ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 57773— 2017 (ИСО 19157: 2013)

ПРОСТРАНСТВЕННЫЕ ДАННЫЕ

Качество данных

(ISO 19157:2013, Geographic information — Data quality, MOD)

Издание официальное

Москва
Стандартинформ
2017

gost.e-ecolog.ru

Предисловие

1 ПОДГОТОВЛЕН Федеральным государственным бюджетным учреждением «Федеральный научно-технический центр геодезии, картографии и инфраструктуры пространственных данных» (ФГБУ «Центр геодезии, картографии и ИПД») на основе собственного перевода на русский язык англоязычной версии стандарта, указанного в пункте 4

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 394 «Географическая информация/геоматика»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 6 октября 2017 г. № 1367-ст

4 Настоящий стандарт является модифицированным по отношению к международному стандарту ИСО 19157:2013 «Географическая информация — Качество данных» (ISO 19157:2013 «Geographic information — Data quality, MOD») путем включения дополнительных фраз, слов, ссылок и внесения изменений по отношению к тексту применяемого международного стандарта, которые выделены курсивом, подчеркиванием сплошной горизонтальной линией, а также невключения отдельных структурных элементов, ссылок и дополнительных элементов. Объяснения причин внесения этих технических отклонений приведены во введении.

Сопоставление структуры настоящего стандарта со структурой указанного международного стандарта приведено в дополнительном приложении ДА.

Наименование настоящего стандарта изменено относительно наименования указанного международного стандарта для приведения в соответствие с ГОСТ Р 1.5—2012 (пункт 3.5).

Сведения о соответствии ссылочных национальных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте, приведены в дополнительном приложении ДБ

5 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 ежемесячном информационном указателе «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

© Стандартинформ, 2017

В Российской Федерации 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 Соответствие . . . 1
3 Нормативные ссылки . . . 1
4 Термины и определения . . . 2
5 Сокращенные термины . . . 3
6 Обзор качества данных . . . 3
7 Компоненты качества данных . . . 5
7.1 Обзор компонентов . . . 5
7.2 Единица качества данных (data quality unit) . . . 5
7.3 Элементы качества данных (data quality elements) . . . 6
7.4 Дескрипторы элементов качества (descriptors of data quality elements) . . . 8
7.5 Элементы метакачества (metaquality elements) . . . 11
7.6 Дескрипторы элемента метакачества (descriptors of a metaquality element) . . . 12
8 Меры качества данных (data quality measures) . . . 13
8.1 Общие положения . . . 13
8.2 Стандартизированные меры качества данных . . . 13
8.3 Создание определяемых пользователем мер качества данных . . . 13
8.4 Каталог мер качества данных . . . 13
8.5 Список компонентов . . . 13
8.6 Подробное описание компонентов . . . 14
9 Оценка качества данных (data quality evaluation) . . . 15
9.1 Процесс оценки качества данных . . . 15
9.2 Методы оценки качества данных (data quality evaluation methods) . . . 17
9.3 Обобщение и преобразование (aggregation and derivation) . . . 18
10 Отчетность о качестве данных (data quality reporting) . . . 18
10.1 Общие положения . . . 18
10.2 Частные случаи . . . 19
Приложение А (обязательное) Комплекс проверок . . . 20
Приложение В (справочное) Концепции оценки качества данных и их применение . . . 21
Приложение С (обязательное) Словарь данных для оценки качества . . . 25
Приложение D (обязательное) Перечень стандартизированных мер качества данных . . . 38
Приложение Е (справочное) Оценка и выдача информации о качестве данных . . . 86
Приложение F (справочное) Методы выборочной оценки . . . 108
Приложение G (обязательное) Базовые меры качества данных . . . 115
Приложение Н (справочное) Управление мерами качества данных . . . 119
Приложение I (справочное) Руководство по использованию элементов качества . . . 122
Приложение J (справочное) Обобщение результатов оценки качества . . . 129
Приложение ДА (справочное) Сопоставление структуры настоящего стандарта со структурой примененного в нем международного стандарта . . . 131
Приложение ДБ (справочное) Сведения о соответствии ссылочных национальных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте . . . 132
Библиография . . . 133

Введение

Пространственные данные все чаще распространяют, обменивают и используют для целей, отличных от тех, которые предусматривались разработчиком. Информация о качестве доступных данных имеет важнейшее значение для процесса отбора наборов данных, в которых ценность данных напрямую связана с их качеством. Пользователям пространственных данных представляется множество наборов данных на выбор. В связи с этим возникает необходимость сравнения качества наборов данных с целью определения тех, которые наилучшим образом отвечают потребностям пользователя.

Цель описания качества пространственных данных состоит в облегчении сравнения и выбора набора данных, которые наилучшим образом удовлетворяют потребностям или требованиям приложений. Полные описания качества наборов данных будут способствовать распространению, обмену и использованию соответствующих наборов данных. Информация о качестве пространственных данных позволяет их разработчику оценить, насколько точно данный набор отвечает критериям, изложенным в спецификации продукта, и помогает пользователям данных оценить способность продукта удовлетворять требованиям его конкретного приложения. Для осуществления такой оценки используется ряд четко определенных процедур в последовательном порядке.

С целью облегчения сравнения важно, чтобы результаты в отчетах информации о качестве данных были выражены в сопоставимом виде, а также чтобы было общее понимание использованных мер качества данных. Эти меры качества обеспечивают дескрипторы качества пространственных данных путем сравнения с данными предметной области. Использование несовместимых мер делает сравнение качества данных невозможным. Настоящий стандарт устанавливает требования к компонентам и структуре мер качества данных и определяет общие базовые меры качества данных.

В настоящем стандарте отмечается, что мнение производителя и мнение пользователя о качестве данных могут не совпадать. Уровень соответствия качеству может быть установлен с помощью спецификации на продукт производителя данных или требований к качеству пользователя данных. Если пользователю данных требуется больше информации о качестве данных, чем предусмотрено их производителем, то пользователь данных может ознакомиться с ходом процесса оценки качества данных у производителя, чтобы получить дополнительную информацию. В этом случае требования пользователя данных рассматриваются как спецификация продукта с целью использования в процессе производства данных.

Цель настоящего стандарта — установить принципы описания качества пространственных данных и концепций обработки информации о качестве пространственных данных, а также последовательные и стандартные методы определения и выдачи информации о качестве наборов данных. Задачей стандарта также является обеспечение руководящими принципами процедур оценки количественной информации о качестве пространственных данных.

Настоящий стандарт является модифицированным по отношению к международному стандарту ИСО 19157:2013 «Географическая информация — Качество данных» (ISO 19157:2013 «Geographic information — Data quality»). В настоящий стандарт внесены изменения, необходимые для приведения его в соответствие современным международным и национальным стандартам, в частности изменен раздел нормативных ссылок, уточнена библиография, в тексте стандарта приведены ссылки на национальные стандарты.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ПРОСТРАНСТВЕННЫЕ ДАННЫЕ

Качество данных

Spatial data. Data quality

Дата введения — 2018—06—01

1 Область применения

Настоящий стандарт устанавливает принципы описания качества пространственных данных, в том числе:

- определяет компоненты для описания качества данных;
- устанавливает компоненты и структуру содержания реестра для мер качества данных;
- описывает общие процедуры оценки качества пространственных данных,
- устанавливает принципы составления отчетности о качестве данных.

Настоящий стандарт также определяет меры качества данных для использования при оценке и составлении отчетности об их качестве. Он предназначен для разработчиков данных, обеспечивающих информацию о качестве, для описания и оценки того, насколько набор данных соответствует спецификации продукта, а также для пользователей данных, пытающихся определить, обладают ли конкретные пространственные данные достаточным качеством, необходимым для их частного применения.

Настоящий стандарт не преследует цель определения минимально приемлемых уровней качества пространственных данных.

2 Соответствие

Любой продукт, заявляющий о соответствии настоящему стандарту, должен отвечать всем требованиям, описанным в комплексе проверок, который приведен в приложении А, а именно:

- процесс оценки качества данных должен пройти проверку, изложенную в разделе А.1 приложения А;
- метаданные о качестве данных должны пройти проверку, изложенную в разделах А.2 и А.3 приложения А;
- независимый отчет о качестве должен пройти проверку, изложенную в разделе А.4 приложения А;
- меры качества данных должны пройти проверку, изложенную в разделе А.5 приложения А.

