

*Data
Protection is
People
Protection*



РУКОВОДСТВО ПО ЗАЩИТЕ ДАННЫХ И КОНФИДЕНЦИАЛЬНОСТИ:

Обезличивание (деидентификация) данных

В этом руководстве приведен пример удаления из набора данных информации, позволяющей установить личность (PII). Существует несколько способов «деидентификации» данных, которые относятся к действиям или методам обработки, помогающим предотвратить идентификацию личности Субъекта данных. Двамя распространенными типами деидентификации являются «Анонимизация» и «Псевдонимизация».

Анонимизация — процесс, посредством которого персональные данные становятся анонимными, а человек (или «субъект данных») больше не может быть идентифицирован: это постоянное изменение данных. Распространенные методы включают в себя удаление информации, позволяющей установить личность, или шифрование значений в определенных наборах ПИ.

Пример: представьте, что у организации есть данные опроса, которые содержат такие поля, как имя, государственный идентификационный код, название населенного пункта, этническая принадлежность, возраст, уровень образования и показатели здоровья. В этом случае удаление имени и государственного идентификационного кода будет первым шагом к обеспечению анонимности данных, поскольку эти «непосредственные характеристики» являются личными данными, которые прямо идентифицируют человека. Останутся «косвенные характеристики», такие как название населенного пункта, этническая принадлежность, возраст, уровень образования, показатели здоровья.

Однако даже если некоторые характеристики кажутся «анонимными», они могут таковыми не быть. Если бы опрос проводился в очень маленькой деревне, где только два жителя имеют определенную этническую принадлежность, и при этом они разного возраста, то использование этих двух косвенных характеристик могло бы позволить идентифицировать этих людей! Процесс, в ходе которого проверяются все характеристики для снижения риска повторной идентификации субъекта данных, называется Статистическим контролем раскрытия информации. Первым шагом в этом процессе является оценка риска раскрытия информации, и в Центре гуманитарных данных есть [учебное онлайн-руководство по проведению оценки риска раскрытия информации](#).

Псевдонимизация, с другой стороны, подразумевает обработку персональных данных, после которой более невозможно установить принадлежность персональных данных конкретному субъекту данных без использования дополнительной информации, такой как ключевой код.

Пример: представьте, что опрос содержит ваше имя, адрес электронной почты, возраст, национальность и место работы. Псевдонимизация берет данные, которые можно идентифицировать конкретно о вас (ваше имя, адрес электронной почты, возраст), и делает их недоступными и отделенными от неидентифицирующих данных, таких как ваша национальность. Псевдонимные данные можно в какой-то момент восстановить, чтобы всю информацию можно было связать с конкретным источником или человеком. Вот почему псевдонимизация требует, чтобы дополнительная информация хранилась отдельно, и были приняты технические и организационные меры, обеспечивающие невозможность установления взаимосвязи указанных данных с субъектом данных.

Что выбрать: анонимизацию или псевдонимизацию?

Анонимизация, как правило, более безопасна и снижает риск раскрытия ПИ. Однако иногда этот процесс может слишком обобщать данные, что может сделать их бесполезными для таких программ, как оказание помощи денежными ваучерами. Если говорить о программах здравоохранения, которые включают в себя вакцинацию или другие процедуры, то может возникнуть потребность связаться с человеком для обсуждения последующего лечения. В обоих этих случаях псевдонимизация была бы лучшим вариантом, поскольку вы всегда можете восстановить целостность данных, чтобы идентифицировать человека, когда это необходимо.

Нет единого правильного ответа о том, когда лучше выбрать тот или иной метод. Прежде чем выбрать способ деидентификации ваших данных, важно понять, почему данные были собраны, осознать потенциальные риски, связанные с хранением этих данных, а также требования программы.

Также важно понимать, что методы, используемые как для анонимизации данных, так и для взлома данных, становятся все более изощренными и что [даже обезличенные данные не всегда защищены на сто процентов](#). В случае сомнений обратитесь за помощью к своей команде по обработке данных или IT-команде.

