ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 71674— 2024

Системы искусственного интеллекта в клинической медицине

НАБОР ДАННЫХ В ФОРМАТЕ DICOM ДЛЯ ТЕСТИРОВАНИЯ АЛГОРИТМОВ

Методы обезличивания набора данных и контроля набора данных на отсутствие персональных данных

Издание официальное

Москва
Российский институт стандартизации
2024

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАН Государственным бюджетным учреждением здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы» (ГБУЗ «НПКЦ ДиТ ДЗМ»)

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 164 «Искусственный интеллект»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 7 октября 2024 г. № 1387-ст

4 ВВЕДЕН ВПЕРВЫЕ

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© Оформление. ФГБУ «Институт стандартизации», 2024

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 1
4 Общие положения 2
5 Методы обезличивания набора данных 3
6 Методы контроля набора данных на отсутствие персональных данных 8
Приложение А (обязательное) Список атрибутов DICOM-файла для обезличивания 10
Библиография 12

Введение

Вопросы конфиденциальности и безопасности данных связаны с вопросами неприкосновенности частной жизни, персональных данных, врачебной тайны, что достигается за счет обезличивания данных. Все меры информационной безопасности должны соответствовать действующим нормативным правовым актам в соответствии с [1].

К персональным данным относятся все сведения, которые прямо или косвенно могут идентифицировать лицо, что соответствует определению, приведенному в [2]. Обезличивание — действия, в результате которых становится невозможным без использования дополнительной информации определить принадлежность персональных данных конкретному субъекту персональных данных. Основная цель обезличивания — минимизация ущерба субъектам персональных данных в случае успешной реализации угрозы информационной безопасности.

При формировании наборов данных в формате DICOM для тестирования систем искусственного интеллекта в лучевой диагностике необходимо соблюдать требования по обезличиванию персональных данных в соответствии с [2]. В [3] и [4] была продемонстрирована важность этапа обезличивания и его контроль качества при подготовке наборов данных из медицинских изображений в формате DICOM, одним из этапов которых является обезличивание и его контроль качества.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Системы искусственного интеллекта в клинической медицине

НАБОР ДАННЫХ В ФОРМАТЕ DICOM ДЛЯ ТЕСТИРОВАНИЯ АЛГОРИТМОВ

Методы обезличивания набора данных и контроля набора данных на отсутствие персональных данных

Artificial intelligence in routine clinical practice. DICOM dataset for algorithm testing.
Dataset de-identification and monitoring for presence of personal data

Дата введения — 2025—01—01

1 Область применения

Настоящий стандарт определяет требования в области применения систем искусственного интеллекта (СИИ) в клинической медицине, а именно для наборов данных в формате Digital Imaging and Communications in Medicine (стандарт обработки, хранения, передачи, печати и визуализации медицинских изображений — DICOM) для тестирования алгоритмов.

Настоящий стандарт определяет методы обезличивания персональных данных пациентов, содержащихся в медицинских изображениях формата DICOM, а также методы контроля результата обезличивания набора данных.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р 59921.0 Системы искусственного интеллекта в клинической медицине. Основные положения

ГОСТ Р 59921.5 Системы искусственного интеллекта в клинической медицине. Часть 5. Требования к структуре и порядку применения набора данных для обучения и тестирования алгоритмов

ГОСТ Р ИСО 12052 Информатизация здоровья. Цифровые изображения и связь в медицине (DICOM), включая управление документооборотом и данными

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого стандарта с учетом всех внесенных в данную версию изменений. Если заменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

3 Термины и определения

В настоящем стандарте применены термины по ГОСТ Р 59921.0, а также следующие термины с соответствующими определениями:

Издание официальное

3.1

де-идентификация (de-identification): Общее название любого процесса удаления связи между совокупностью идентифицирующих данных и субъектом данных.

[ГОСТ Р 55036—2012, пункт 3.18]

3.2 **маскирование:** Процесс обработки изображения, посредством которого происходит удаление персональных данных за счет наложения маски на выбранные области исходного изображения.

3.3

метаданные (metadata): Информация о ресурсе.

[ГОСТ Р 57668—2017, пункт 4.10]

3.4 **набор данных** (data set): Совокупность данных, прошедших предварительную подготовку (обработку) в соответствии с требованиями законодательства Российской Федерации об информации, информационных технологиях и о защите информации и сформированных для разработки и тестирования программного обеспечения на основе искусственного интеллекта.