3 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р 50779.72—99 Статистические методы. Процедуры выборочного контроля по альтернативному признаку. Часть 2. Планы выборочного контроля отдельных партий на основе предельного качества LQ

ГОСТ Р 57668—2017 Пространственные данные. Метаданные. Часть 1. Основные положения

ГОСТ Р 57656—2017 Пространственные данные. Метаданные. Часть 2. Расширения для изображений и матричных данных

ГОСТ Р 57657—2017 Пространственные данные. Спецификация информационного продукта

ГОСТ Р ИСО 2859-3—2009 Статистические методы. Процедуры выборочного контроля по альтернативному признаку. Часть 3. Контроль с пропуском партий

Издание официальное

ГОСТ Р ИСО 3951-1—2015 Статистические методы. Процедуры выборочного контроля по количественному признаку. Часть 1. Требования к одноступенчатым планам на основе AQL при контроле последовательных партий по единственной характеристике и единственной AQL

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 ежемесячно издаваемого информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный документ, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого документа с учетом всех внесенных в данную версию изменений. Если заменен ссылочный документ, на который дана датированная ссылка, то рекомендуется использовать версию этого документа с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный документ, 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.

4 Термины и определения

4.1 **точность** (accuracy): Степень соответствия результата проверки или измерений принятому истинному значению.

Примечание — В настоящем стандарте истинным значением может выступать эталонное значение, принятое за истинное.

4.2 **каталог** (catalogue): Набор элементов (4.18) либо электронный или бумажный документ, который содержит информацию о наборе элементов.

4.3 **соответствие** (conformance): Выполнение заданных требований.

4.4 **уровень соответствия качества** (conformance quality level): Пороговое значение или набор пороговых значений для результирующих оценок качества данных (4.21), по которым определяют степень соответствия набора данных (4.8) спецификации информационного продукта (4.6) или потребностям пользователя.

4.5 **корректность** (correctness): Соответствие предметной области (4.24).

4.6 **спецификация информационного продукта** (data product specification): Детальное описание набора данных (4.8) или ряда наборов данных, а также дополнительная информация, которая обеспечивает его (их) создание, поставку и использование другой стороной.

4.7 **базовая мера качества данных** (data quality basic measure): Универсальная мера качества данных (4.21), используемая как основа для специализированных мер качества данных.

Примечание — Базовые меры качества данных — абстрактные типы данных. Непосредственное их использование при составлении отчета о качестве данных не представляется возможным.

4.8 **набор данных** (dataset): Идентифицируемая совокупность данных.

Примечание — Набором данных может быть также малая группа данных, которая, хотя и имеет некоторые ограничения в виде пространственной протяженности или типа объекта (4.15), физически расположена в пределах более крупного набора данных. Теоретически набором данных могут быть отдельный объект (4.11) или атрибут объекта (4.12), расположенные в пределах большого набора данных. Печатная копия топографической или морской карты может рассматриваться как набор данных.

4.9 **комплект наборов данных** (dataset series). Совокупность наборов данных (4.8), обладающих одинаковыми характеристиками.

4.10 **метод прямой оценки** (direct evaluation method): Метод оценки качества (4.21) набора данных (4.8), основанный на проверке элементов (4.18) набора данных.

4.11 **пространственный объект, объект** (feature): Абстракция явления реального мира.

Примечание — Объект может быть представлен как тип или экземпляр. Тип объекта (4.15) или экземпляр объекта (4.13) используются в зависимости от назначения.

4.12 **атрибут пространственного объекта** (feature attribute): Характеристика объекта (4.11).

Примечание — Атрибут объекта имеет имя, характеризуется определенным типом данных и имеет область допустимых значений. Атрибут экземпляра объекта (4.13) также имеет значение атрибута, принадлежащее области допустимых значений.

4.13 **экземпляр объекта** (feature instance): Объект данного типа (4.15), имеющий определенные значения атрибутов объекта (4.12).

4.14 **операция с объектом** (feature operation): Операция, которую может выполнить любой экземпляр некоторого типа пространственного объекта.

4.15 **тип объекта** (feature type): Класс объектов (4.11), имеющих общие характеристики.

4.16 ***пространственные данные*** *(spatial data): Данные о пространственных объектах, включающие сведения об их форме, местоположении и свойствах, в том числе представленные с использованием координат.*

4.17 **метод косвенной оценки** (indirect evaluation method): Метод оценки качества (4.21) набора данных (4.8), основанный на дополнительных знаниях.

Примечание — Примером дополнительных знаний служит происхождение набора данных, таких как метод производства или исходные данные.

4.18 **элемент** (item): То, что может быть описано и рассмотрено отдельно.

Примечание — Элемент может являться любой частью набора данных (4.8), такой как объект (4.11), отношения объектов, атрибут объекта (4.12) или их комбинация.

4.19 **метаданные** (metadata): Данные о данных.

4.20 **метакачество** (metaquality): Информация, описывающая качество (4.21) оценки качества данных.

4.21 **качество** (quality): Степень соответствия совокупности собственных характеристик заявленным требованиям.

4.22 **реестр** (register): Набор файлов, содержащий идентификаторы элементов (4.18) с описанием соответствующих элементов.

4.23 **самостоятельный отчет о качестве** (standalone quality report): Свободный текстовый документ, содержащий подробную информацию о примененных методах оценки качества данных (4.21), результатах и мерах.

4.24 **предметная область** (universe of discourse): Отображение реального или гипотетического мира, которое включает все, что представляет интерес.

5 Сокращенные термины

5.1 Аббревиатуры

ADQR	(aggregated data quality results)	—	агрегированные результаты качества данных;
AQL	(acceptance quality limit)	—	допустимый уровень качества;
RMSE	(root mean square error)	—	среднеквадратическая ошибка;
UML	(Unified Modeling Language)	—	унифицированный язык моделирования;
XML	(Extensible Markup Language)	—	расширяемый язык разметки.

5.2 Пакетные аббревиатуры

Для обозначения пакета, содержащего класс, предусмотрено использование аббревиатур. Те сокращения, которые предшествуют именам класса, соединяются символом "_". Ниже приводится перечень таких аббревиатур:

CI	(Citation)	—	ссылки;
CT	(Catalogues)	—	каталоги;
DQ	(Data Quality)	—	качество данных;
DQM	(Data Quality Measure)	—	мера качества данных;
EX	(Extent)	—	пространственно-временная протяженность;
GF	(General Feature)	—	основной объект;
MD	(Metadata)	—	метаданные,
QE	(Quality Extanded)	—	расширенное качество данных;
RE	(Registration)	—	регистрация.

6 Обзор качества данных

Работа с качеством данных включает:

- изучение концепций качества в отношении пространственных данных. В приложении В представлено описание концепций качества данных, применяемых при определении компонентов для описания качества пространственных данных;

- определение уровней соответствия качества данных спецификации информационного продукта или потребностям пользователя. Разработка спецификации информационного продукта описана в *ГОСТ Р 57657;*
- определение аспектов качества в схемах приложения;
- оценку качества данных;
- составление отчетности о качестве данных.

Примечание — Создание схем приложения описано в [1].

Оценка качества данных может быть применена к комплекту наборов данных, набору данных или поднабору данных в наборе данных, обладающих общими характеристиками, так что их качество может быть оценено.

Качество данных должно быть описано с помощью элементов качества данных. Элементы качества данных и их дескрипторы применяются для описания степени соответствия набора данных критериям, изложенным в спецификации информационного продукта или требованиях пользователя, и предоставляют количественную информацию о качестве.

Если информация о качестве описывает данные, которые были созданы без детальной спецификации информационного продукта или с наличием спецификации, но при отсутствии в ней количественных мер и дескрипторов, то оценка элемента данных может быть осуществлена посредством неколичественного субъективного подхода в виде описательного результата для каждого элемента.

Некоторая информация о качестве данных может быть получена на основе элементов назначения, использования и происхождения. Такая информация выдается в виде метаданных согласно *ГОСТ Р 57668.*

Примечание — Назначение описывает обоснование создания набора данных и содержит информацию о его предполагаемом использовании, которое может не совпадать с фактическим использованием набора данных. Использование описывает приложения, в которых использовался набор данных как разработчиком данных, так и другими пользователями данных. Происхождение описывает историю набора данных и производит подсчет жизненного цикла набора данных, начиная со сбора и приобретения, включая составление и преобразование к их текущему виду. Эта общая, неколичественная информация является наглядной для пользователей и может помочь в оценке качества набора данных, особенно в случаях особого применения, отличного от предполагаемого (см. также 9.2.3).

Настоящий стандарт определяет, что количественные элементы качества данных могут иметь ассоциативное качество, которое называется метакачеством. Метакачество описывает качество результатов оценки качества данных по определенным характеристикам.

Примечание — Концепция метакачества описана в 7.5.

На рисунке 1 представлен обзор информации по качеству данных.

