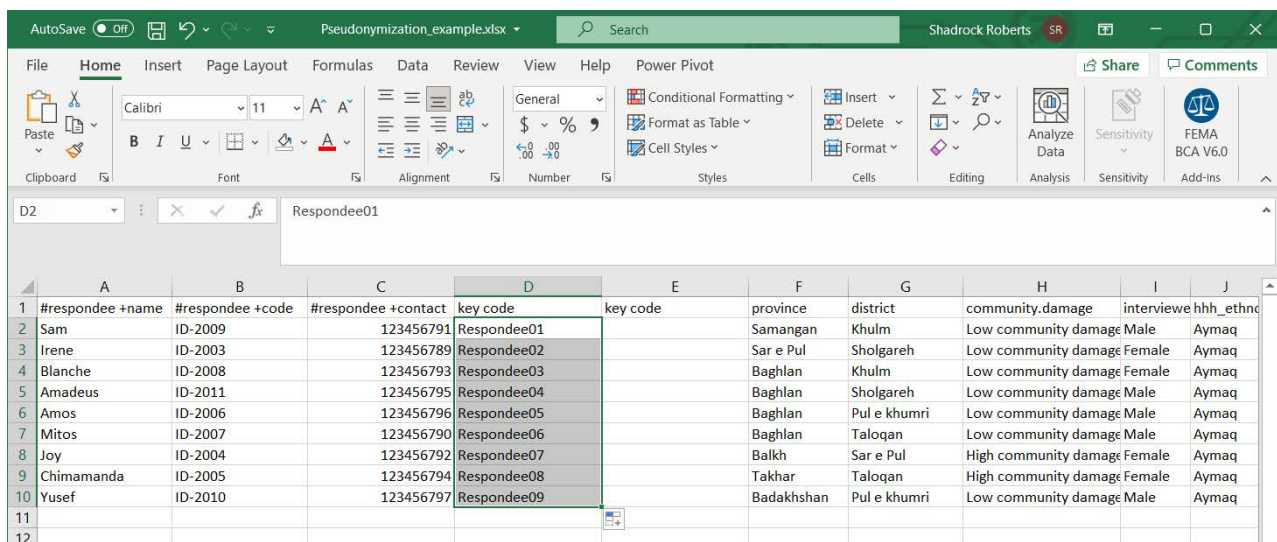
## Крок 2 — створіть нові стовпці для ключового коду

Ми будемо використовувати ключовий код, значення, яке ми генеруємо, щоб розбити ІО. Оскільки всі прямі ідентифікатори згруповані разом, ми створимо два нових стовпці між стовпцем C, **№ респондента** + **контакт** і стовпцем D, **область**. У Excel ми робимо це, виділяючи стовпець праворуч від того місця, куди ми хочемо вставити нові стовпці. Кладніть правою кнопкою миші на стовпці та оберіть **"Вставити"**. Повторіть цей процес ще раз, щоб створити ще один пустий стовпець.



## Крок 3 — створіть ключовий код

Почніть з іменування нових стовпців. Ми будемо використовувати "ключовий код" у кожному з них: кожен стовпець буде містити однакові значення. Зараз саме час оновити всі метадані цього набору даних, щоб пояснити, що означає **ключовий код**! Далі ми скористаємося **функцією Excel "Автозаповнення"** для створення простого коду. Введіть **Респондент01** у першу комірку. Далі виділіть цю комірку, натисніть на маркер перетягування в правому нижньому кутку комірки та перетягніть його вниз до кінця набору даних. Це автоматично заповнить остаточний номер кожного запису, щоб кожен респондент тепер мав новий код.