Примечание — См. [5].

3.5 **обезличивание персональных данных:** Действия, в результате которых становится невозможным без использования дополнительной информации определить принадлежность персональных данных конкретному субъекту персональных данных.

Примечание — См. [2].

3.6 **обратный процесс обезличивания:** Действия, в результате которых обезличенные данные принимают вид, позволяющий определить их принадлежность к конкретному субъекту персональных данных, то есть становятся персональными данными.

3.7 **оператор:** Государственный орган, муниципальный орган, юридическое или физическое лицо, самостоятельно или совместно с другими лицами организующие и (или) осуществляющее обработку персональных данных, а также определяющее цели обработки персональных данных, состав персональных данных, подлежащих обработке, действия (операции), совершаемые с персональными данными.

Примечание — См. [2].

4 Общие положения

Обезличивание набора данных, состоящего из медицинских изображений в формате DICOM, выполняют с целью минимизации ущерба субъектам персональных данных в случае успешной реализации угрозы информационной безопасности, например при передаче набора данных третьей стороне для тестирования СИИ.

Возможная схема реализации процесса обезличивания набора данных при необходимости их передачи третьей стороне приведена на рисунке 1.

В соответствии с [2], [6] и [7] к свойствам обезличенных данных относят следующие параметры:

- полноту (сохранение всей информации о конкретных субъектах или группах субъектов, которая имелась до обезличивания);
- структурированность (сохранение структурных связей между обезличенными данными конкретного субъекта или группы субъектов, соответствующих связям, имеющимся до обезличивания);
- релевантность (возможность обработки запросов по обработке персональных данных и получения ответов в одинаковой семантической форме);
- семантическую целостность (сохранение семантики персональных данных при их обезличивании);
- применимость (возможность решения задач обработки персональных данных, стоящих перед оператором, осуществляющим обезличивание персональных данных, обрабатываемых в информационных системах персональных данных, в том числе созданных и функционирующих в рамках реализации федеральных целевых программ (далее — оператор, операторы), без предварительного обезличивания всего объема записей о субъектах);
- анонимность (невозможность однозначной идентификации субъектов данных, полученных в результате обезличивания, без применения дополнительной информации).

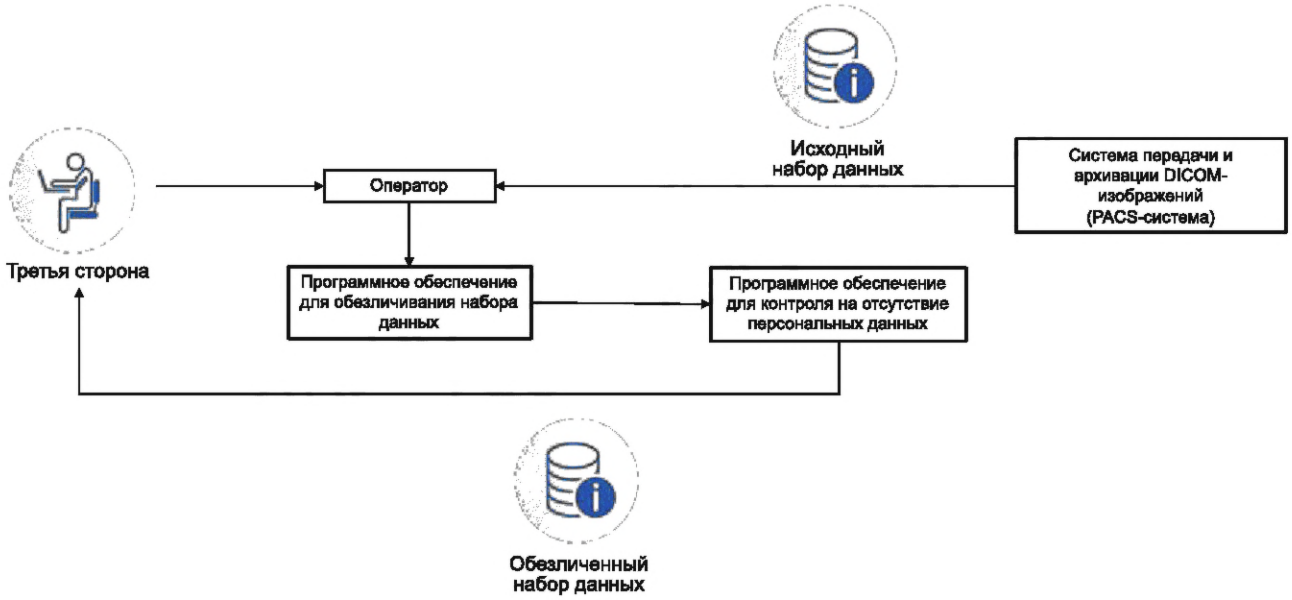


Рисунок 1 — Пример схемы процесса обезличивания набора данных при необходимости их передачи третьей стороне