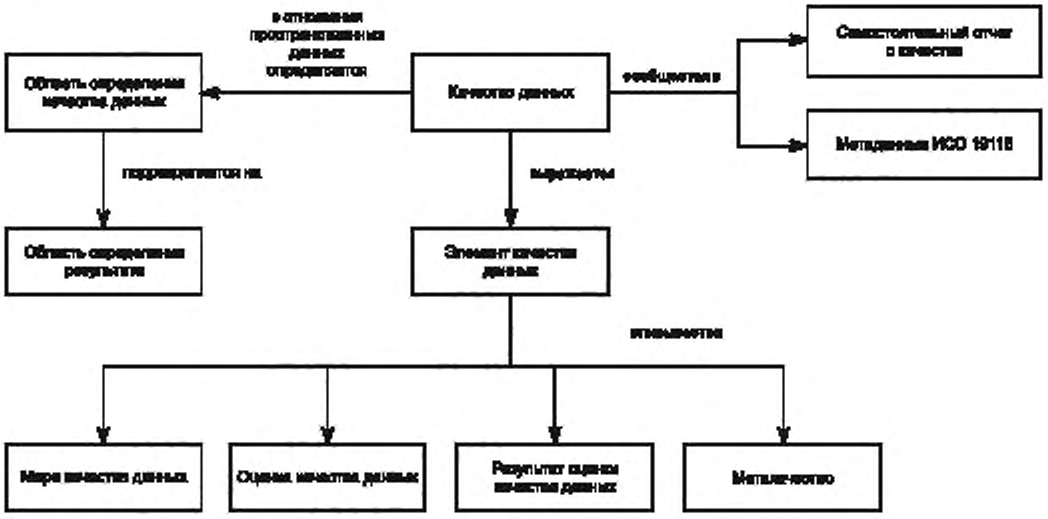


Рисунок 1 — Концептуальная модель качества пространственных данных

7 Компоненты качества данных

7.1 Обзор компонентов

Компоненты качества данных описаны в разделе 7. На рисунке 2 представлен обзор компонентов и связей между ними. Подробнее о компонентах и их атрибутах см. в словаре данных в приложении С.

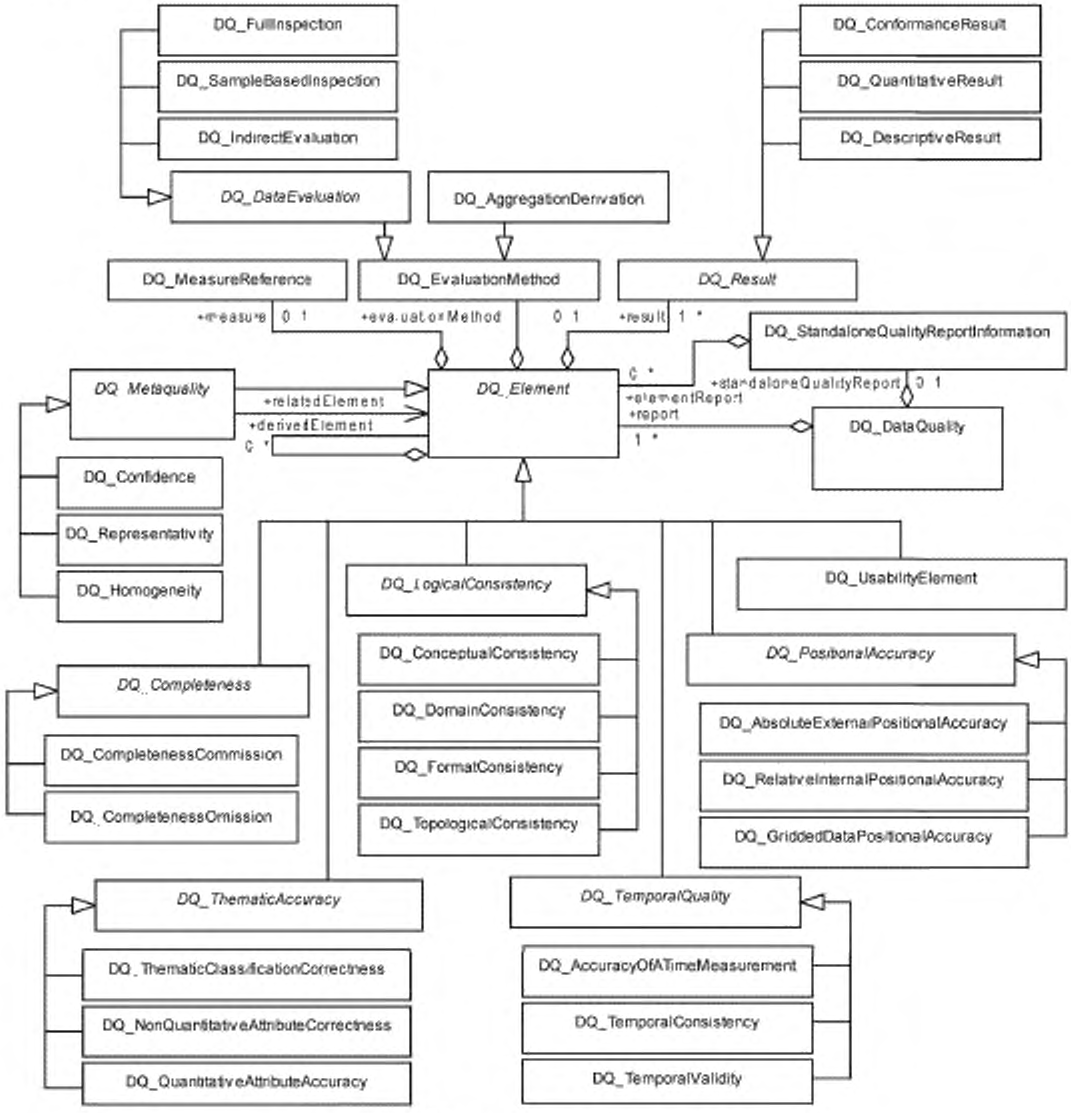


Рисунок 2 — Обзор компонентов качества данных

7.2 Единица качества данных (data quality unit)

Для описания качества пространственных данных могут рассматриваться различные элементы качества и различные поднаборы данных, с целью описания которых используются единицы качества данных. Единица качества данных — это совокупность области определения и элементов качества данных (см. рисунок 3).

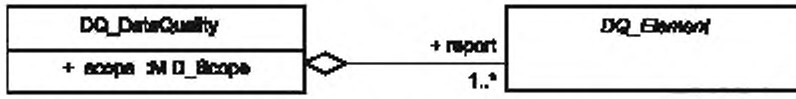


Рисунок 3 — Единица качества данных

Область определения единиц(ы) качества данных задает протяженность, пространственные, и/или временные, и/или общие характеристики, идентифицирующие данные, качество которых должно быть оценено.

Одна область определения качества данных должна быть предусмотрена для каждой единицы качества данных. Один отчет о качестве данных (метаданные или самостоятельный отчет по оценке качества) может включать несколько единиц качества данных, так как часто области для отдельных элементов качества данных различны. Эти различные области могут быть, например, пространственно разделены, перекрываться или даже совпадать.

Ниже приведены примеры того, что определяет область определения качества данных (см. также MD_Scope *в ГОСТ Р 57656*):

а) комплект наборов данных;

б) набор данных;

в) поднабор данных, определяемый одной или несколькими из следующих характеристик:

1) типы элементов (наборы типов объектов, атрибутов объектов, операций с объектами или отношений объектов);

2) конкретные элементы (наборы экземпляров объектов, значения атрибутов или экземпляры отношений объектов);

3) географическая протяженность;

4) временная протяженность (заданные временные рамки и точность временных рамок).

7.3 Элементы качества данных (data quality elements)

7.3.1 Общие положения

Элемент качества данных — это компонент, описывающий определенный аспект качества пространственных данных, организованный по различным категориям. Данные категории представлены на рисунке 4.