## Крок 4 — дублювання ключового коду та видалення формул

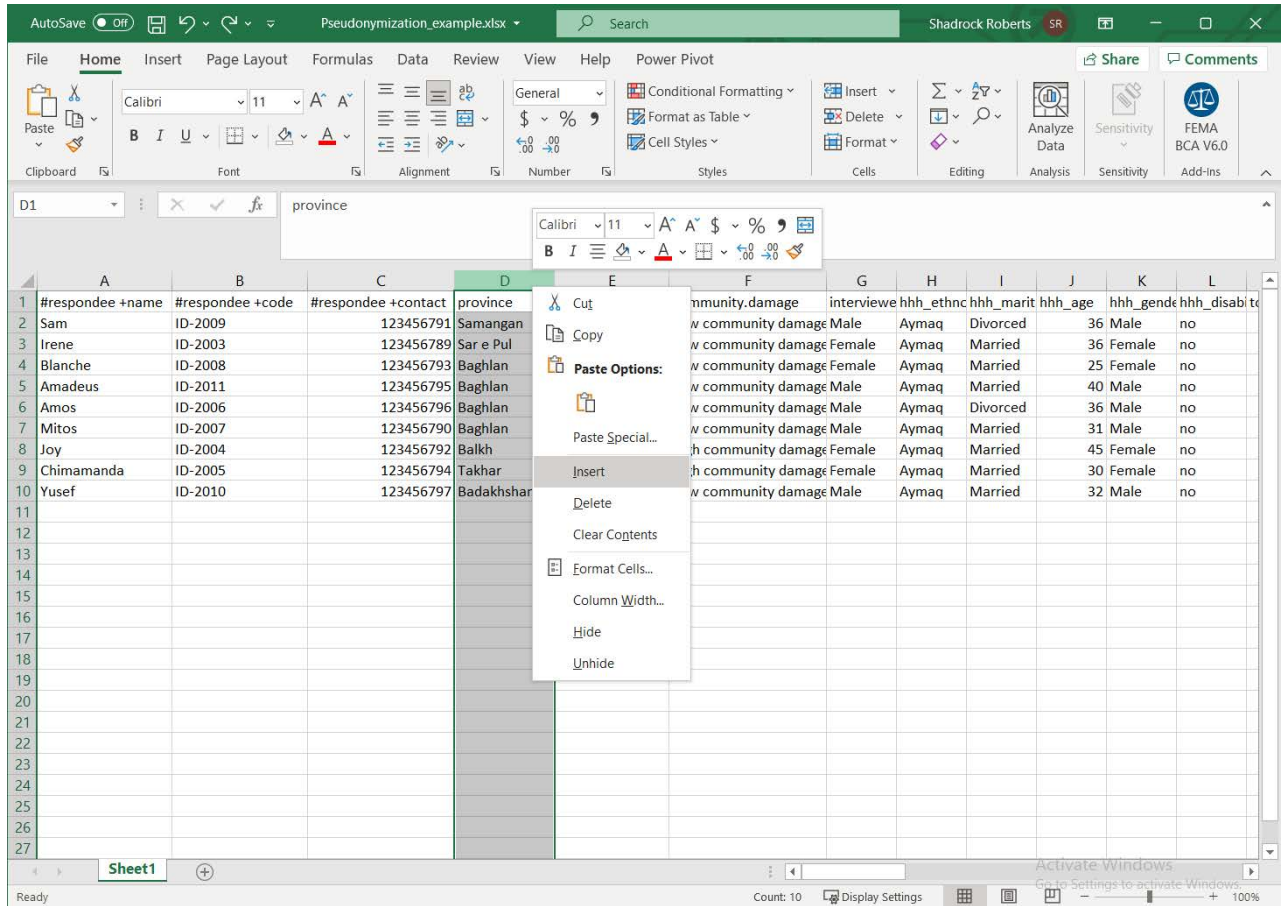
Тепер скопіюємо ключовий код і вставимо його в сусідній стовпець. Зробити це можна за допомогою основних клавіатурних команд, зокрема **ctrl + C**, або виділіть комірки, які потрібно скопіювати, клацніть на них правою кнопкою миші та оберіть **"Копіювати"**. У сусідньому стовпці виділіть комірки, в які потрібно вставити новий ключовий код, клацніть правою кнопкою миші та оберіть **"Вставити"**. Ми вирішили спеціально вставити лише значення. Якщо для створення нового коду ви використовували формулу, то важливо буде зберегти *лише значення* для використання як коду ключа!

The screenshot shows the Microsoft Excel interface with the 'Home' tab selected. A data table is visible with columns A through J. The 'key code' column (D) is highlighted, and a context menu is open over it, showing options like 'Paste', 'Paste Values', and 'Other Paste Options'. The table contains data for 10 respondents, including names, IDs, contact numbers, and codes.

