Приложение

к Приказу Государственного
агентства по защите персональных
данных при Кабинете Министров
Кыргызской Республики
от «14» января 2025 года № 4

Порядок обезличивания персональных данных для проведения статистических, социологических, исторических, медицинских и других научных и практических исследований

1. Общие положения

1. Настоящий Порядок устанавливает порядок обезличивания персональных данных в соответствии со статьей 26 и 27 Закона Кыргызской Республики «Об информации персонального характера» с целью использования в статистических, социологических, исторических, медицинских и других научных и практических исследованиях.

2. Настоящий Порядок применяется ко всем операциям обезличивания персональных данных, осуществляемым держателями (обладателями) массивов персональных данных.

3. В случае обезличивания персональных данных режим конфиденциальности персональных данных снимается. Если набор данных не содержит персональной информации, его использование или раскрытие не может нарушать конфиденциальность.

4. В настоящем Порядке применяются следующие термины и определения:

Информация персонального характера (персональные данные) - зафиксированная информация на материальном носителе о конкретном человеке, отождествленная с конкретным человеком или которая может быть отождествлена с конкретным человеком, позволяющая идентифицировать этого человека прямо или косвенно, посредством ссылки на один или несколько факторов, специфичных для его биологической, экономической, культурной, гражданской или социальной идентичности.

К персональным данным относятся биографические и опознавательные данные, личные характеристики, сведения о семейном положении, финансовом положении, состоянии здоровья и прочее.

Субъект персональных данных (субъект) - физическое лицо, к которому относятся соответствующие персональные данные.

Держатель (обладатель) массива персональных данных - органы государственной власти, органы местного самоуправления и юридические лица, на которые возложены полномочия определять цели, категории

персональных данных и контролировать сбор, хранение, обработку и использование персональных данных в соответствии с настоящим Законом.

Обезличивание персональных данных – изъятие из персональных данных той их части, которая позволяет отождествить их с конкретным человеком.

Прямые идентификаторы — информация, которая непосредственно указывает на личность человека (имя, адрес, номер телефона, номер социального страхования и т. д.).

Косвенные или квази-идентификаторы (quasi-identifier) — информация, которая сама по себе не идентифицирует человека, но в сочетании с другой информацией может быть использована для его идентификации. Это может быть, например, дата рождения, почтовый индекс, пол, этническая принадлежность, а также другие характеристики, которые могут быть использованы для чтения информации о человеке.

2. Цели обезличивания персональных данных

5. Целями обезличивания являются:

1) Защита конфиденциальности: обеспечение конфиденциальности персональных данных и предотвращение их идентификации;

2) Соответствие законодательству: соблюдение национальных и международных стандартов защиты данных;

3) Безопасный обмен данными: обезличенные данные могут безопасно обмениваться между держателями (обладателями) массивов персональных данных без риска нарушения конфиденциальности;

4) Улучшение качества данных для исследований: обеспечение возможности использования больших объемов данных для анализа без нарушения прав субъектов данных;

5) Снижение рисков утечек: уменьшение вероятности утечек персональной информации и угроз безопасности.

3. Задачи организации порядка обезличивания персональных данных

6. Задачами организации порядка являются:

1) Идентификация и классификация данных: определение и классификация данных, требующих обезличивания, особенно чувствительных категорий, таких как медицинские, финансовые и биометрические данные, на основе их значимости для личности и уровня конфиденциальности;

2) Разработка и применение методов обезличивания: определение подходящих методов обезличивания для различных типов данных и контекстов их использования, обеспечивающих баланс между защитой персональных данных и сохранением полезности данных. Включение

современных подходов, таких как дифференциальная конфиденциальность и многоуровневая анонимизация;

3) Мониторинг и аудит: установление системы регулярного мониторинга и аудита процессов обезличивания для выявления потенциальных угроз, улучшения механизмов защиты данных и обеспечения соответствия действующему законодательству и международным стандартам;

4) Обучение и повышение осведомленности сотрудников: проведение обучения и информационных кампаний среди сотрудников организации для повышения их осведомленности о важности, методах и правилах обезличивания данных;

5) Процедуры реагирования на инциденты: разработка и внедрение процедур реагирования на инциденты, связанные с утечкой обезличенных данных, для обеспечения своевременного устранения последствий и минимизации ущерба.