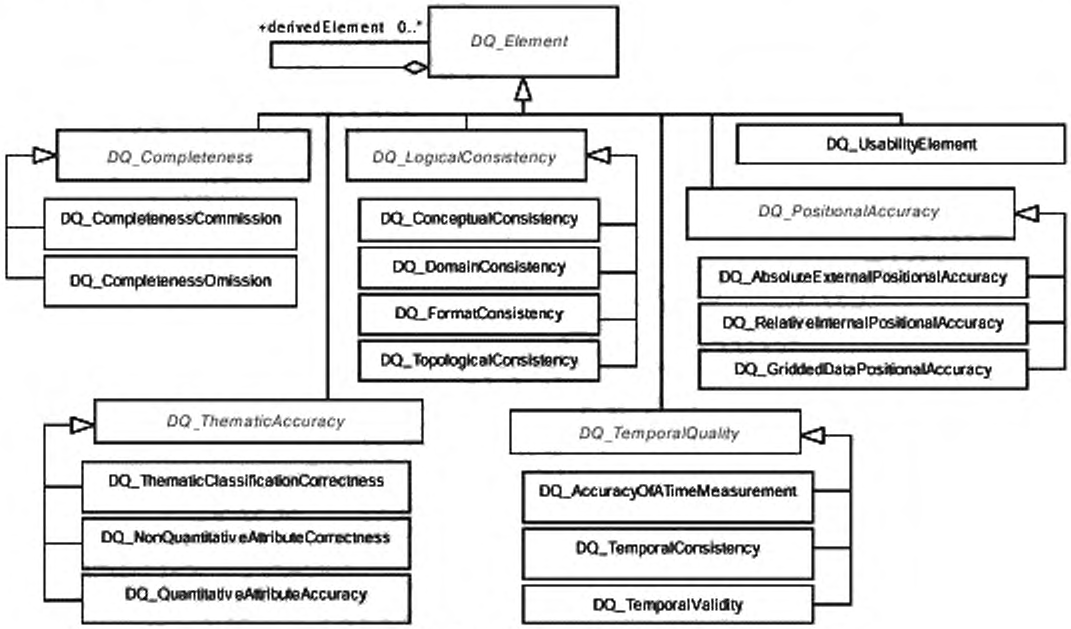


Рисунок 4 — Обзор элементов качества данных

7.3.2 Полнота (completeness)

Полнота определяется наличием и отсутствием объектов, их атрибутов и отношений. Она состоит из двух элементов качества данных:

- присутствие (commission): избыточность данных в наборе данных;
- отсутствие (omission): отсутствие данных в наборе данных.

7.3.3 Логическая согласованность (logical consistency)

Под логической согласованностью понимают степень соответствия логических правил структуры данных, атрибутов и отношений (структура данных может быть концептуальной, логической или физической). Если эти логические правила документально оформлены в другом источнике (например, в спецификации на информационный продукт), то необходимо ссылаться на этот источник (например, при оценке качества данных). Логическая согласованность состоит из четырех элементов качества данных:

- концептуальная согласованность (conceptual consistency): соответствие правилам концептуальной схемы;
- доменная согласованность (domain consistency): соответствие значений атрибутов области допустимых значений;
- согласованность по формату (format consistency): степень, с которой данные хранятся в соответствии с физической структурой набора данных;
- топологическая согласованность (topological consistency): корректность представления закодированных топологических характеристик набора данных.

7.3.4 Позиционная точность (positional accuracy)

Под позиционной точностью понимают точность положения объектов внутри пространственной системы координат. Она состоит из трех элементов качества:

- абсолютная или внешняя точность (absolute or external accuracy): степень соответствия заявленных значений координат значениям координат, принятым в качестве правильных или являющимся правильными;
- относительная или внутренняя точность (relative or internal accuracy): степень соответствия относительного положения объектов в наборе данных их соответствующим исходным положениям, принятым в качестве правильных или являющимся правильными;
- позиционная точность матричных данных (gridded data positional accuracy): соответствие значений пространственного позиционирования матричных данных значениям, принятым в качестве правильных или являющимся правильными.

7.3.5 Тематическая точность (thematic accuracy)

Под тематической точностью понимают точность количественных атрибутов и корректность неколичественных атрибутов, классификаций объектов и их отношений. Она состоит из трех элементов качества.

- правильность классификации (classification correctness): соответствие классов объектов или их атрибутов предметной области (например, реальной ситуации или эталонному набору данных);
- правильность неколичественных атрибутов (non-quantitative attribute correctness): определение, является ли неколичественный атрибут правильным или неправильным;
- точность количественных атрибутов (quantitative attribute accuracy): степень соответствия значения количественного атрибута значению, принятому в качестве правильного или являющемуся правильным.

7.3.6 Временное качество (temporal quality)

Под временным качеством понимают качество временных атрибутов и временных отношений объектов. Оно состоит из трех элементов качества:

- точность измерения времени (accuracy of a time measurement): степень соответствия заявленных временных измерений значениям, принятым в качестве правильных или являющимся правильными;
- согласованность по времени (temporal consistency): правильность временного порядка событий;
- временная достоверность (temporal validity): достоверность данных по отношению ко времени.

Примечание — В качестве измерения времени может выступать определенный момент времени или период.

Пример — 33 марта — пример неверных данных.

7.3.7 Элемент применимости (usability element)

Применимость основана на требованиях пользователя. Все элементы качества могут быть использованы для оценки применимости. Оценка применимости может основываться на конкретных

требованиях пользователей, которые не могут быть описаны с использованием элементов качества, описанных выше. В этом случае должен использоваться элемент применимости с целью предоставления конкретной информации о качестве в отношении пригодности определенного набора данных для конкретного приложения или удовлетворения набору требований.

При использовании элемента применимости рекомендуется использовать все приемлемые дескрипторы элементов качества (см. 7.4) и определять меры качества в соответствии с разделом 8 или приложением D в целях предоставления детальной информации по оценке.

Пример — С помощью данного элемента производитель данных может продемонстрировать, насколько набор данных пригоден для различных указанных способов использования. Данный элемент может быть применен для подтверждения соответствия набора данных конкретной спецификации.

7.4 Дескрипторы элементов качества (descriptors of data quality elements)

7.4.1 Общие положения

Оценка элемента качества данных осуществляется с помощью:

- меры (measure): тип оценки;
- метода оценки (evaluation method): процедура, используемая для оценки меры;
- результата (result): итог оценки.

Данные дескрипторы представлены на рисунке 5 и описаны в 7.4.2, 7.4.3 и 7.4.4.

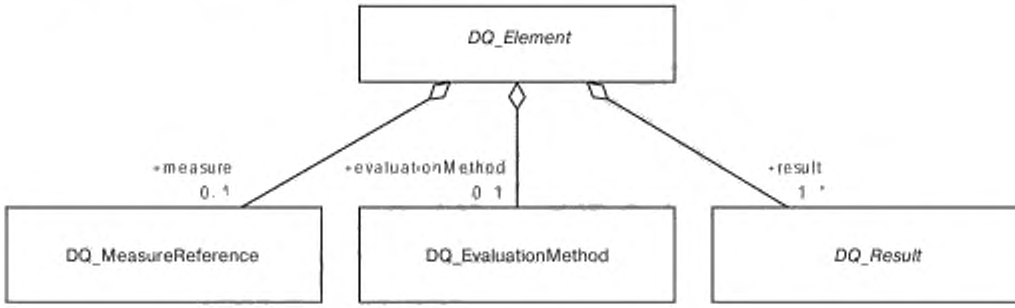


Рисунок 5 — Дескрипторы элементов качества

7.4.2 Мера (measure)

Элемент качества данных должен ссылаться только на одну меру посредством ссылки (см. рисунок 6), указывая идентификатор меры, полностью описанный в другом месте (DQM_Measure. measureIdentifier, см. 8.6.1), и/или указывая имя и краткое описание меры.

Примечание — Полное описание можно найти в реестре или каталоге мер, которые могут образовывать часть спецификации информационного продукта или самостоятельного отчета по качеству.

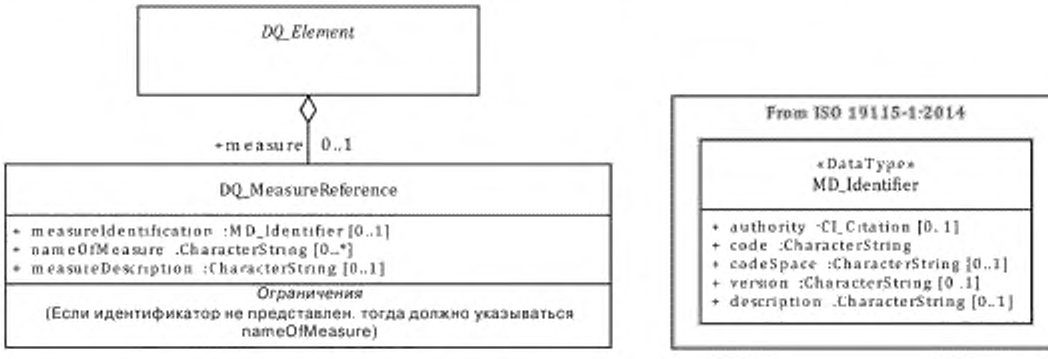


Рисунок 6 — Ссылки на меры качества данных

Меры качества данных более подробно описаны в разделе 8 настоящего стандарта. В приложении D приведен список стандартизированных мер качества данных.

Пример — Процент значений атрибутов, которые являются верными.