DICOM-файл медицинского изображения представляет собой объектно-ориентированный файл, включающий информацию о пациенте, проведенном исследовании лучевой диагностики. Данная информация записана в значения соответствующих атрибутов по ГОСТ Р ИСО 12052.

Необходимо выполнить обнаружение всех DICOM атрибутов, содержащих персональные данные.

К персональным данным в соответствии с [8]—[10] относят атрибуты DICOM-файла, приведенные в таблице А.1.

5 Методы обезличивания набора данных

5.1 Общее описание методов

Стандартные методы обезличивания медицинских изображений в формате DICOM реализуют с помощью замены или уничтожения атрибутов, содержащих персональную информацию.

Необходимо выполнить обнаружение всех DICOM атрибутов, содержащих персональные данные, используя данные таблицы А.1, а также изучив данные других DICOM атрибутов.

Состав наборов атрибутов для обезличивания должен быть определен в зависимости от целей тестирования СИИ в соответствии с ГОСТ Р 59921.5.

Существует несколько методов обезличивания данных. Методы, оперирующие с атрибутами DICOM-файла приведены в 5.4.1—5.4.4. Метод обезличивания персональных данных, сохраненных на медицинском изображении, приведен в 5.4.5.

В частных случаях медицинские изображения могут содержать информацию, позволяющую реконструировать фотографию пациента. Это возможно выполнить, например, по данным медицинских изображений головы, полученных с помощью компьютерной томографии и магнитно-резонансной томографии. В таком случае необходимо обеспечить внесение изменений, не затрагивающих целевую область (например, головной мозг), но исключающих возможность восстановления фотографии пациента. Описание данных методов обезличивания не входит в область настоящего стандарта.

Дополнительно следует обращать внимание на медицинские изображения пациентов с инородными телами, позволяющие легко идентифицировать личность (например, отдельные виды пирсинга и т. д.), и исключать данные изображения из наборов данных.

5.2 Требования к квалификации персонала

Оператор, выполняющий обезличивание наборов данных, должен соответствовать требованиям [2], [6] и обладать:

- навыками по определению персональных данных;

- умением определять категории персональных данных;
- умением определять условия, при которых набор персональных данных позволяет точно идентифицировать человека;
- знаниями о принципах, условиях и порядке обработки персональных данных;
- знанием требований по обезличиванию персональных данных;
- умением эксплуатировать средства вычислительной техники;
- умением эксплуатировать программные средства по обезличиванию персональных данных.

5.3 Требования к программному обеспечению, осуществляющему обезличивание

Программное обеспечение, применяемое в процессе обезличивания, должно обеспечивать:
- конфиденциальность персональных данных;
- работу с атрибутами DICOM-файлов;
- осуществление одного из методов обезличивания по 5.4.1—5.4.5;
- возможность внесения изменений и поддержку актуальности таблиц соответствия, которые формируют в процессе обезличивания.

Примечание — Следует руководствоваться [6] и [7].

Оператор в зависимости от условий определяет уровень защищенности программного обеспечения, применяемого в процессе обезличивания персональных данных в соответствии с [11].

5.4 Описание метода обезличивания набора данных

5.4.1 Метод введения идентификаторов

Метод введения идентификаторов реализуют путем замены персональных данных, позволяющих идентифицировать субъект, их идентификаторами и созданием таблицы соответствия (справочника идентификаторов).

Применение данного метода позволяет получить обезличенные данные, обладающие следующими свойствами:
- полнотой (информация, позволяющая идентифицировать субъектов персональных данных, не удаляется, а переносится в таблицу соответствия);
- структурированностью (каждому идентификатору после процедуры обезличивания однозначно соответствует свой набор данных);
- семантической ценностью (вид представления данных не меняется, данные лишь переносятся в таблицу соответствия).