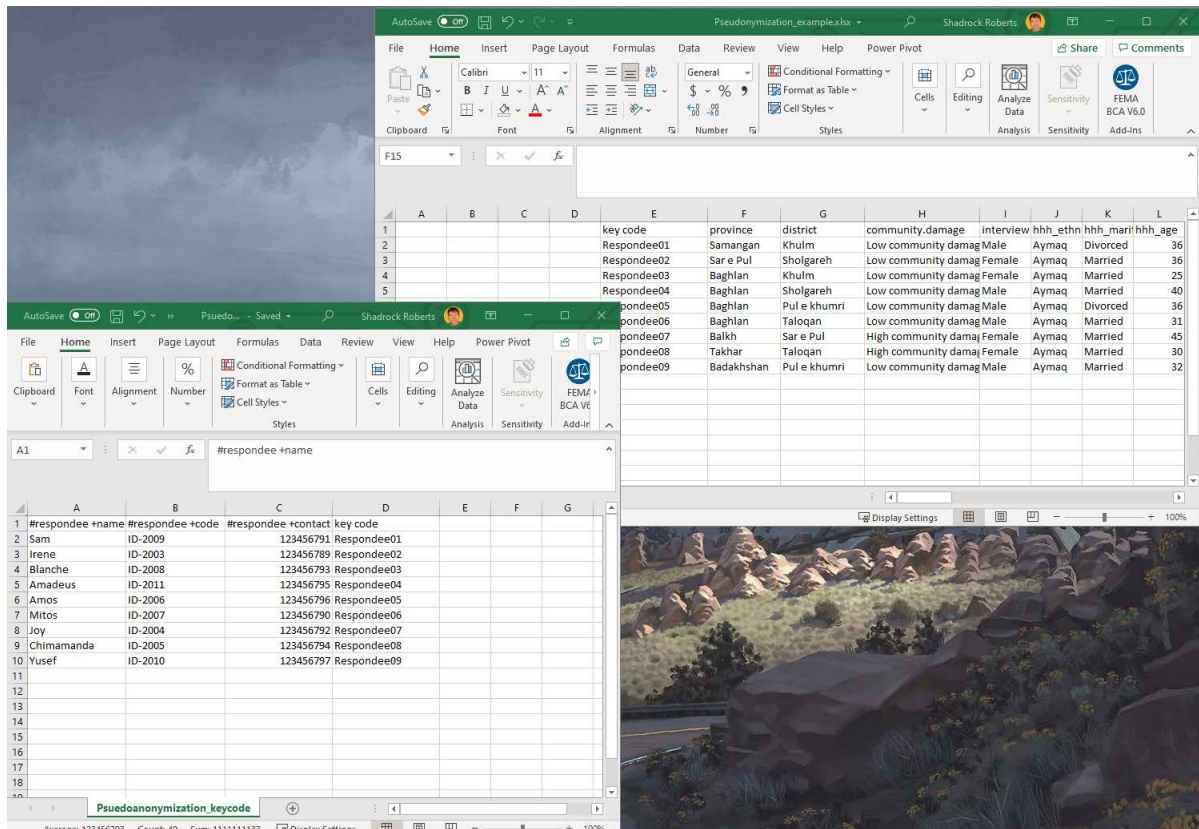
| #respondent +name | #respondent +code | #respondent +contact | key code    | key code    | province   | district     | community.damage     | interviewee | hhh_ethnc |
|-------------------|-------------------|----------------------|-------------|-------------|------------|--------------|----------------------|-------------|-----------|
| Sam               | ID-2009           | 123456791            | Respondee01 | Respondee01 | Samangan   | Khulm        | Low community damage | Male        | Aymaq     |
| Irene             | ID-2003           | 123456789            | Respondee02 | Respondee02 | Sar e Pul  | Sholgareh    | Low community damage | Female      | Aymaq     |
| Blanche           | ID-2008           | 123456793            | Respondee03 | Respondee03 | Baghlan    | Khulm        | Low community damage | Female      | Aymaq     |
| Amadeus           | ID-2011           | 123456795            | Respondee04 | Respondee04 | Baghlan    | Sholgareh    | Low community damage | Male        | Aymaq     |
| Amos              | ID-2006           | 123456796            | Respondee05 | Respondee05 | Baghlan    | Pul e khumri | Low community damage | Male        | Aymaq     |
| Mitos             | ID-2007           | 123456790            | Respondee06 | Respondee06 | Baghlan    | Taloqan      | Low community damage | Male        | Aymaq     |
| Joy               | ID-2004           | 123456792            | Respondee07 | Respondee07 | Balkh      | Sar e Pul    | damage               | Female      | Aymaq     |
| Chimamanda        | ID-2005           | 123456794            | Respondee08 | Respondee08 | Takhar     | Taloqan      | damage               | Female      | Aymaq     |
| Yusef             | ID-2010           | 123456797            | Respondee09 | Respondee09 | Badakhshan | Pul e khum   | damage               | Male        | Aymaq     |

## Крок 5 — відокремте прямі та непрямі ідентифікатори

Виділіть стовпці, які містять прямі ідентифікатори з ІО, а також один із стовпців із ключовим кодом. У цьому прикладі ми виділяємо стовпці A-D. Клацніть на них правою кнопкою миші та оберіть "Вирізати".



Далі відкрийте нову електронну таблицю та вставте ці значення за допомогою комбінації клавіш **ctrl + V** або іншим способом. Збережіть нову електронну таблицю. Тепер у вас є дві електронні таблиці: одна з них містить непрямі ідентифікатори, тоді як новий аркуш містить прямі ідентифікатори з ІО. Обидва набори даних містять ключовий код для кожного запису в даних, щоб усі дані можна було об'єднати, коли це необхідно.