Настоящий стандарт устанавливает, что для измерения качества набора данных используются различные методы. Одной меры качества данных может быть недостаточно для полной оценки качества данных, задаваемых областью определения качества данных, а также для предоставления мер качества для всевозможных вариантов использования набора данных. Комбинация мер качества данных может давать полезную информацию. Несколько мер качества данных могут быть описаны для данных, заданных одной областью определения качества данных. В этом случае отчет о качестве данных должен включать один экземпляр DQ_Element для каждой примененной меры.

7.4.3 Метод оценки (evaluation method)

Метод оценки качества данных описывает те процедуры и методы, которые применяются к пространственным данным для получения результата оценки качества данных (см. рисунок 7). Для различных элементов качества данных часто используются различные методы оценки.

Для каждой примененной меры качества должен быть включен метод оценки качества. Метод оценки качества данных используется для описания методологии, используемой для применения меры качества данных для данных, задаваемых некоторой областью их определения, или для описания документации, в которой такая методология описывается.

Примечание — Оценка качества данных подробнее описана в разделе 9.

Пример — Примерами такой документации являются спецификации на информационный продукт, опубликованные статьи или утвержденные отраслевые стандарты.

Для каждой оценки следует указывать дату или диапазон дат. Если оценка проводилась в непоследовательном временном порядке, то следует указывать каждую отдельную дату. Даты изложены в соответствии с требованиями [2].

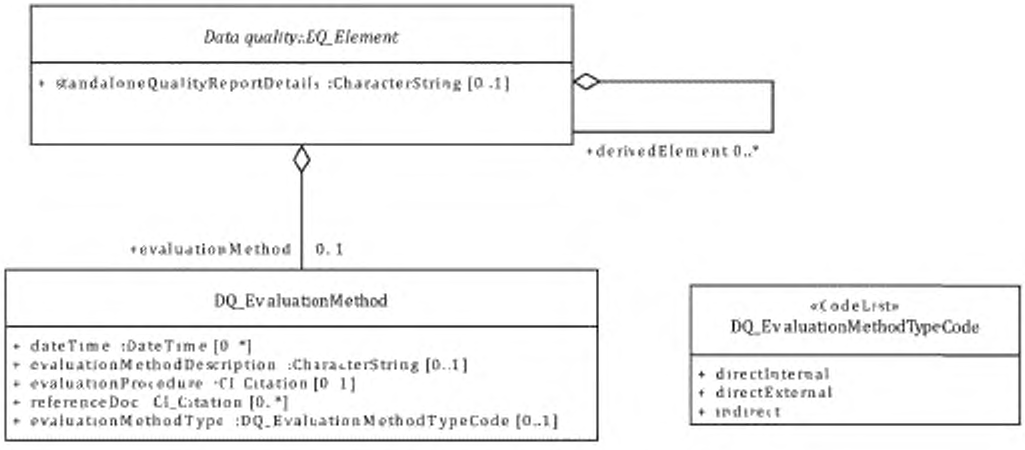


Рисунок 7 — Метод оценки качества данных

7.4.4 Результат (result)

7.4.4.1 Общие положения

Для каждого элемента качества данных должно быть обеспечено не менее одной результирующей оценки качества. Это может быть количественный результат, результат соответствия, описательный результат или результат покрытия (см. также рисунок 8).

Примечание — Для одного элемента качества данных могут применяться различные типы результатов.

Для различных частей набора данных качество часто различается. Поэтому несколько оценок могут быть применены для одного элемента качества данных для более полного и более детального описания количественной информации о качестве. Чтобы избежать повторных описаний мер и проце-

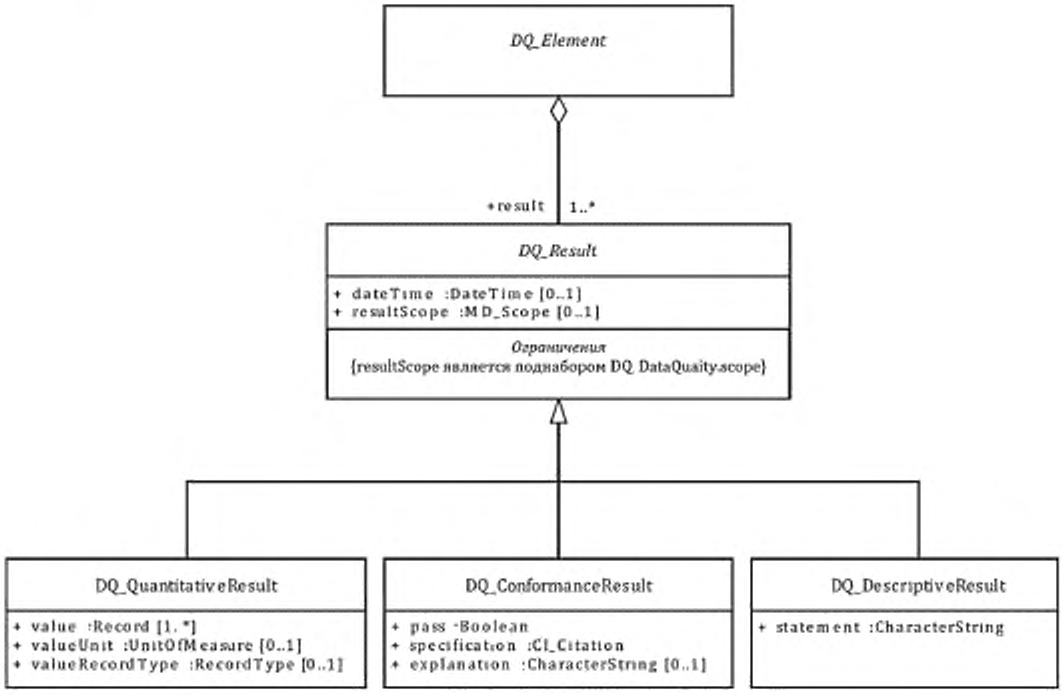


Рисунок 8 — Результат оценки качества

дур оценки для нескольких экземпляров элемента качества данных (DQ_Element), можно использовать несколько результатов с областью применения индивидуальных результатов.

Примечание — Результирующая область — это подгруппа области определения качества данных (см. 7.2).

Пример — Набор данных содержит объекты идентичного типа, положение которых было установлено с помощью отличающихся методов и с различной точностью определения местоположения. Однако для всего набора данных применяются одни и те же процедура и мера оценки качества, которые дают различные результаты в зависимости от метода сбора данных. В этом случае может быть желательно иметь несколько результатов с отдельными результирующими областями (зона, охватываемая каждым методом сбора данных) и одной областью определения качества данных (набор данных).

7.4.4.2 Количественный результат (quantitative result)

Количественным результатом может быть одно или несколько значений, в зависимости от значений атрибутов valueType и valueStructure, определенных в описании применяемой меры.

Атрибут valueRecordType используется для описания того, как valueType и valueStructure, определенные данной мерой, реализованы для получения значения количественного результата.

Примечание — Атрибут valueRecordType принадлежит к типу RecordType, который является универсальным типом данных согласно [3]. Его значение меняется в зависимости от того, какое решение по реализации используется для получения количественного результата. Пример реализации XML для RecordType дан в [4].

Пример 1 — Использование XML: простой пример: value = 5, valueRecordType = gco:Integer, valueUnit = «metre.»

Пример 2 — В рамках описания меры, valueType — целочисленная переменная, а valueStructure: матрица (n×n). Значение атрибута количественного результата обеспечивает результат самой матрицы в пределах числового кодирования, использующего определенный тип XML, называемый MatrixType (например). В XML атрибут valueRecordType обеспечивает описание типа MatrixType. Если применяется другая кодировка, то атрибут valueRecordType изменяется для обеспечения описания

типа Matrix в другой кодировке, и, соответственно, изменяется реализация вычисления значения атрибута, но само значение остается прежним.

Для каждого результата должна быть установлена отдельная единица значения в соответствующих случаях.

Пример 3 — Возможные единицы расстояния: метр, сантиметр, миллиметр.

Пример 4 — Мера «уровень избыточных элементов» (см. таблицу D.3) используется для оценки количества избыточных элементов в наборе данных по отношению к количеству элементов, которые должны быть представлены. Результат количественного значения имеет тип значения Real. В этом случае единица значения указывается в процентах, значение умножается на 100. В этом примере единицей значения является «процент», «%».

7.4.4.3 Результат соответствия (conformance result)

Результат соответствия — это результат сравнения значения или набора значений, полученных в результате применения некоторой меры качества к данным, заданным областью определения качества, с учетом установленного приемлемого уровня соответствия качества.

Когда уровень соответствия качества определен, его сравнивают с полученным результатом с целью выявления того факта, что качество данных удовлетворяет заданному уровню качества.