☆ Важность

Недавние утечки данных в Международном комитете Красного Креста, взломы электронной почты в Агентстве США по международному развитию и неправильный обмен данными со стороны Верховного комиссара ООН по делам беженцев указывают на несколько путей, через которые гуманитарные данные подвергаются риску. Данные опросов домохозяйств, оценки потребностей и другие формы микроданных составляют все более значительный объем информации в гуманитарном секторе. Эти типы данных имеют решающее значение для определения потребностей и перспектив участников программы и сообществ, в которых мы работаем, но эти данные также представляют риски. Понимание того, как оценивать и управлять конфиденциальностью этих данных, абсолютно необходимо для обеспечения их безопасного, этичного и эффективного использования в различных контекстах реагирования.

Среди преимуществ использования анонимных данных по сравнению с личными данными - следующие:

- › защита от неправомерного раскрытия персональных данных;
- › к анонимным данным применяется меньше юридических ограничений; и
- › это позволяет организациям создавать открытые или общедоступные данные, соблюдая при этом свои обязательства по защите данных.

📄 Принципы

Обезличивание и обработка персональных данных, осуществляемая гуманитарными организациями, должна соответствовать следующим принципам.

- › **Справедливость и законность обработки:** методы должны соответствовать региональному, национальному или местному законодательству или политикам, которые могут ограничивать тип данных, подходящих для обезличивания, и возможности использования определенных технологий. Любая обработка персональных данных должна быть прозрачной для вовлеченных субъектов данных.
- › **Ограничение цели:** гуманитарные организации должны определить и указать конкретные цели, для которых обрабатываются данные. Эти цели должны быть явными и законными.
- › **Соразмерность:** гарантия, что каждое конкретное действие, связанное с обработкой персональных данных, соответствует заявленной цели. Например: собирается ли только минимально необходимый объем данных? Принимаются ли соответствующие технические и организационные меры для снижения рисков, связанных с обработкой данных?
- › **Технологические изменения:** новые наборы данных и новые инструменты для их анализа быстро меняются и совершенствуются, как и средства взлома или кражи данных. Важно понимать новые и возникающие риски для ваших данных и непрерывно корректировать свои методы и практики соответствующим образом.

Псевдонимизация

Это пример одного из способов деидентификации данных в электронной таблице. Существует множество способов деидентификации, и в этом примере используется «код ключа» для удаления личной информации, содержащейся в прямых идентификаторах, и сохранения ее в отдельном файле. Персональные данные, позволяющие идентифицировать личность (PII) — это информация, которая может быть использована для установления личности человека. Типичными примерами являются имя, адрес, номер телефона, дата рождения и номер социального страхования или государственный идентификационный код.



Инструкции

Вы можете следовать этим [Инструкциям по псевдонимизации](#), чтобы пошагово рассмотреть базовый пример псевдонимизации набора данных. В упражнении используется [демонстрационный набор данных, который находится в папке данных онлайн-руководства](#).

После псевдонимизации демонстрационного набора данных вы можете продолжить работу с учебным пособием Центра гуманитарных данных [для проведения оценки риска раскрытия информации](#).

Шаг 1. Определите PII

Начните с идентификации PII в данных. В идеале у вас должны быть метаданные — данные или документ, определяющий ваши данные, — чтобы помочь вам понять, какие поля содержат PII. В демонстрационных данных есть три столбца, которые содержат потенциальные PII:

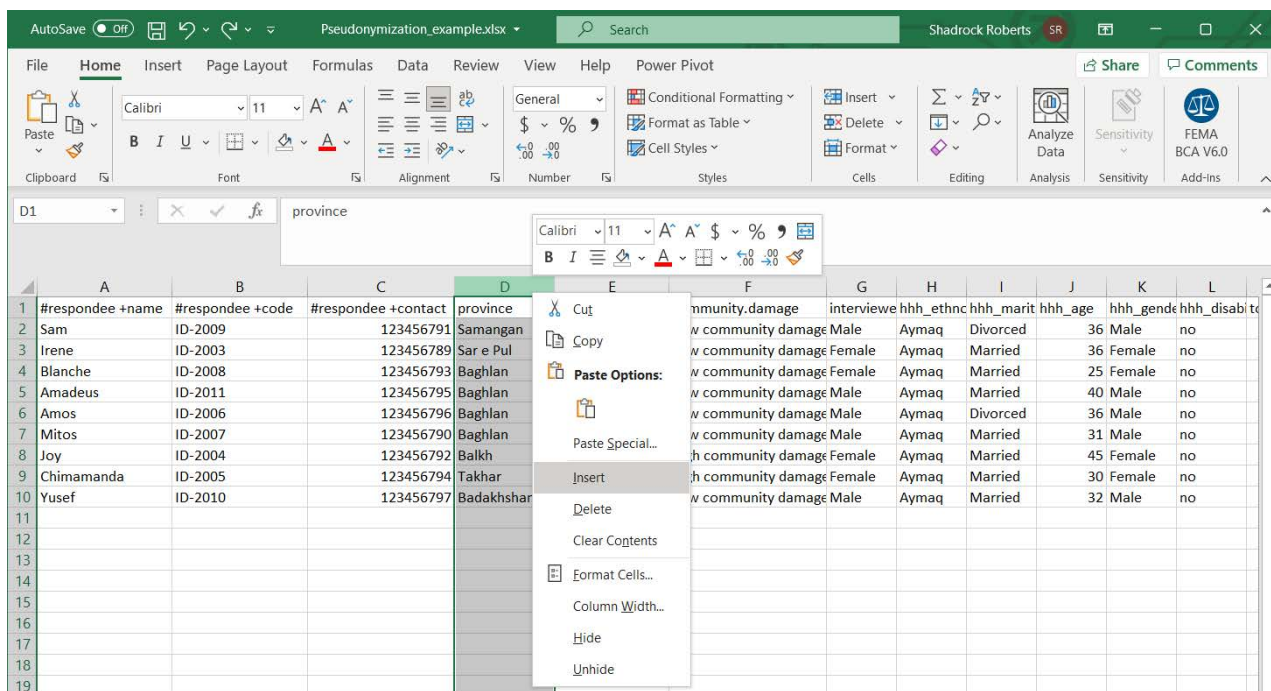
- › #респондент +имя - содержит имя.
- › #респондент +код - вероятно, содержит какой-либо идентификационный код.
- › #респондент +контакт - возможно содержит номер мобильного телефона

Каждый из этих прямых идентификаторов использует [Гуманитарный язык обмена для маркировки данных](#).

	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_disablt
2	Sam	ID-2009	123456791	Samangan	Khulum	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	Khulum	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												

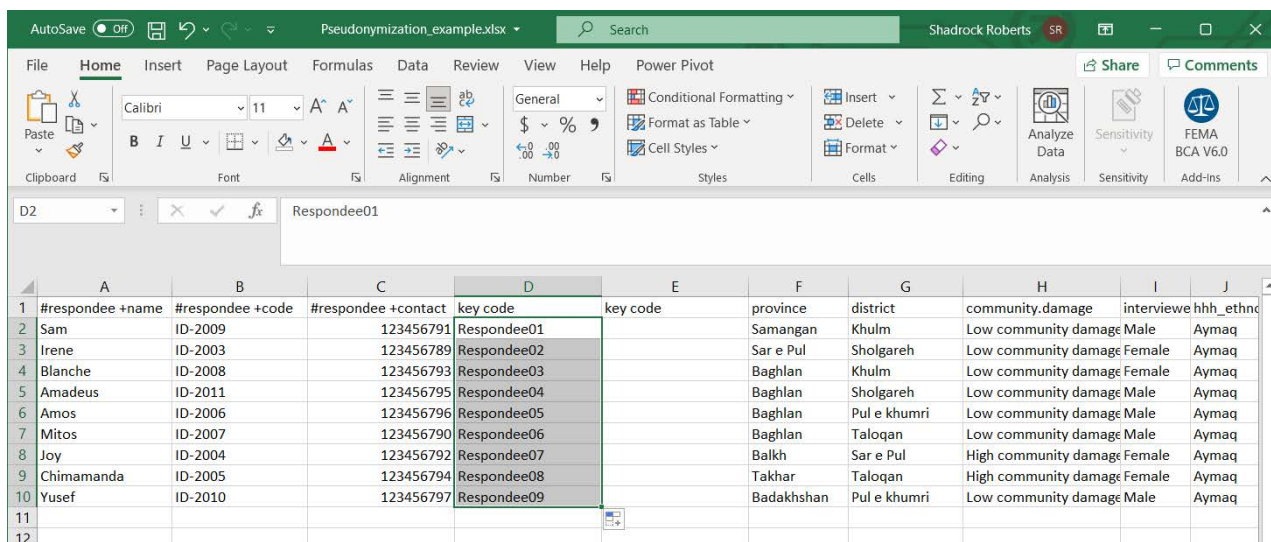
Шаг 2. Создайте новые столбцы для кода ключа