4. Принципы обезличивания персональных данных

7. Обезличивание персональных данных осуществляется с соблюдением следующих принципов:

1) Процедуры обезличивания должны основываться на технических и организационных мерах, обеспечивающих невозможность восстановления персональных данных;

2) Изъятие из данных той их части, которая позволяет идентифицировать физическое лицо;

3) Удаление прямых и косвенных идентификаторов;

4) Обезличивание должно исключать возможность идентификации субъекта персональных данных;

5) При проведении процедуры обезличивания персональных данных держатель (обладатель) массивов персональных данных выбирает наиболее перспективное и рациональное для практического применения метод обезличивания персональных данных;

6) Процессы обезличивания должны быть прозрачными для субъектов данных и контролирующих органов, с четким определением ответственности внутри организации и назначением лиц, ответственных за выполнение процедур обезличивания. Организации должны быть готовы объяснить, какие методы обезличивания были использованы, и демонстрировать их эффективность;

7) Устойчивость к повторной идентификации и этичность. Применение методов, устойчивых к атакам повторной идентификации, с регулярным пересмотром и обновлением методов в соответствии с технологическими и законодательными изменениями. Все действия должны осуществляться с соблюдением прав и законных интересов субъектов данных.

8) Применение нескольких методов обезличивания в комплексе для усиления защиты данных, гарантируя комплексную защиту персональной информации.

5. Условия обезличивания персональных данных

8. Обезличивание персональных данных проводится с целью ведения статистических, социологических, исторических, медицинских и других научных и практических исследований данных, а также в целях обработки больших данных в случае достижения целей обработки персональных данных или в режиме реального времени при условии, что такие данные обрабатываются в обезличенном виде, а у держателя (обладателя) массивов персональных данных, как и у обработчика имеются законные основания на такую обработку.

Обезличивание персональных данных осуществляется при помощи действий, в результате которых становится невозможным без использования дополнительной информации определить принадлежность персональных данных конкретному субъекту персональных данных, в случае утраты необходимости в достижении этих целей, если иное не предусмотрено законодательством.

9. Держатель (обладатель) массива персональных данных принимает решение о необходимости обезличивания персональных данных исходя из целей и потребностей по обезличиванию персональных данных при наличии законных оснований.

10. Держатель (обладатель) массива персональных данных, осуществляет обезличивание выбранных персональных данных, с соблюдением условий конфиденциальности обезличиваемых персональных данных.

11. Ответственность за надлежащую организацию и проведение обезличивания лежит на держателе (обладателя) массива персональных данных. Держатели (обладатели) массива персональных данных организуют и проводят обезличивание, обеспечивая соблюдение всех необходимых мер безопасности.

12. Держатель (обладатель) массива персональных данных, должны проводить регулярные проверки и мониторинг эффективности применяемых методов обезличивания персональных данных с целью подтверждения их соответствия установленным требованиям и обеспечения должного уровня защиты персональной информации.

6. Методы обезличивания

13. Обезличивание применяется в сфере защиты данных и конфиденциальности, позволяя использовать данные в анализе или публикации, не раскрывая личную информацию о субъектах данных.

14. При обезличивании персональных данных применяются следующие методы:

- метод введения идентификаторов (условных обозначений) – это замена части сведений (значений персональных данных) идентификаторами с созданием таблицы (справочника) соответствия идентификаторов исходным данным, т.е. применение шифрования с кодами;

- метод изменения состава или семантики – изменение состава или семантики персональных данных путем замены результатами статистической обработки, преобразования, обобщения или удаления части сведений;

- метод декомпозиции данных – разбиение множества (массива) персональных данных на несколько подмножеств (частей) с последующим раздельным хранением подмножеств, каждый из которых по отдельности не несет значимой информации и практического смысла;

- метод перемешивания данных – перестановка отдельных записей, а также групп записей в массиве персональных данных, по результатам которой невозможно дальнейшее установление принадлежности персональных данных конкретному лицу;

- метод использования генератора случайных чисел – обезличивание персональных данных заключается в замене идентифицирующих данных случайно сгенерированными значениями;

- метод агрегации данных – объединение данных в группы или наборы, что делает невозможным идентификацию отдельного субъекта данных. Информация отображается на уровне группы, а не отдельных лиц;

- метод добавления шума – изменение атрибутов в наборе данных таким образом, чтобы они были менее точными, сохраняя при этом общее распределение;

- метод размывания (для изображений) – техника, направленная на использование приближенных значений данных, чтобы уменьшить точность данных, снижая возможность идентификации личности;

- метод обобщения данных – приведение данных к более общим категориям (например, замена точной даты рождения на год рождения);

- другие методы и их комбинации.

Методы обезличивания могут использоваться индивидуально или комбинироваться для повышения эффективности обезличивания данных и предотвращения возможной идентификации субъектов данных.