Обезличенные наборы данных, полученные в результате применения данного метода, не будут обладать свойством релевантности, поскольку в запросе и в ответе на запрос изменяется вид представления персональных данных, которые были заменены идентификаторами.

Применение данного метода сохраняет в записях связи между атрибутами обезличенных данных, соответствующие связям между атрибутами персональных данных.

Каждому значению идентификатора должно соответствовать одно значение атрибута и каждому значению атрибута должно соответствовать одно значение идентификатора.

Таблицы соответствия (дополнительные данные) создают для каждого атрибута персональных данных, значения которых заменяются идентификаторами.

На этапе реализации процедуры обезличивания определяют следующие параметры:
- перечень таблиц соответствия (перечень атрибутов, для которых происходит замена значений идентификаторами);
- правила вычисления идентификаторов — наборов символов, однозначно соответствующих значениям атрибутов персональных данных субъекта;
- объемы таблицы соответствия — количество строк таблицы соответствия, содержащих идентификатор и соответствующее ему значение.

В качестве атрибутов, значение которых заменяют идентификаторами, выбирают атрибуты, однозначно идентифицирующие субъекта персональных данных.

Количество идентификаторов и объем таблиц соответствия должны быть равны исходному количеству субъектов персональных данных. Возможны случаи, когда идентификатор вычисляют в зависимости от значения соответствующего атрибута. Однако должна быть обеспечена безопасность вычисления.

Пример — Создание таблицы соответствия значений атрибутов (персональных данных) и идентификаторов, где в качестве идентификаторов будут выступать хеш-функции значений атрибутов. Применение хеш-функции гарантирует, что не будет возможности преобразовать в персональные данные вычисленный идентификатор, хранящийся в открытом виде (см. [10]).

Таблицы соответствия должны храниться в соответствии с требованиями нормативных правовых актов.

5.4.2 Метод изменения состава или семантики

Метод изменения состава или семантики реализуют путем обобщения, изменения значений атрибутов персональных данных или удаления части сведений, позволяющих идентифицировать субъект.

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

- структурированностью — связь между отдельными значениями атрибутов персональных данных субъекта не нарушается;
- анонимностью — удаление или обобщение части данных приводит к неоднозначности при идентификации с использованием обезличенных данных;
- частичной релевантностью, поскольку в определенных случаях возможно получить семантическое соответствие поискового запроса и полученного ответа на запрос;
- применимостью, поскольку оператор может осуществлять обработку, не требующую идентификации всего объема данных о субъекте.

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.

Семантическую целостность обезличенных данных обеспечивают только при условии проведения изменений в составе персональных данных, сохраняющих семантику данных. Изменения должны учитывать специфику задач обработки, стоящих перед оператором.

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.

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

Процедура реализации данного метода должна содержать правила удаления либо замены значений персональных данных субъектов на новые значения, вычисляемые по заданным правилам.

При замене значений атрибутов на новые необходимо устанавливать правила обратной замены, если это необходимо для обезличивания.

На этапе реализации процедуры обезличивания необходимо определить следующие параметры:

- перечень атрибутов персональных данных, подлежащих удалению;
- перечень атрибутов персональных данных, подлежащих замене на новые значения;
- правила вычисления значений для замены (обратной замены) персональных данных субъектов.

Программная реализация процедуры должна обеспечить возможность внесения изменений и дополнений в состав обезличенных данных, динамическое вычисление значений для замены при занесении новых субъектов, проверку и поддержку актуальности данных.

5.4.3 Метод декомпозиции

Метод декомпозиции реализуют 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:

- полнотой — информация о субъектах персональных данных не удаляется, а переносится в другое хранилище;
- структурированностью — сохраняется связь между записями в разных хранилищах, что позволяет однозначно сопоставлять их;
- семантической целостностью — семантика и вид представления данных о субъекте не изменяется;
- релевантностью, поскольку возможно получить семантическое соответствие поискового запроса и полученного ответа на запрос;

- применимостью, поскольку оператор может осуществлять обработку данных, расположенных в одном хранилище, как независимо от другого, так и при совместном их использовании без идентификации всего объема обезличенных данных.

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

Процедура реализации метода производит разделение исходного массива персональных данных на несколько частей, каждая из которых содержит заданный набор атрибутов всех субъектов. Сведения, содержащиеся в каждой части, не позволяют идентифицировать субъекты персональных данных.