## Наступні кроки

Обидва файли містять ключовий код, який дозволить знову об'єднати їх. Одним зі способів зробити це в Excel є використання [функції VLOOKUP](#) для автоматичного заповнення комірок на основі значення інших комірок. У цьому випадку можна заповнити пусті комірки у вихідному файлі відсутньою ІО на основі значення **ключового коду**.

Оскільки новий файл містить прямі ідентифікатори з ІО, його потрібно надійно зберігати. Один чудовий спосіб зробити це — зашифрувати файл і використовувати хмарне сховище, щоб обмежити кількість тих, хто має доступ до файлу (дивіться посібники з найкращих практик шифрування та обміну файлами).

**Пам'ятайте:** хоча оригінальна електронна таблиця була знеособлена шляхом видалення прямих ідентифікаторів, які містять очевидну ІО, інші непрямі ідентифікатори можуть бути об'єднані з іншими даними або проаналізовані таким чином, що це дозволить ідентифікувати особу.



З цієї причини обидва файли слід надійно зберігати. Якщо ви хочете надати ширший доступ до оригінального файлу, без ІО, вкрай важливо виконати *оцінку ризиків розкриття даних*, щоб забезпечити мінімальний ризик повторної ідентифікації даних. Центр гуманітарних даних має [навчальний онлайн-посібник для проведення оцінки ризиків розкриття даних](#) із використанням [статистичного програмного забезпечення "R" із відкритим вихідним кодом](#). Крім того, веб-сторінка [Poverty Action Lab "Знеособлення для публікації даних"](#) містить чудове обговорення знеособлення даних і зразок коду для [статистичного програмного забезпечення Stata](#). Співробітники Mercy Corps мають внутрішній доступ до [Проекта настанов від T4D](#), де є додаткові формули Excel.

Нарешті, всі ці кроки разом допомагають зменшити ризик розкриття ІО, тому вони повинні бути перераховані в ОВК (дивіться [посібник з оцінки впливу на конфіденційність](#)), щоб інші розуміли, яким чином ці дані захищені.

---

## Додаткова підтримка

Знеособлення даних є частиною практик належного керування даними та більшого життєвого циклу даних, що є загальною діяльністю для індивідуального збору даних як складової програми або відгуку. Наступні ресурси є відмінним початком для повнішого розуміння управління вашими даними відповідальним чином.

- › "Інструментарій відповідальності за дані" від Cash Learning Partnership призначений спеціально для тих, хто використовує готівку та ваучери, але є золотим стандартом управління відповідальним використанням даних. Інструментарій доступний [англійською](#), [арабською](#), [французькою](#) й [іспанською](#) мовами.
- › "[Набір для початківців про дані для працівників гуманітарної сфери](#)" від Electronic Cash Transfer Learning Action Network містить низку порад щодо даних для розуміння різних аспектів практик належного керування даними та їх захисту.
- › "[Довідник із захисту даних у гуманітарній діяльності](#)" від Міжнародного Комітету Червоного Хреста є докладним посібником майже з усіх аспектів гуманітарних даних. Глава 2 спеціально присвячена знеособленню даних.
- › "[Довідник спеціаліста із сучасного розвитку](#)" від Engine Room — хороший огляд питання даних у контексті міжнародної діяльності з розвитку. Розділ про "[Обмін даними](#)" спеціально присвячений знеособленню.

## КОНТАКТИ

ХІЗЕР ЛАВ (HEATHER LOVE)

Директор із глобального захисту та конфіденційності даних | IT  
hlove@mercycorps.org

ШЕДРОК РОБЕРТС (SHADROCK ROBERTS)

Спеціаліст із захисту даних | IT  
shroberts@mercycorps.org

## Про Mercy Corps

Mercy Corps — провідна всесвітня організація, яка спирається на віру в те, що кращий світ можливий. Під час катастроф і негод, у більш ніж 40 країнах по всьому світу ми співпрацюємо, щоби ввести сміливі рішення в дію, допомагаючи людям перемогти негаразди й побудувати сильніші громади зсередини. І зараз, і на майбутнє.



### Всесвітня штаб-квартира

45 SW Ankeny Street  
Портленд, Орегон 97204  
888 842 0842  
[mercycorps.org](https://mercycorps.org)

### Європейська штаб-квартира

40 Sciences  
Единбург EH9 1NJ  
Шотландія, Великобританія  
+44 131 662 5160  
[mercycorps.org.uk](https://mercycorps.org.uk)