7. Метод введения идентификаторов (условных обозначений)

15. Метод введения идентификаторов (условных обозначений) при обезличивании персональных данных реализуется путем процесса замены идентифицирующих данных субъекта на уникальные псевдонимы или условные обозначения.

Персональные данные, такие как имя, фамилия, адрес и другие уникальные идентификаторы, заменяются на уникальные условные обозначения (псевдонимы). Псевдонимы не содержат личной информации, но могут быть использованы для дальнейшей связи с реальными данными в случае необходимости.

Важно, чтобы данные, позволяющие идентифицировать субъектов, и сами псевдонимы хранились в разных базах данных. Это минимизирует риск утечки личной информации, даже если одна из баз данных будет скомпрометирована.

Применение данного метода позволяет получить обезличенные данные обладающие следующими свойствами:

– полнота: информация, позволяющая идентифицировать субъектов персональных данных, не удаляется, а переносится в таблицу соответствия;

– структурированность: каждому идентификатору после процедуры обезличивания однозначно соответствует свой набор данных;

– семантическая целостность: вид представления данных не меняется, они лишь переносятся в таблицу соответствия.

Доступ к ключам или информации, которая может восстановить связь между идентификатором и отображением личности, должен быть строго ограничен. Лишь ограниченный круг лиц, имеющих разрешение, может иметь доступ к этим ключам. Такой доступ должен контролироваться и документироваться.

Метод введения идентификаторов (условных обозначений) целесообразно применять при небольшом количестве атрибутов персональных данных и небольшом объеме массива персональных данных, в связи с тем, что объем справочников будет напрямую зависеть от этих параметров. Вычислительная эффективность метода значительно снижается при частом внесении изменений в состав данных и значения атрибутов.

Для максимальной защиты метод введения идентификаторов (условных обозначений) необходимо сочетать с другими методами, такими как шифрование и строгие политики управления доступом к ключам.

Пример обезличивания по данному методу указан в Образце 1.

8. Метод изменения состава или семантики

16. Метод реализуется путем обобщения, изменения значений атрибутов персональных данных или удаления части сведений, позволяющих идентифицировать субъекта.

Применение данного метода позволяет получить обезличенные данные, обладающие следующими свойствами:

– структурированность: связь между отдельными значениями атрибутов персональных данных субъекта не нарушается;

– анонимность: удаление или обобщение части данных приводит к неоднозначности при идентификации с использованием обезличенных данных.

Полученные обезличенные данные могут обладать свойством полноты только при проведении изменений в составе персональных данных, гарантирующих сохранность данных. При удалении части сведений, полученные обезличенные данные утрачивают свойство полноты.

Семантическая целостность полученных данных обеспечивается только при условии проведения изменений в составе персональных данных, сохраняющих семантику данных. Изменения должны учитывать специфику задач обработки, стоящих перед держателем (обладателем) данных.

При выделении атрибутов персональных данных необходимо учитывать возможность проведения обезличивания с использованием данных атрибутов. При простом изменении значений отдельных атрибутов обезличивание может не произойти, поскольку произойдет только изменение состава персональных данных.

Пример обезличивания по данному методу указан в Образце 2.

9. Метод декомпозиции данных

17. Метод реализуется путем разделения множества атрибутов персональных данных на несколько подмножеств и создания таблиц, устанавливающих связи между подмножествами (таблицы связей), с последующим раздельным хранением записей, соответствующих подмножествам этих атрибутов.

Применение данного метода позволит получить обезличенные данные, обладающие следующими свойствами:

– полнота: информация о субъектах персональных данных не удаляется, а переносится в другое хранилище;

– структурированность: сохраняется связь между записями в разных хранилищах, что позволяет однозначно сопоставлять их;

– семантическая целостность: семантика и вид представления данных о субъекте не изменяется;

– применимость: держатель (обладатель) массивов персональных данных может осуществлять обработку данных, расположенных в одном хранилище, как независимо от другого, так и при совместном их использовании.

Анонимность обеспечивается только при достаточно сложных связях между хранилищами и защите хранилищ от несанкционированного доступа.

Применение данного метода позволяет сохранить в записях каждого хранилища связи между атрибутами обезличенных данных, соответствующие связям между атрибутами персональных данных.

Метод декомпозиции целесообразно применять при большом количестве атрибутов персональных данных, но при достаточно редком внесении изменений в состав данных и значения атрибутов.

Пример обезличивания по данному методу указан в Образце 3.