На этапе реализации процедуры обезличивания необходимо определить следующие параметры:

- перечень атрибутов, содержащихся в каждом подмножестве персональных данных;
- таблицы связей между подмножествами персональных данных;
- адреса хранения подмножеств персональных данных.

Правила разделения исходного массива данных определяют таким образом, чтобы каждая из раздельно хранимых частей не содержала сведений, позволяющих однозначно идентифицировать субъект персональных данных.

Программная реализация процедуры должна обеспечивать согласованное внесение изменений и дополнений во все подмножества и таблицы связей, поиск данных о субъекте во всех подмножествах, поддержку актуальности таблиц связей, проверку полноты данных (согласование подмножеств).

5.4.4 Метод перемешивания

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

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

- полнотой (вся информация о субъектах персональных данных сохраняется);
- структурированностью (связи между данными полностью восстанавливают при идентификации);
- семантической целостностью (семантика и вид представления данных о субъекте не изменяется);
- анонимностью (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);
- релевантностью (поскольку возможно получить семантическое соответствие поискового запроса и полученного ответа на запрос);
- применимостью [поскольку при наличии доступа к дополнительной (служебной) информации оператор может осуществлять обработку как отдельных записей о субъектах, так и всех данных без идентификации всего объема обезличенных данных].

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

На этапе реализации процедуры обезличивания необходимо определить следующие параметры:

- набор параметров алгоритма перемешивания (дополнительные данные для обезличивания);
- значения параметров алгоритма перемешивания (дополнительные данные для обезличивания).

Выбор параметров перемешивания зависит от алгоритма перемешивания, требуемой стойкости к атакам и объема обезличенных персональных данных.

Программная реализация процедуры должна обеспечивать возможность внесения изменений и дополнений в состав обезличенных данных, добавление новых пользователей, поддержку актуальности данных и возможность повторного перемешивания с новыми параметрами без предварительного обезличивания.

5.4.5 Метод обезличивания данных на медицинских изображениях

Часть персональных данных может быть закодирована непосредственно на участках изображений (например, в дозовых отчетах, вторичных объемных реконструкциях и др.), как показано на рисунке 2.

Удаление персональной информации, внедренной в изображение, основано на методе оптического распознавания текста (optical character recognition; OCR) (см. [12]).

Общая схема выполнения процедуры обезличивания медицинских изображений в формате DICOM в случае наличия персональных данных на участках изображения приведена на рисунке 3 (см. [13]). Для реализации данного метода обезличивания необходимо найти участки изображения, содержащие персональные данные с использованием метода OCR (см. [14]).

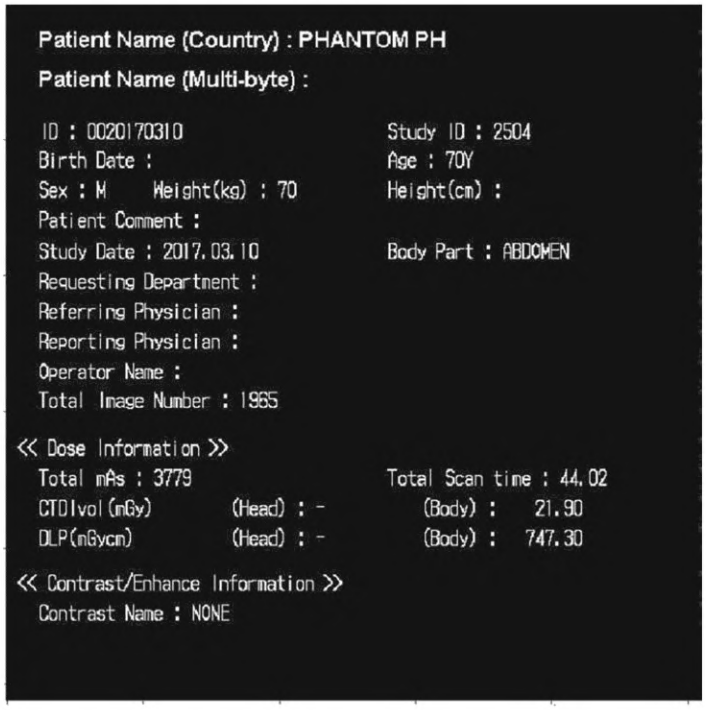


Рисунок 2 — Пример медицинского изображения с внедренной приватной информацией

Для обезличивания данных областей выполняют наложение бинарной маски (маскирование), которая, например, имеет значения 0 — для областей, содержащих персональных данные, и 1 — для области, выделенной методом OCR как содержащей персональные данные. Пример осуществления данного метода приведен на рисунке 4.