Результат соответствия может быть установлен для каждой меры. Уровень соответствия качества определяется в соответствующей документации, такой как спецификация к информационному продукту или спецификация требований, определенных пользователем. При определении соответствия необходимо ссылаться на соответствующую нормативно-техническую документацию и указывать примененный уровень соответствия качества.

Для одной меры предусмотрено более одного результата соответствия качества данных, если процедура оценки противоречит уровню соответствия, установленного в различных источниках.

7.4.4.4 Описательный результат (descriptive result)

В некоторых случаях (например, при тематических и геонаучных исследованиях) проведение количественной результирующей оценки для элемента качества данных не представляется возможным. Тогда субъективная оценка элемента может быть выражена в текстовом формате в качестве описательного результата качества данных.

Пример — Относительная точность положения выше между геологическим объектом и соседним объектом на топографической карте (дороги, реки, озера и т. д.), чем абсолютная точность положения самого геологического объекта.

Данный описательный результат также может быть использован для проведения короткого синтетического описания результата оценки качества данных, сопровождения полного количественного результата или полной его замены в случае, если количественные значения не могут быть предоставлены.

7.4.4.5 Результат покрытия (coverage result)

Результат покрытия — это результат оценки качества данных, формализованных в виде покрытия. Он описан в *ГОСТ Р 57656*.

7.5 Элементы метакачества (metaquality elements)

Элементы метакачества представляют собой набор количественных и качественных отчетных данных об оценке качества и его результате. Знания о качестве и пригодности метода оценки, примененных мерах и данном результате могут иметь такое же значение, как и сам результат.

Пример оценки метакачества см. в Е.5.3 приложения Е.

Метакачество может быть описано с использованием следующих элементов, представленных на рисунке 9:

- достоверность (confidence): надежность результата качества данных.

Примечание — Количественные показатели достоверности могут быть получены посредством статистических параметров, таких как стандартное отклонение или доверительный интервал с заданным уровнем достоверности.

Пример — Достоверность зависит в первую очередь от используемого метода и его надежности, а также в меньшей мере от соответствующей генеральной совокупности;

- репрезентативность (representativity): степень репрезентативности данных в пределах области определения качества данных в соответствии с результатом выборки.

Примечание — Статистический метод, основанный на выборке, может считаться таким же надежным, как и глобальный метод, когда охвачены все географические зоны и рассматриваемые временные диапазоны, а генеральная совокупность достаточно велика. Важен не только объем выборки, который имеет решающее значение, но и тот факт, насколько точно отражается текущее состояние данных. См. также 9.2.2 и приложение F;

- гомогенность (homogeneity): ожидаемая или тестируемая однородность результатов, полученных для оценки качества данных.

Примечание — Однородность заключается в сравнении результатов оценки нескольких сегментов глобального набора данных. Это сравнение может быть выражено, например, посредством среднеквадратических ошибок. Гомогенность невозможно оценить в случае общего процесса, так как результат носит глобальный характер.

Примечание — Такие тесты часто проводятся при сборе данных различными операторами в зависимости от зоны или времени сбора данных.

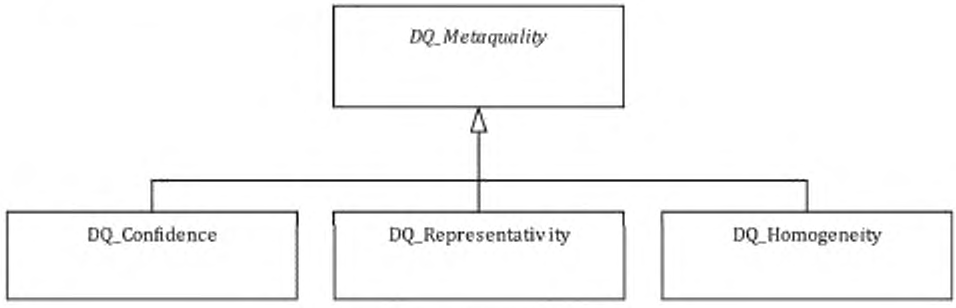


Рисунок 9 — Элементы метакачества

7.6 Дескрипторы элемента метакачества (descriptors of a metaquality element)

Для описания элемента метакачества используются те же дескрипторы, что и для элемента качества [мера, метод оценки и результат (см. 7.4 и рисунок 10)]. Дополнительно должен применяться следующий дескриптор:

- связанный элемент качества.

Примечание — Связанный элемент качества — это элемент, к которому применяется элемент метакачества.

Пример оценки метакачества см. в Е.5.3 приложения Е.

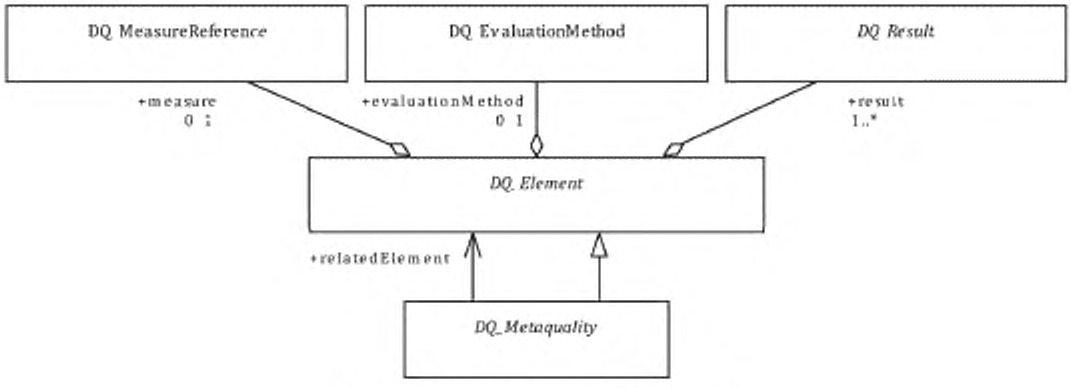


Рисунок 10 — Дескрипторы метакачества

8 Меры качества данных (data quality measures)

8.1 Общие положения

Для облегчения сравнения набора данных необходимо, чтобы результаты в отчетах по качеству данных представлялись в сравнительном виде, а также чтобы имелось общее представление об использованных мерах качества данных. Чтобы сделать оценки качества данных и отчеты о качестве данных (в виде метаданных или самостоятельного отчета по оценке качества) из различных источников сопоставимыми, должны использоваться по возможности стандартизированные меры качества данных, описанные в приложении D.

8.2 Стандартизированные меры качества данных

Перечень стандартизированных мер качества данных приведен в приложении D. Каждая мера качества данных этого списка содержит все необходимые компоненты, указанные в разделе 8. Для каждого элемента качества данных предусмотрено применение нескольких мер. Какие из них использовать, зависит от типа данных и их предполагаемого назначения. Меры из этого перечня должны быть использованы в рамках реализации данного стандарта.

Любой реестр, установленный для управления стандартизированными мерами качества данных, приведен в [5].

8.3 Создание определяемых пользователем мер качества данных

Вследствие природы качества и пространственных данных перечень стандартизированных мер качества данных не может быть полным. Возможны случаи, когда пользователь настоящего стандарта вынужден разрабатывать другие меры качества данных. По возможности эти меры должны быть определены с помощью базовых мер качества данных, описанных в приложении G, а сама мера — в соответствии со структурой, приведенной в разделе 8.

8.4 Каталог мер качества данных

Для более полного описания мер, указанных в отчете по оценке качества данных, каталоги мер качества данных могут быть предоставлены совместно с метаданными или в онлайн-режиме.

Каталог может содержать набор мер, применяемых в одном или нескольких отчетах со всеми необходимыми компонентами для мер качества данных, указанных в настоящем стандарте.

Каталог (как и реестр) позволяет пользователю описывать меру и хранить информацию для того, чтобы иметь возможность ссылаться на него при каждой необходимости, вместо повторного описания меры в отчете по качеству данных.

Приложение H описывает структуру каталога мер. В [4] представлен XML-механизм для ассоциирования каталога с набором метаданных.

8.5 Список компонентов

Каждая мера качества данных описывается следующими компонентами:

- идентификатор меры (8.6.1);
- название (8.6.2);
- краткое название (8.6.3);
- название элемента (8.6.4);
- базовая мера (8.6.5);
- определение (8.6.6);
- описание (8.6.7);
- параметр (8.6.8);
- тип значения (8.6.9);
- структура значения (8.6.10);
- ссылка на источник (8.6.11);
- пример (8.6.12).

На рисунке 11 представлены компоненты меры качества данных.