10. Метод перемешивания данных

18. Метод реализуется путем перемешивания (перестановки) отдельных значений или групп значений атрибутов персональных данных между собой.

Применение данного метода позволит получить обезличенные данные, обладающие следующими свойствами:

– полнота: вся информация о субъектах персональных данных сохраняется;

– семантическая целостность: семантика и вид представления данных о субъекте не изменяется;

– анонимность: данные перемешиваются по каждому отдельному атрибуту записи о субъекте, что не позволяет без доступа к дополнительной (служебной) информации определить принадлежность тех или иных данных конкретному субъекту.

Применение данного метода не позволяет сохранить в записях связи между атрибутами обезличенных данных, соответствующие связям между атрибутами персональных данных.

Метод перемешивания целесообразно применять при большом количестве атрибутов персональных данных и большом объеме массива персональных данных

Метод перемешивания эффективен при необходимости сложной обработки персональных данных, частом внесении изменений в значения атрибутов.

Пример обезличивания по данному методу указан в Образце 4.

11. Метод использования генератора случайных чисел

19. Метод использования генератора случайных чисел применяется путем присвоения к атрибутам данных случайных чисел. Вместо реальных персональных данных (например, имя, фамилия, адрес, серия и номер паспорта или другие идентификаторы), используются случайные числа или символы, сгенерированные программным способом.

Случайно сгенерированные значения не имеют связи с реальными данными, что делает их практически невозможными для реидентификации без дополнительных данных.

Случайные числа могут генерироваться для каждого отдельного случая обработки данных. Они могут быть одноразовыми (генерируются для одноразового использования) или многократными (применяются в различных процессах, но всегда остаются случайными).

Для обеспечения высокой степени защиты персональных данных необходимо использовать надежные генераторы случайных чисел, которые обеспечивают высокую энтропию, чтобы случайные значения были действительно непредсказуемыми и уникальными.

Пример обезличивания по данному методу указан в Образце 5.

12. Метод агрегации данных

20. Метод агрегации при обезличивании данных направлен на защиту персональной информации, при котором данные об отдельных лицах объединяются в группы или обобщаются по категориям, что делает невозможным идентификацию конкретных субъектов. В процессе агрегации точные индивидуальные значения заменяются средними, суммарными или иными статистическими показателями.

Индивидуальные данные субъектов не отображаются, вместо этого используются агрегированные показатели, такие как среднее значение, медиана, максимумы и минимумы, или другие статистические метрики.

Значения атрибутов обобщаются, что каждый индивид имеет одно и то же значение. Например, путем уменьшения детализации местоположения с города до страны большее число субъектов данных. Индивидуальные даты рождения могут быть обобщены в диапазон дат или сгруппированы по месяцам или годам. Другие числовые атрибуты (например, зарплата, вес, рост или доза лекарства) могут быть обобщены интервальными значениями (например, вес 70 - 80 кг).

Метод агрегации является эффективным способом обезличивания данных, позволяющим обеспечить высокий уровень конфиденциальности, особенно при обработке больших массивов информации.

Пример обезличивания по данному методу указан в Образце 6.

13. Метод добавления шума

21. Метод добавления шума при обезличивании персональных данных применяется путем намеренного добавления к исходным данным случайные искажения или ошибки (шум) для снижения риска их реидентификации.

При добавлении шума важно, чтобы отклонение не было слишком большим, иначе это может исказить результаты анализа. Процесс добавления шума регулируется, чтобы сохранить баланс между обезличиванием и сохранением точности данных.

Шум может быть добавлен в виде случайных чисел, перестановки значений, изменения категорий, округления или других техник в зависимости от типа данных. Например, для числовых данных может использоваться добавление случайных значений, а для категориальных - случайная замена категории.

Добавление шума необходимо сочетать с другими методами обезличивания такими как удаление очевидных атрибутов и квазиидентификаторов. Уровень шума должен зависеть от необходимости требуемого уровня информации.

Пример обезличивания по данному методу указан в Образце 7.

14. Метод размывания (для изображений)

22. Метод размывания (для изображений) применяется путем обработки визуальных данных, при котором изображение или часть изображения (например, лица, регистрационные номера, документы) подвергается размытой фильтрации для предотвращения идентификации людей или объектов.

Степень размытости можно регулировать, чтобы сделать идентификацию невозможной.

Пример обезличивания по данному методу указан в Образце 8.

15. Метод обобщения данных

23. Метод обобщения данных применяется путем преобразования конкретных данных, позволяющих идентифицировать личность, в более общие или агрегированные формы. Это снижает вероятность, что конкретные данные могут быть использованы для реидентификации человека, и при этом сохраняет полезность данных для анализа и исследований.