ПД — персональные данные

Рисунок 3 — Схема выполнения обезличивания данных на медицинских изображениях

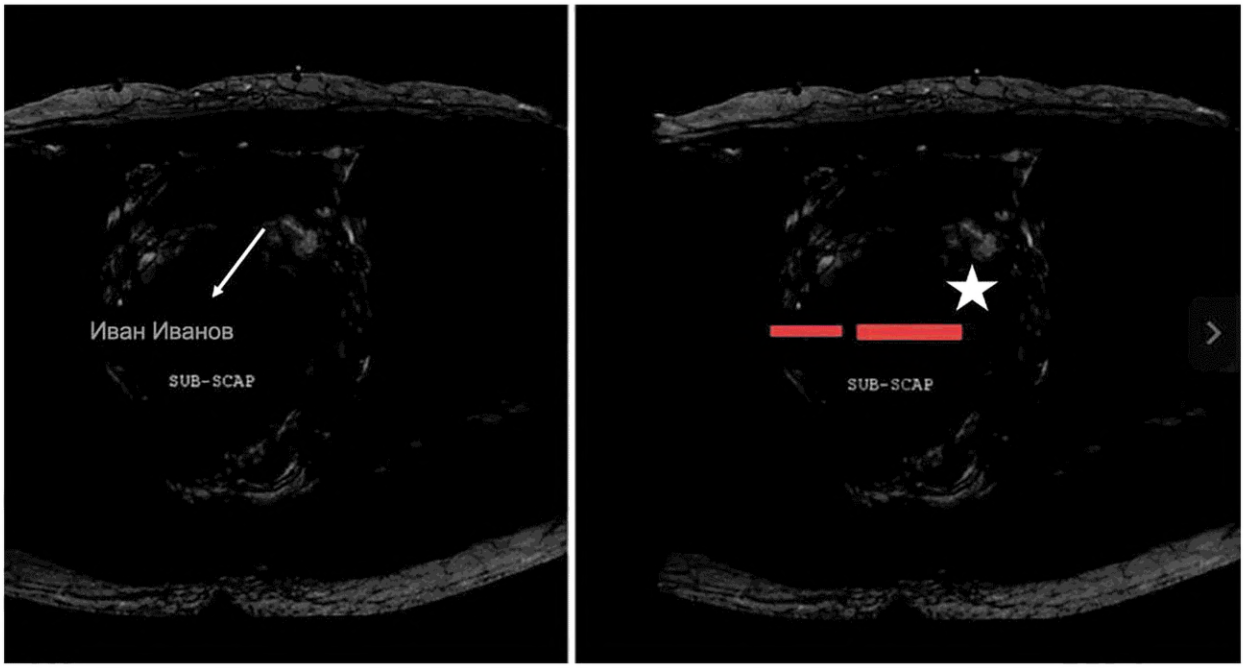


Рисунок 4 — Пример применения наложения маски на области с персональными данными: слева — исходное изображение, стрелкой обозначено положение имени пациента; справа — обезличенное изображение, звездой обозначено положение участков изображения с наложением маски (см. [13])

5.5 Предоставление результатов

Тип используемого варианта обезличивания должен быть записан в обезличенном файле в DICOM-атрибут «De-identification Method» с номером группы и элемента — [0012, 0063].

Метаданные набора данных оформляют в соответствии с ГОСТ Р 59921.5.

В метаданные обезличенного набора данных необходимо включить описание обезличивания (см. [15]), включающее следующую информацию:

- какие атрибуты были удалены во время обезличивания;
- какие атрибуты были заменены фиктивными значениями и как генерируют фиктивные значения;
- какие атрибуты были заменены на идентификаторы атрибутов для последующей повторной идентификации, а также любые соответствующие сведения о том, как выбираются ключи для выполнения шифрования;
- область, в которой приложение может гарантировать ссылочную целостность значений замены для ссылок;
- какие атрибуты и значения атрибутов вставляют во время защиты DICOM-атрибута класса пары сервис-объект («Service-Object Pair»);
- какие синтаксисы передачи поддерживаются для обезличивания набора данных зашифрованных атрибутов.