Мы будем использовать код ключа, значение, которого мы сгенерируем, чтобы раскрыть PII. Поскольку все прямые идентификаторы сгруппированы вместе, мы создадим два новых столбца между столбцами C, **#респондент +контакт** и D, **область**. В Excel мы делаем это, выделяя столбец справа от того места, куда мы хотим вставить новые столбцы, щелкаем правой кнопкой мыши на столбце и выбираем **Вставить**. Повторите этот процесс еще раз, чтобы создать еще один пустой столбец.



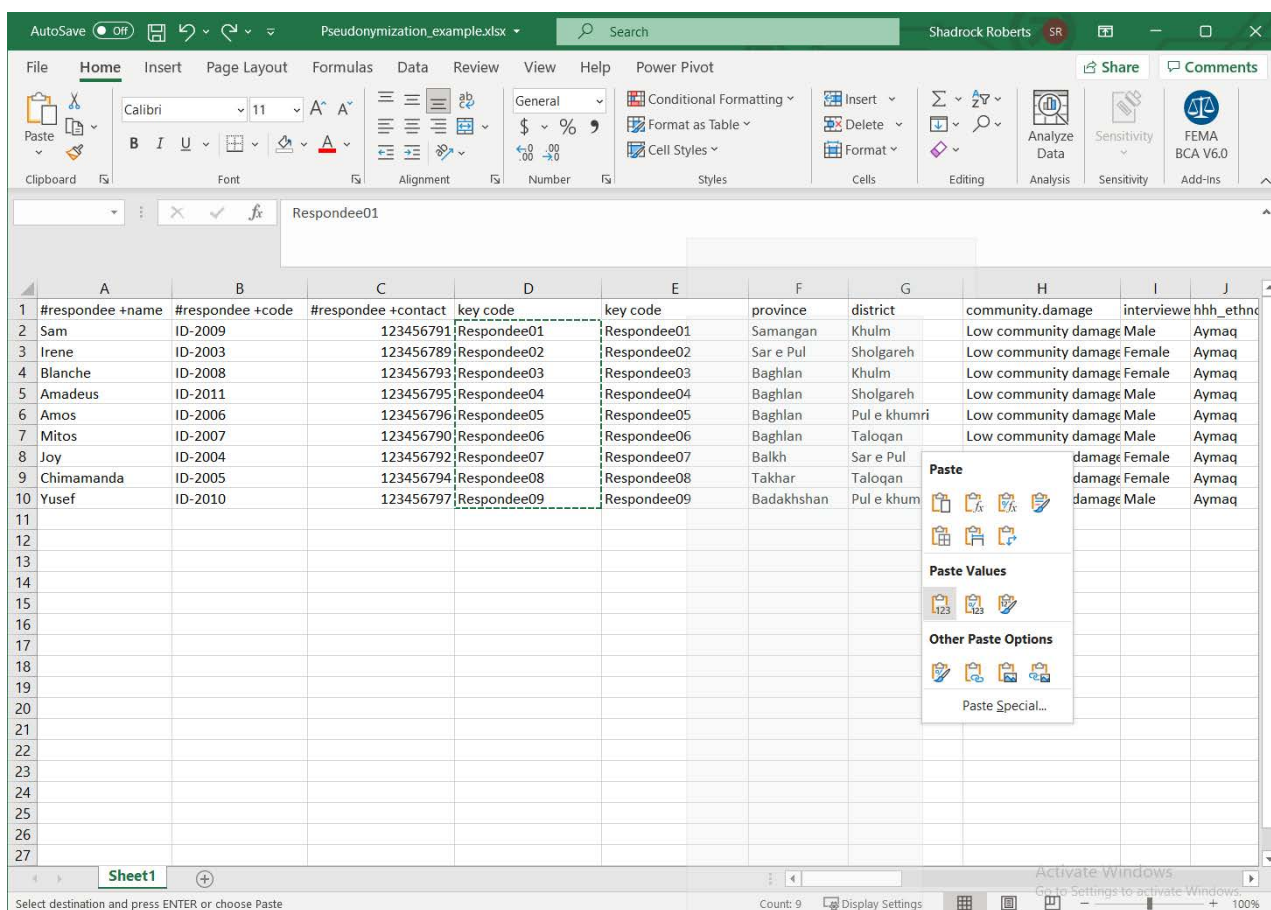
Шаг 3 — Создайте код ключа

Начните с названия новых столбцов. Мы будем использовать «ключевой код» в каждом из них: каждый столбец будет содержать одинаковые значения. Сейчас самое время обновить любые метаданные об этом наборе данных, чтобы объяснить, что означает «ключевой код»! Далее мы воспользуемся **функцией Excel Автозаполнение** для создания простого кода. Введите **Респондент01** в первой ячейке. Затем выделите эту ячейку, щелкните указатель перемещения в правом нижнем углу ячейки и перетащите вниз до конца набора данных. Это автоматически заполнит окончательный номер каждой записи, присвоив каждому респонденту новый код.



Шаг 4. Продублируйте код ключа и удалите формулы

Теперь мы скопируем код ключа и вставим его в соседний столбец. Это можно сделать с помощью основных клавиатурных команд, таких как **ctrl + C** или выделить ячейки, которые нужно скопировать, щелкнуть по ним правой кнопкой мыши и выбрать **Копировать**. В соседнем столбце выделите ячейки, в которые вы хотите вставить новый код ключа, щелкните правой кнопкой мыши и выберите **Вставить**. Мы специально выбрали вставить только значения. Если вы использовали формулу для создания нового кода, важно сохранить только значения для использования в качестве кода ключа!



The screenshot shows the Microsoft Excel interface with a spreadsheet titled 'Pseudonymization_example.xlsx'. The spreadsheet has columns labeled A through J. A 'Paste' context menu is open over the 'key code' column (column D). The menu options include 'Paste', 'Paste Values', and 'Other Paste Options'. The spreadsheet data is as follows:

	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 khumri	damage	Male	Aymaq
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