Метод предполагает замену детализированной информации более обобщенными категориями. Например, вместо указания точного возраста человека можно использовать диапазон возрастов («от 30 до 40 лет»), а вместо точного адреса - указание только города или региона проживания.

Уровень обобщения может быть адаптирован в зависимости от цели обработки данных. Например, в одном случае может потребоваться более общее обобщение (например, на уровне страны), в другом — более детализированное (например, на уровне города).

При применении метода обобщения следует учитывать баланс между уровнем обобщения и потерей точности данных, а также необходимость дополнительных мер защиты для минимизации риска реидентификации.

Пример обезличивания по данному методу указан в Образце 9.

16. Заключительные положения

24. Приведенный выше список методов не является исчерпывающим, а представляет собой общее руководство по оценке возможности идентификации конкретного набора данных, подвергшегося обезличиванию в соответствии с различными доступными методами.

Держателем (обладателем) массива персональных данных могут использоваться собственные алгоритмы обезличивания персональных данных, которые позволяют обеспечить невозможность идентификации субъектов персональных данных при получении доступа к его обезличенным данным.

25. Допускается использование держателем всех средств, которые «с достаточной степенью вероятности» могут быть использованы для обезличивания, принимая во внимание то, что при современном уровне развития технологий, стало «вероятным использование наиболее разумного подхода» (учитывая вычислительные мощности и доступные инструменты).

26. Держателем (обладателем) массива персональных данных принимается локальный акт по обезличиванию персональных данных с указанием порядка и методики обезличивания, сотрудника ответственного за обезличивание персональных данных с установлением ответственности за соблюдение конфиденциальности обезличивания персональных данных, а также указанием обратного алгоритма персонификации персональных данных.

При использовании обезличенных персональных данных в автоматизированном режиме в структуре больших данных, когда используются массивы персональных данных нескольких держателей (обладателей) массивов персональных данных не допускается смешение таких массивов между собой. При этом у держателей должна быть возможность персонифицировать конкретный набор обезличенных персональных данных для реализации целей, закрепленных нормативными правовыми актами за держателем (обладателем) массива персональных данных.

27. При использовании держателем (обладателем) массива персональных данных процедуры обезличивания не допускается совместное хранение персональных данных и обезличенных данных.

28. Обработка персональных данных до осуществления процедур обезличивания и после выполнения операций реидентификации должна осуществляться в соответствии с действующим законодательством Кыргызской Республики с применением мер по обеспечению безопасности персональных данных.

Дополнительные источники по осуществлению обезличивания персональных данных:

Стандарты:

1. ISO/IEC 27000:2014 Information technology -- Security techniques -- Information security management systems -- Overview and vocabulary

2. 2. ISO/IEC 24760-1:2011 Information technology -- Security techniques -- A framework for identity management -- Part 1: Terminology and concepts

3. ISO/TS 25237:2008(E) Health Informatics — Pseudonymization. ISO, Geneva, Switzerland. 2008. This ISO Technical Specification describes how privacy sensitive information can be de-identified using a “pseudonymization service” that replaces direct identifiers with pseudonyms. It is provides a set of terms and definitions that are considered authoritative for this document

4. ISO/IEC 20889:2018 Privacy enhancing data de-identification terminology and classification of techniques;

5. ISO/IEC 29100 Information technology — Security techniques — Privacy framework;

Официальные публикации:

1. Opinion 05/2014 on Anonymisation Techniques, Article 29 Data Protection Working Party, 0829/14/EN WP216, Adopted on 10 April 2014 Introduction to anonymisation Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance;
https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

2. Серия X: Сети передачи данных, взаимосвязь открытых систем и безопасность Безопасные приложения и услуги (1) – Безопасность веб-среды Структура процесса деидентификации для поставщиков услуг электросвязи; Рекомендация МСЭ-Т X.1148;

3. NISTIR 8053 De-Identification of Personal Information Simson L. Garfinkel https://nvlpubs.nist.gov/nistpubs/ir/2015/nist.ir.8053.pdf

Пример обезличивания по методу введения идентификаторов (условных обозначений)

Таблица соответствий (ФЛ ↔ Идентификатор):

Фамилия	Дата Рождения	Уникальный Идентификатор
Асанов	01.01.1990	ID_001
Азаматов	15.05.1985	ID_002

База прочих данных (Идентификатор ↔ Прочие данные):

Уникальный Идентификатор	Имя	Адрес	Телефон
ID_001	Асанов	Город X, ул.1	123-456-7890
ID_002	Азаматов	Город Y, ул.2	987-654-3210

1. **Исходные данные**:
   - Личные данные (например, имя, фамилия, адрес, паспортные данные).
   - Эти данные являются идентифицирующими и должны быть защищены.