Сформированные во время процесса обезличивания наборов данных таблицы должны быть сохранены в защищенном месте, доступ к которому имеет только персонал, уполномоченный оператором на основании нормативных правовых актов.

Обезличенные наборы данных необходимо хранить отдельно от наборов данных, позволяющих произвести процесс, обратный обезличиванию.

6 Методы контроля набора данных на отсутствие персональных данных

6.1 Назначение

Методы контроля набора данных на отсутствие персональных данных применяют с целью подтверждения обезличивания персональных данных, указанных в нормативных правовых актах, а также установленных в требованиях оператора.

6.2 Требования к квалификации персонала

Требования должны соответствовать 5.2.

В случае возникновения конфликта между персоналом, выполняющим обезличивание, и персоналом, проводящим контроль, необходимо привлечение третьей стороны (или еще одного сотрудника, выполняющего контроль). Решение будет принято по результатам консенсуса или голосования данных лиц.

6.3 Требования к программному обеспечению

Программное обеспечение, применяемое для контроля набора данных на отсутствие персональных данных, должно обеспечивать:

- конфиденциальность обезличенных персональных данных;
- просмотр атрибутов DICOM-файлов;
- просмотр изображений DICOM-файлов.

Оператор в зависимости от условий определяет уровень защищенности программного обеспечения, применяемого для контроля набора данных на отсутствие персональных данных в соответствии с [11].

6.4 Описание метода контроля набора данных на отсутствие персональных данных

Метод контроля набора данных может быть основан на визуальном контроле атрибутов DICOM-файлов, а также может быть произведен при помощи средств автоматизации с использованием программного обеспечения.

6.5 Предоставление результатов

Результаты контроля набора данных на отсутствие персональных данных должны быть оформлены в виде протокола с указанием соответствия/несоответствия требованиям к обезличенным данным.

Приложение А
(обязательное)

Список атрибутов DICOM-файла для обезличивания

Расширенный список атрибутов DICOM-файла для обезличивания приведен в таблице А.1.

Таблица А.1 — Расширенный список атрибутов DICOM-файла для обезличивания

№	Номер группы и элемента	Имя атрибута
1	0008,0020	StudyDate
2	0008,0021	SeriesDate
3	0008,0022	AcquisitionDate
4	0008,0023	ContentDate
5	0008,0024	OverlayDate
6	0008,0025	CurveDate
7	0008,002A	AcquisitionDatetime
8	0008,0030	StudyTime
9	0008,0031	SeriesTime
10	0008,0032	AcquisitionTime
11	0008,0033	ContentTime
12	0008,0034	OverlayTime
13	0008,0035	CurveTime
14	0008,0050	AccessionNumber
15	0008,0080	InstitutionName
16	0008,0081	InstitutionAddress
17	0008,0090	ReferringPhysiciansName
18	0008,0092	ReferringPhysiciansAddress
19	0008,0094	ReferringPhysiciansTelephoneNumber
20	0008,0096	ReferringPhysicianIDSequence
21	0008,1040	InstitutionalDepartmentName
22	0008,1048	PhysicianOfRecord
23	0008,1049	PhysicianOfRecordIDSequence
24	0008,1050	PerformingPhysiciansName
25	0008,1052	PerformingPhysicianIDSequence
26	0008,1060	NameOfPhysicianReadingStudy
27	0008,1062	PhysicianReadingStudyIDSequence
28	0008,1070	OperatorsName
29	0010,0010	PatientsName
30	0010,0020	PatientID
31	0010,0021	IssuerOfPatientID
32	0010,0030	PatientsBirthDate
33	0010,0032	PatientsBirthTime
34	0010,0040	PatientsSex
35	0010,1000	OtherPatientIDs
36	0010,1001	OtherPatientNames
37	0010,1005	PatientsBirthName

Окончание таблицы А.1

№	Номер группы и элемента	Имя атрибута
38	0010,1010	PatientsAge
39	0010,1040	PatientsAddress
40	0010,1060	PatientsMothersBirthName
41	0010,2150	CountryOfResidence
42	0010,2152	RegionOfResidence
43	0010,2154	PatientsTelephoneNumbers
44	0020,0010	StudyID
45	0038,0300	CurrentPatientLocation
46	0038,0400	PatientsInstitutionResidence
47	0040,A120	DateTime
48	0040,A121	Date
49	0040,A122	Time
50	0040,A123	PersonName
51	0010,1002	Other Patient IDs Sequence
52	0010,0022	Type of Patient ID
53	0010,1090	Medical Record Locator
54	0010,1100	Referenced Patient Photo Sequence