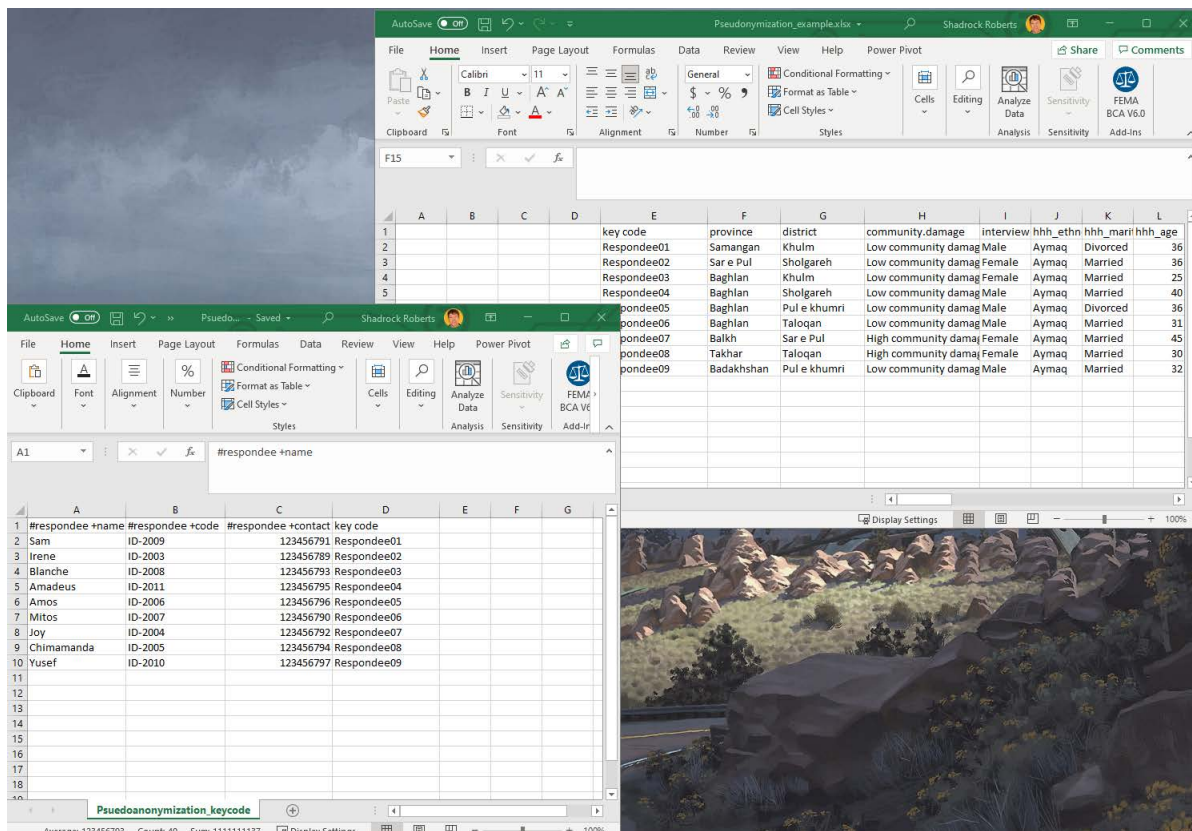
Шаг 5. Разделите прямые и косвенные идентификаторы

Выделите столбцы, содержащие прямые идентификаторы с PII, вместе с одним из столбцов кода ключа. В этом примере мы выделяем столбцы A-D. Щелкните их правой кнопкой мыши и выберите **Вырезать**.

The screenshot shows the Microsoft Excel interface with a spreadsheet titled 'Pseudonymization_example.xlsx'. The ribbon is set to 'Home'. A data table is visible with columns A through L. Columns A, B, C, and D are highlighted in grey, indicating they are selected. A context menu is open over column D, showing options such as Cut, Copy, Paste Options, Paste Special..., Insert, Delete, Clear Contents, Format Cells..., Column Width..., Hide, and Unhide. The data in the table includes names, IDs, contact numbers, provinces, and various demographic and status attributes.

#respondee +name	#respondee +code	#respondee +contact	province	community.damage	interview	hhh_ethnc	hhh_marit	hhh_age	hhh_gende	hhh_disablt
Sam	ID-2009	123456791	Samangan	w community damage	Male	Aymaq	Divorced	36	Male	no
Irene	ID-2003	123456789	Sar e Pul	w community damage	Female	Aymaq	Married	36	Female	no
Blanche	ID-2008	123456793	Baghlan	w community damage	Female	Aymaq	Married	25	Female	no
Amadeus	ID-2011	123456795	Baghlan	w community damage	Male	Aymaq	Married	40	Male	no
Amos	ID-2006	123456796	Baghlan	w community damage	Male	Aymaq	Divorced	36	Male	no
Mitos	ID-2007	123456790	Baghlan	w community damage	Male	Aymaq	Married	31	Male	no
Joy	ID-2004	123456792	Balkh	w community damage	Female	Aymaq	Married	45	Female	no
Chimamanda	ID-2005	123456794	Takhar	w community damage	Female	Aymaq	Married	30	Female	no
Yusef	ID-2010	123456797	Badakhshar	w community damage	Male	Aymaq	Married	32	Male	no

Затем откройте новую таблицу и вставьте эти значения с помощью сочетания клавиш **ctrl + V** или другим способом. Сохраните новую таблицу. Теперь у вас есть две таблицы: одна из них содержит косвенные идентификаторы, а новая таблица содержит прямые идентификаторы с ПИ. Оба набора данных содержат ключевой код для каждой записи в данных, чтобы при необходимости все данные можно было объединить.