2. **Процесс замены идентифицирующих данных**:
   - Идентифицирующие данные заменяются на уникальные псевдонимы или условные обозначения.
   - Пример:
     - Имя: Асанов Данияр → Псевдоним: ID_001
     - Фамилия: Асангазиева Бурул → Псевдоним: ID_002

3. **Создание таблицы соответствий**:
   - Создается отдельная таблица, в которой хранится соответствие между псевдонимами и оригинальными данными.

Уникальный Псевдоним	Имя	Фамилия	Адрес
ID_001	Асан	Асанов	Город X, ул. 1
ID_002	Үсөн	Үсөнов	Город Y, ул. 2

4. **Разделение хранения данных**:
   - **Таблица соответствий** (с идентифицирующими данными):
     - Хранится отдельно от базы прочих данных.

- **База обезличенных данных**:
  - Содержит только псевдонимы без возможности идентификации.

Уникальный Псевдоним
ID_001
ID_002

5. **Ограниченный доступ к таблице соответствий**:
   - Доступ к ключам и таблице соответствий должен быть строго ограничен.
   - Лишь ограниченный круг лиц (например, администраторы) имеет доступ к этой информации.

6. **Сопоставление с другими методами защиты**:
   - Для повышения уровня безопасности метод введения идентификаторов можно сочетать с:
     - Шифрованием.
     - Строгими политиками управления доступом.

Этот процесс помогает защитить персональные данные, сохраняя возможность их анализа без риска идентификации конкретных физических лиц.

Пример обезличивания по методу изменения состава или семантики

1. **Исходные данные**:
   - Личные данные, которые подлежат обработке.

Имя	Фамилия	Адрес	Номер телефона
Асан	Асанов	Город X, ул. 1	123-456-7890
Айгүл	Айбекова	Город Y, ул. 2	987-654-3210

2. **Изменение значений атрибутов или удаление данных**:
   - Изменение или обобщение значений для обеспечения анонимности.

Имя	Фамилия	Адрес	Номер телефона
Аноним	**Аноним**	**Город X**	**Номер 1**
Аноним	**Аноним**	**Город Y**	**Номер 2**

3. **Структурированность и анонимность**:
   - Связь между атрибутами сохраняется, но данные становятся неоднозначными.
   - Обобщение может привести к следующим результатам:

Атрибут	Значение
Имя	Аноним
Фамилия	Аноним
Город	Город X/Y

4. **Потеря полноты**:
   - При удалении значений теряется детальность данных.
   - Пример:

Атрибут	Значение
Имя	Удалено
Фамилия	Удалено
Город	Удалено

5. **Семантическая целостность**:
   - Изменения должны сохранять семантику данных для использования в обработке.
   - Например, вместо полного адреса оставляем только город:

Атрибут	Значение
Город	Город X

6. **Заключительный результат**:
   - Обезличенные данные, сохраняющие структуру, но не позволяющие идентифицировать субъекты.

Обезличенные данные
Аноним, Аноним, Город X

Этот процесс демонстрирует, как обобщение и изменение значений атрибутов позволяют получить обезличенные данные, сохраняя при этом их структурированность и анонимность.

Пример обезличивания по методу декомпозиции данных

1. **Исходные данные**:
   - Личные данные, содержащие множество атрибутов.

Имя	Фамилия	Адрес	Номер телефона	Дата рождения
Асан	Асанов	Город X, ул. 1	123-456-7890	01.01.1990
Айгүл	Айбекова	Город Y, ул. 2	987-654-3210	15.05.1985

2. **Разделение атрибутов на подмножества**:
   - Атрибуты делятся на несколько подмножеств.

Подмножество 1:

Имя	Фамилия
Асан	Асанов
Айгүл	Айбекова

Подмножество 2:

Адрес	Номер телефона
Город X, ул. 1	123-456-7890
Город Y, ул. 2	987-654-3210

3. **Создание таблиц связей**:
   - Устанавливаются связи между подмножествами в отдельных таблицах.

Уникальный Идентификатор	Подмножество 1 (Имя, Фамилия)	Подмножество 2 (Адрес,Номер телефона)
ID_001	Асан, Асанов	Город X, ул. 1, 123-456-7890
ID_002	Айгүл, Айбекова	Город Y, ул. 2, 987-654-3210

4. **Раздельное хранение записей**:
   - Каждое подмножество хранится отдельно, но связи между ними сохраняются.