Библиография

[1] Федеральный закон от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации»

[2] Федеральный закон от 27 июля 2006 г. № 152-ФЗ «О персональных данных»

[3] МР 80 Регламент подготовки наборов данных с описанием подходов к формированию репрезентативной выборки данных. Часть 1. Методические рекомендации [препринт] / сост. С.П. Морозов, А.В. Владзимирский, А.Е. Андрейченко [и др.] // Серия «Лучшие практики лучевой и инструментальной диагностики». — Вып. 103. — М. : ГБУЗ «НПКЦ ДиТ ДЗМ», 2021. — 40 с.

[4] Павлов Н.А., Андрейченко А.Е., Владзимирский А.В., и др. Эталонные медицинские датасеты (MosMedData) для независимой внешней оценки алгоритмов на основе искусственного интеллекта в диагностике // Digital Diagnostics. 2021. Т. 2, No1. с.49—65

[5] Указ Президента Российской Федерации от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации»

[6] Требования и методы по обезличиванию персональных данных, обрабатываемых в информационных системах персональных данных, в том числе созданных и функционирующих в рамках реализации федеральных целевых программ (утверждены приказом Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций от 5 сентября 2013 г. № 996)

[7] Методические рекомендации по применению приказа Роскомнадзора от 5 сентября 2013 г. № 996 «Об утверждении требований и методов по обезличиванию персональных данных» (утверждены руководителем Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций 13 декабря 2013 г.)

[8] DICOM : [сайт]. — United States, 2021. — URL: gost.e-ecolog.ru sect_C.2.2.html (дата обращения: 03.07.2021 г.)

[9] Aryanto K. Y. E., Oudkerk M., van Ooijen P. M. A. Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy //European radiology. — 2015. — Vol. 25. — № 12. — p. 3685—3695

[10] Ноздрина, А.А. Метод обезличивания персональных данных, основанный на введении идентификаторов и хешировании / А.А. Ноздрина, Д.В. Применко // Молодежь и новые информационные технологии: Всероссийская научно-практическая конференция молодых ученых в рамках Программы развития деятельности студенческих объединений Череповецкого государственного университета «РАЙОН IT», Череповец, 17—18 ноября 2016 года. — Череповец: Череповецкий государственный университет, 2016. — с. 64—67

[11] Постановление Правительства Российской Федерации от 01.11.2012 № 1119 «Об утверждении требований к защите персональных данных при их обработке в информационных системах персональных данных»

[12] Ray Smith. An overview of the tesseract OCR engine. In Proc. in Int. Conference on Document Analysis and Recognition (ICDAR), pages 629—633, 2007

[13] Kin G., Tsui W., Chan T. Automatic selective removal of embedded patient information from image content of DICOM files. Am. J. Roentgenol. 2012; 198 (4): 769—772

[14] Новик В.П., Кульберг Н.С., Арзамасов К.М., Четвериков С.Ф., Хоружая А.Н., Козлов Д.В., Кремнева Е.И. Распознавание областей текста с персональными данными на диагностических изображениях. Медицинская визуализация. 2023;27(4):150—158

[15] DICOM : [сайт]. — United States, 2021. — URL: gost.e-ecolog.ru html#table_E.1-1. — Текст. Изображение: электронные

УДК 615.841:006.354 ОКС 11.040.01

Ключевые слова: система искусственного интеллекта, искусственный интеллект, лучевая диагностика, обезличивание, наборы данных, DICOM

Редактор *М.В. Митрофанова*
Технический редактор *И.Е. Черепкова*
Корректор *И.А. Королева*
Компьютерная верстка *Е.О. Асташина*

Сдано в набор 08.10.2024. Подписано в печать 28.10.2024. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 2,32. Уч.-изд. л. 1,80.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении в ФГБУ «Институт стандартизации»
для комплектования Федерального информационного фонда стандартов,
117418 Москва, Нахимовский пр-т, д. 31, к. 2.
gost.e-ecolog.ru gost.e-ecolog.ru