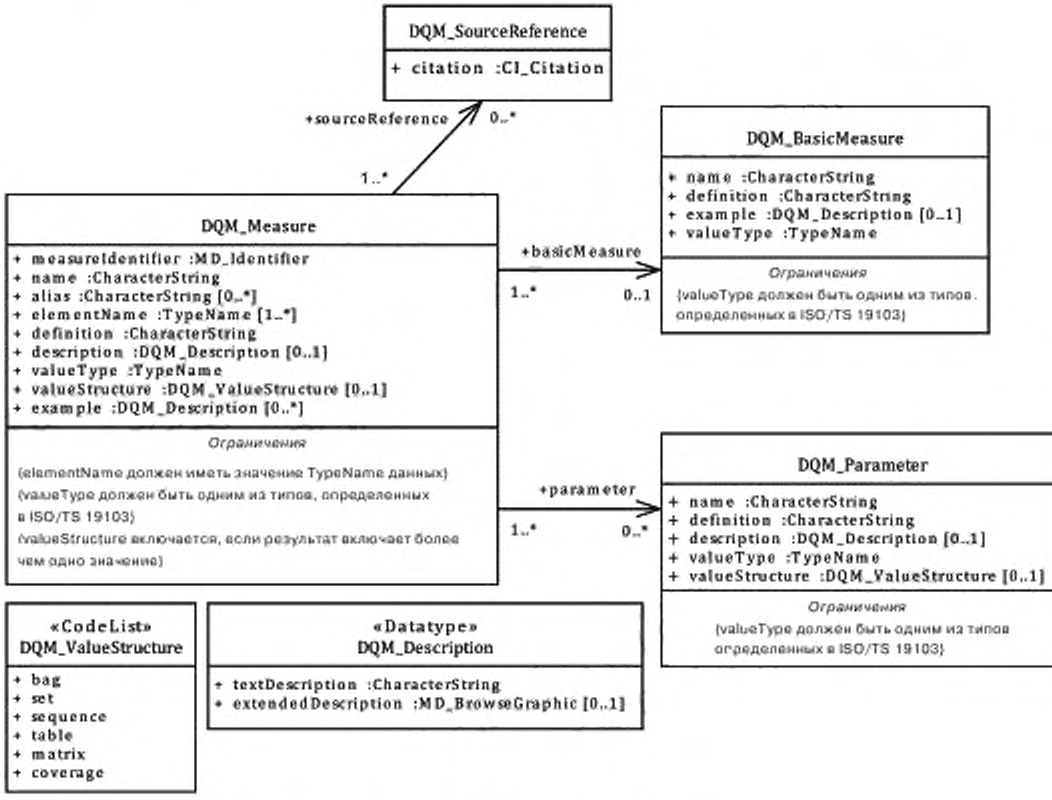


Рисунок 11 — Меры качества данных

8.6 Подробное описание компонентов

8.6.1 Идентификатор меры (measure identifier)

Идентификатор — это значение, уникально идентифицирующее меру в пространстве имен.

Примечание — Данный идентификатор устанавливает ссылки на меру качества данных в элементах качества данных (см. 7.4.2).

8.6.2 Название (name)

Название — это название меры.

Примечание — Если мера уже имеет общепринятое название, то должно использоваться это название. Если же такого названия не существует, то следует выбрать такое название, которое отражает характер этой меры.

8.6.3 Условное название (alias)

Условное название — это другое отличительное наименование для той же меры качества данных. Им может быть иное общепринятое наименование либо аббревиатура или же краткое имя. Возможно использование более одного условного названия.

8.6.4 Название элемента (element name)

Название элемента — это название элемента качества данных (см. 7.3 и 7.5), к которым применяется некая мера. Возможно использование более одного имени элемента.

8.6.5 Базовая мера (basic measure)

Если какая-либо мера основана на одной из базовых мер, то она должна быть описана его именем, определением и типом значения. Базовые меры идентифицируются по своим именам.

Множество мер основано на подсчете ошибочных элементов. Существуют также некоторые меры, связанные с неопределенностью числовых значений. Для того чтобы избежать повторений, наиболее

распространенные методы построения расчетных мер, а также общие статистические показатели для одно- и двумерных случайных величин должны быть установлены в виде базовых мер.

При необходимости создания новых мер необходимо также использовать базовые меры. Например, для выдачи информации по незамкнутой мозаичной поверхности или других мер в зависимости от приложения.

Примечание — В приложении G приведены базовые меры.

8.6.6 Определение (definition)

Определение является фундаментальным понятием меры.

Примечание — Если мера является производной от базовой меры, то такое определение основывается на определении базовой меры и специфицируется для этой меры.

8.6.7 Описание (description)

Описание — это описание меры, включающее методы подсчета с учетом всех формул и/или иллюстрации, необходимые для получения результата применения этой меры.

Если мера использует концепцию ошибок, то должно быть указано, каким образом элемент классифицируется как некорректный. В этом случае качество может быть описано только как корректное или некорректное.

8.6.8 Параметр (parameter)

Параметр — это вспомогательная переменная, используемая мерой. Она должна включать в себя имя, определение и тип значения. Может использоваться более одного параметра.

Примечание — Пример параметра см. в таблице D.66 приложения D.

8.6.9 Тип значения (value type)

Тип значения — это тип данных, используемый для представления результата меры. Должны использоваться типы данных, определенные в [3].

8.6.10 Структура значения (value structure)

Результат может состоять из множества значений. В таких случаях результат должен быть структурирован с использованием структуры значения, как указано в С.3.3 приложения С.

8.6.11 Ссылка на источник (source reference)

Ссылка на источник — это ссылка на документацию с описанием меры.

Когда мера, для которой предоставляется дополнительная информация из внешнего источника, добавляется в перечень стандартизированных мер, то здесь может быть указана ссылка на этот источник.

8.6.12 Пример (example)

Пример — это пример применения меры или результата, полученного при использовании этой меры. Может быть приведено более одного примера.

9 Оценка качества данных (data quality evaluation)

9.1 Процесс оценки качества данных

9.1.1 Введение

Процессы оценки качества используются на различных фазах жизненного цикла продукта, при этом в каждой фазе их задачи различаются. Здесь рассматриваются следующие фазы жизненного цикла: спецификация, производство, поставка, использование и обновление.

Процесс оценки качества данных представляет собой последовательность этапов для получения результата качества данных.

9.1.2 Последовательность выполнения процесса

Процесс оценки качества представляет собой последовательность шагов, выполняемых для получения результата оценки качества. Рисунок 12 иллюстрирует возможную последовательность действий для оценки качества данных. См. также приложение Е, в котором дано описание концепции оценки и составления отчетности по качеству данных.

Когда оцениваемые пространственные данные неоднородны и разные части имеют разное качество, то тесты должны быть выполнены соответственно для различных частей.

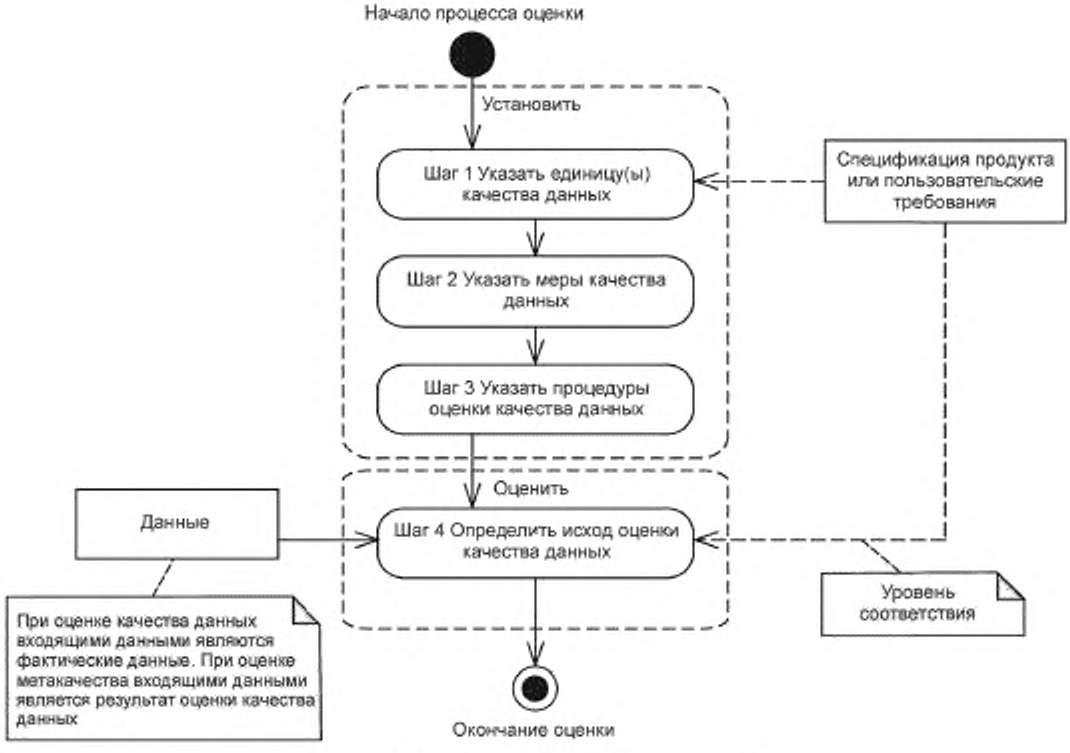


Рисунок 12 — Процесс оценки качества данных

9.1.3 Этапы процесса (process steps)

В таблице 1 представлены этапы процесса.