Хранилище 1 (Имя, Фамилия):

Уникальный Идентификатор	Имя	Фамилия
ID_001	Асан	Асанов
ID_002	Айгүл	Айбекова

Хранилище 2 (Адрес, Номер телефона):

Уникальный Идентификатор	Адрес	Номер телефона

| ID_001 | Город X, ул. 1 | 123-456-7890 |
| ID_002 | Город Y, ул. 2 | 987-654-3210 |

5. **Обеспечение анонимности**:
   - Анонимность достигается сложными связями между хранилищами и защитой данных.

[Хранилище 1] <--- Связь ---> [Хранилище 2]

6. **Сохранение семантической целостности**:
   - Связи между атрибутами остаются неизменными, что позволяет сохранить семантику данных.

Связь между атрибутами
ID_001 соответствует Асан Асанова
ID_001 соответствует Город X, ул. 1

Пример обезличивания по методу перемешивания данных

1. **Исходные данные**:
   - Начальные данные, содержащие идентифицирующую информацию.

Имя	Фамилия	Адрес	Номер телефона
Асан	Асанов	Город X, ул. 1	123-456-7890
Айгүл	Айбекова	Город Y, ул. 2	987-654-3210

2. **Перемешивание атрибутов**:
   - Атрибуты перемешиваются или изменяются таким образом, чтобы связи между данными стали неочевидными.

Пример перемешивания:

Имя	Фамилия	Адрес	Номер телефона
Айбекова	Асан	Город Y, ул. 2	987-654-3210
Асанов	Айгүл	Город X, ул. 1	123-456-7890

3. **Создание нового набора данных**:

Имя	Фамилия	Адрес	Номер телефона
Айбекова	Асан	Город Y, ул. 2	987-654-3210
Асанов	Айгүл	Город X, ул. 1	123-456-7890

4. **Сохранение структуры данных**:
   - Связи между атрибутами сохраняются, но они не указывают на реальные идентичности.

Связь между атрибутами
[Айбекова, Асан] - [Город Y, ул. 2]
[Асанов, Айгүл] - [Город X, ул. 1]

5. **Ограничение доступа к оригинальным данным**:
   - Доступ к исходным данным должен быть строго ограничен.

Пример обезличивания по методу использования генератора случайных чисел

1. **Исходные данные**:
   - Начальные данные, содержащие идентифицирующую информацию.

Имя	Фамилия	Адрес	Номер телефона
Асанов	Асанов	Город X, ул. 1	123-456-7890
Айгүл	Айгүлова	Город Y, ул. 2	987-654-3210

2. **Генерация случайных чисел**:
   - Генерируем случайные значения для замены идентифицирующих данных.

Генерация случайных значений
ID_001 → 5463
ID_002 → 9281

3. **Замена идентифицирующих данных**:
   - Заменяем оригинальные данные на сгенерированные случайные значения.

Случайные данные			
Имя	Фамилия	Адрес	Номер телефона
5463	9281	Город Z, ул. 3	321-654-9870
1234	5678	Город W, ул. 4	654-321-0987

4. **Создание таблицы соответствий**:
   - Создаем таблицу соответствий, если требуется восстановление идентифицирующих данных.

Уникальный ID	Имя	Фамилия	Адрес
5463	Асанов	Асанова	Город X, ул. 1
9281	Айгүл	Айбекова	Город Y, ул. 2

5. **Ограничение доступа к оригинальным данным**:
   - Доступ к исходным данным должен быть строго ограничен.

Пример обезличивания по методу агрегации данных

1. **Исходные данные**:
   - Начальные данные, содержащие индивидуальные записи.

Имя	Фамилия	Возраст	Город
Асан	Асанов	30	Город X
Айгүл	Айбекова	25	Город Y
Сыргак	Саидов	40	Город X

2. **Агрегация данных**:
   - Объединение данных по определённым категориям, например, по городу или возрасту.

Агрегированные данные по городу:

Город	Средний возраст	Количество людей
Город X	35	2
Город Y	25	1

3. **Сохранение анонимности**:
   - В агрегированных данных нет возможности идентифицировать отдельных лиц.

Агрегированные данные
Город: Город X, Людей: 2
Средний возраст: 35

4. **Представление результатов**:
   - Агрегированные данные могут быть представлены в отчетах или визуализациях.

[График]

Город X: 35 лет, 2 человека
Город Y: 25 лет, 1 человек

5. **Заключительный результат**:
   - Полученные данные представляют общую картину, не позволяя идентифицировать конкретных людей.