Дальнейшие шаги

Оба файла содержат ключевой код, который позволит их собрать вместе. Один из способов сделать это в Excel — использовать [функцию VLOOKUP](#) для автоматического заполнения ячеек на основе значений других ячеек. В этом случае вы можете заполнить пустые ячейки исходного файла отсутствующими идентификационными данными на основе значения **ключевого кода**.

Поскольку новый файл содержит прямые идентификаторы, содержащие ПИ, он должен храниться в безопасном месте. Отличный способ сделать это — зашифровать файл и использовать облачное хранилище для ограничения доступа к файлу (см. [руководства по Лучшим практикам шифрования и совместного доступа к файлам](#)).

Помните: даже при том, что исходная электронная таблица была деидентифицирована путем удаления прямых идентификаторов, содержащих очевидную ПИ, другие косвенные идентификаторы могут быть объединены с другими данными или проанализированы таким образом, чтобы можно было идентифицировать человека.

По этой причине оба файла должны храниться в безопасном месте. Если вы хотите предоставить более широкий доступ к исходному файлу, не содержащему РИ, крайне важно выполнить оценку риска раскрытия информации, чтобы обеспечить минимальный уровень риска повторной идентификации данных. В Центре гуманитарных данных есть [онлайн-руководство по проведению оценки рисков раскрытия информации](#) с использованием [статистического программного обеспечения с открытым исходным кодом “R”](#). Кроме того, на веб-странице [Poverty Action Lab Деидентификация для публикации данных](#) содержится отличное обсуждение обезличивания данных и пример кода для [статистического программного обеспечения Stata](#). Для внутреннего использования сотрудниками Mercy Corps доступен [Проект руководства от T4D](#), который содержит дополнительные формулы Excel.

Наконец, все эти шаги вместе помогают снизить риски или раскрытие РИ, поэтому они должны быть перечислены в РИА (см. [Руководство по оценке влияния на конфиденциальность](#)), чтобы другие понимали, как эти данные защищаются.

Дальнейшая помощь

Деидентификация данных является частью надлежащей практики управления данными и более широкого жизненного цикла данных, который представляет собой общие действия по сбору отдельных данных в рамках программной деятельности или реагирования. Следующие ресурсы можно использовать как эффективные отправные точки для более полного понимания ответственного управления вашими данными.

- › Инструментарий ответственного использования данных от Cash Learning Partnership разработан специально для специалистов, работающих с наличными и ваучерами, и он является золотым стандартом в руководстве по ответственному данным. Инструментарий доступен на [английском](#), [арабском](#), [французском](#) и [испанском](#).
- › [Стартовый комплект данных для полевого гуманитарного персонала](#) Сообщества для активного обучения электронным денежным переводам (Electronic Cash Transfer Learning Action Network) содержит материалы для разъяснения различных аспектов надлежащего управления данными и методов защиты.
- › [Руководство по защите данных в гуманитарной деятельности](#) Международного комитета Красного Креста представляет собой подробное руководство практически по всем аспектам гуманитарных данных. Глава 2 посвящена обезличиванию данных.
- › [Справочник современного специалиста в области развития](#) от Engine Room — хороший обзор данных в контексте международной деятельности в области развития. В разделе [Совместный доступ к данным](#) конкретно рассматривается деидентификация.

КОНТАКТЫ

ХИЗЕР ЛАВ (HEATHER LOVE)

Директор по глобальной защите данных и конфиденциальности | IT
hlove@mercycorps.org

ШЕДРОК РОБЕРТС (SHADROCK ROBERTS)

Специалист по защите данных | IT
shroberts@mercycorps.org

О Mercy Corps

Mercy Corps — ведущая международная организация, в основе работы которой лежит вера в возможность сделать этот мир лучше. В бедствиях и в трудностях, в более чем 40 странах мира мы трудимся, чтобы претворять в жизнь смелые решения, помогая людям преодолевать невзгоды и строить более сильные изнутри сообщества. Сейчас и для будущего.



Международная штаб-квартира

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

Европейская штаб-квартира

40 Sciences
Edinburgh EH9 1NJ
Scotland, UK
+44 131 662 5160
mercycorps.org.uk