Итоговые агрегированные данные

| Город X: 2 человека, 35 лет |
| Город Y: 1 человек, 25 лет |

Пример обезличивания по методу добавления шума

1. **Исходные данные**:
   - Начальные данные, содержащие конкретные значения.

Имя	Возраст	Зарплата
Айбек	30	50000
Айгүл	25	60000
Сыргак	40	70000

2. **Определение уровня шума**:
   - Устанавливаем уровень шума, который будет добавляться к данным.

Уровень шума: ±5 лет для возраста, ±5000 для зарплаты

3. **Добавление шума**:
   - Применяем шум к данным.

Имя	Возраст	Зарплата
Асан	28 (25-35)	49000 (44500-53500)
Айгүл	27 (22-32)	61000 (56000-66000)
Сыргак	39 (34-44)	69500 (64500-74500)

4. **Результирующие данные**:
   - Получаем новые значения, которые сложно связать с конкретными личностями.

Имя	Возраст	Зарплата
Асан	28	49000
Айгүл	27	61000
Сыргак	39	69500

5. **Заключение о защитных свойствах**:
   - Данные теперь менее идентифицируемы, но сохраняют полезные характеристики.

Итоговые данные с шумом
Возраст: 28, Зарплата: 49000
Возраст: 27, Зарплата: 61000
Возраст: 39, Зарплата: 69500

Пример обезличивания по методу размывания (для изображений)

1. Исходное изображение:

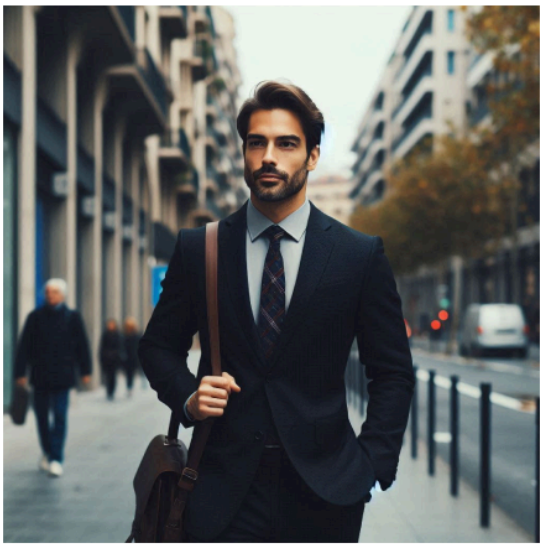
2. Ядро свертки (например, 3x3):
1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

3. Свертка:

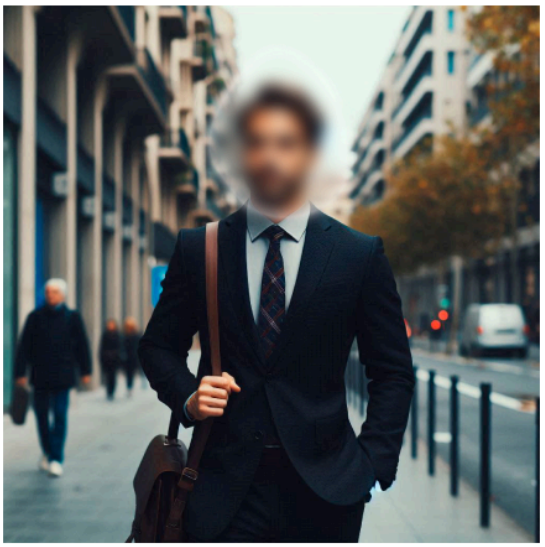
- Показать, как ядро перемещается по изображению, заменяя каждый центральный пиксель на среднее значение.

4. Размытое изображение:

1. Оригинальное изображение



2. Обезличенное изображение по методу размывания



Образец 9

Пример обезличивания по методу обобщения данных

1. **Сбор данных**:
   - **Описание:** данные собираются из различных источников, например, опросов, датчиков, баз данных.

2. **Предварительная обработка данных**:
   - Удаление шумов, обработка пропусков, нормализация и преобразование данных.

3. **Анализ данных**:
   - Применение статистических методов, алгоритмов для выявления паттернов.

4. **Обобщение данных**:
   - Агрегирование данных с помощью различных методов (средние значения, медианы и т.д.).

5. **Интерпретация результатов**:
   - Выявление значимых выводов и выводы на основе обобщенных данных.

6. **Принятие решений**:
   - Использование полученных данных для определения стратегий и принятия решений.