Таблица 1 — Этапы процесса

Этап процесса	Процесс	Описание
1	Указать единицу(ы) качества данных	Единица качества данных состоит из области определения и элемента (элементов) качества (см. 7.2). Следует использовать все элементы качества, релевантные данным, для которых будет описываться качество. Примечание — Тестируемые элементы качества данных описаны в 7.3, в приложении I приведены рекомендации по использованию элементов качества
2	Указать меры качества данных	В случае применимости для каждого элемента качества данных следует определять меру. В приложении D приведен список мер качества данных
3	Указать процедуры оценки качества данных	Процедура оценки качества данных заключается в применении одного или более методов оценки
4	Определить итог оценки качества данных	Результатом является итог применения оценки
В случае невозможности определения мер возможно использование описательного результата		

Оценка метакачества может осуществляться после получения итога оценки качества. Последовательность, описанная выше, также применима к оценке метакачества с учетом следующих этапов процесса: указать элемент метакачества и оценку качества, для которых оценивается метакачество, затем указать меру и метод оценки и определить итог оценки метакачества.

9.2 Методы оценки качества данных (data quality evaluation methods)

9.2.1 Классификация методов оценки качества данных

Процедура оценки качества данных включает один или более методов оценки качества данных. Методы оценки качества данных можно разделить на два основных класса: прямая (direct) оценка и косвенная (indirect) оценка. Методы прямой оценки определяют качество данных путем сравнения данных с внутренней и/или внешней информацией. Методы косвенной оценки выводят или оценивают качество данных, используя информацию о данных, такую как происхождение. Применение методов прямой оценки предпочтительнее косвенной оценки. Методы прямой оценки, в свою очередь, классифицируются по источнику информации, необходимой для выполнения оценки внутренней или внешней.

На рисунке 13 представлены классы описываемых методов оценки.

Примечание — Происхождение описано в *ГОСТ Р 57668*.

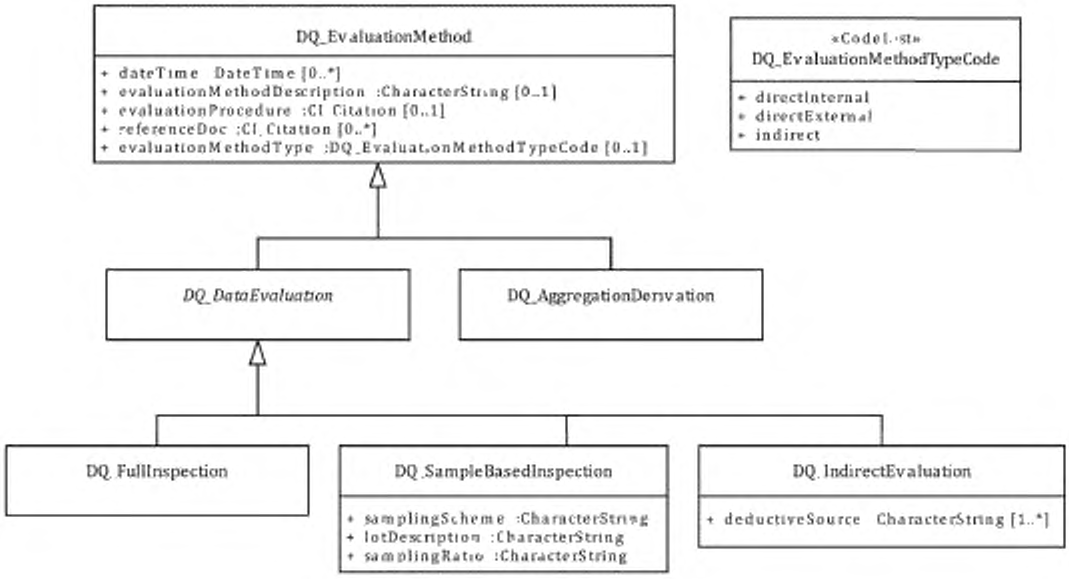


Рисунок 13 — Методы оценки качества данных

9.2.2 Прямая оценка (direct evaluation)

Метод прямой оценки — это метод оценки качества набора данных, основанный на проверке элементов в наборе данных.

Методы прямой оценки могут быть классифицированы на внутренние и внешние. Внутренняя прямая (internal direct) оценка качества данных использует только те данные, которые содержатся в оцениваемом наборе данных. Внешняя прямая (external direct) оценка качества требует применения эталонных данных, внешних по отношению к тестируемому набору данных.

Примечание — Эталонные данные — это данные, принятые как представляющие предметную область.

Как для внешних, так и для внутренних методов оценки может использоваться один из следующих методов проверки:

- полный контроль (full inspection);
- выборочный контроль (sampling).

При полном контроле тестируется каждый элемент в генеральной совокупности, определенной областью качества данных.

Примечание — Полный контроль больше всего подходит для небольших генеральных совокупностей или для тестов, которые могут быть выполнены в автоматическом режиме.

Выборочный контроль выполняется для проверки поднаборов пространственных данных, заданных областью качества данных.

Примечание — Примеры методов выборочной оценки даны в приложении F.

9.2.3 Косвенная оценка (indirect evaluation)

Метод косвенной оценки — это метод оценки качества набора данных на основе внешних знаний или накопленного опыта работы с информационным продуктом и может быть субъективным.

Эти внешние знания могут включать (но не ограничиваться ими) неколичественную информацию по качеству, такую как использование, происхождение и назначение (см. *ГОСТ Р 57668*) или другую отчетную информацию о качестве набора данных или данных, использованных для набора данных. Качество данных может быть оценено, например, на основе знаний об источнике, инструментах и методах, использованных для сбора данных и оцененных по отношению к процедурам и спецификациям, разработанным для этого продукта. Косвенная оценка качества данных также может быть основана только на опыте. Если указывается, что выполнена косвенная оценка, то должно быть указано также, как эта оценка была осуществлена.

В некоторых случаях может быть затруднительно или даже невозможно описать косвенно оцененное качество данных в виде количественного результата. В таких случаях качество данных может быть описано в текстовой форме с использованием описательного результата (см. 7.4.4.4).

9.3 Обобщение и преобразование (aggregation and derivation)

Дополнительные результаты можно получить посредством обобщения или преобразования существующих результатов без проведения новой оценки качества данных.

Обобщение объединяет результаты качества из проведенных оценок качества данных, основанных на различных элементах качества данных или различных областях определения качества данных.

Дополнительные результаты также могут быть выведены из существующих результатов, например, когда результат соответствия получен путем сравнения количественного результата с уровнем соответствия. Это полезно, например, если результат выражен не уровнем соответствия.

Примечания

1 Обобщение может быть применено для обобщения результатов различных элементов качества с целью описания их соответствия спецификации информационного продукта.

2 Подробнее об обобщении см. в приложении J. Как излагать информацию по обобщению, описано в 10.2.1 и приложении Е.

3 Как излагать информацию по преобразованию, описано в 10.2.2 и приложении Е.

Пример — Если уровень достоверности результата равен 95 % и уровень достоверности уровня соответствия 99 %, то результат может быть пересчитан, чтобы иметь такой же уровень достоверности уровня соответствия.

10 Отчетность о качестве данных (data quality reporting)

10.1 Общие положения

Качество данных должно быть представлено в виде метаданных в соответствии с разделом 7, разделом 10, приложением С, а также *ГОСТ Р 57668* и *ГОСТ Р 57656*.

С целью предоставления более подробной информации, помимо метаданных, может быть дополнительно создан самостоятельный отчет по оценке качества. Его структура определяется в свободной форме. Однако самостоятельный отчет по оценке качества не должен подменять метаданные. Метаданные должны содержать ссылку на самостоятельный отчет по оценке качества, если таковой имеется (см. рисунок 14).

Примечания

1 Более подробную информацию о предоставлении отчета по качеству данных, а также дополнительной роли между метаданными и самостоятельными отчетами по качеству см. также в В.4.3.2 приложения В.

2 Примеры отчетности по качеству данных приведены в Е.4 приложения Е.

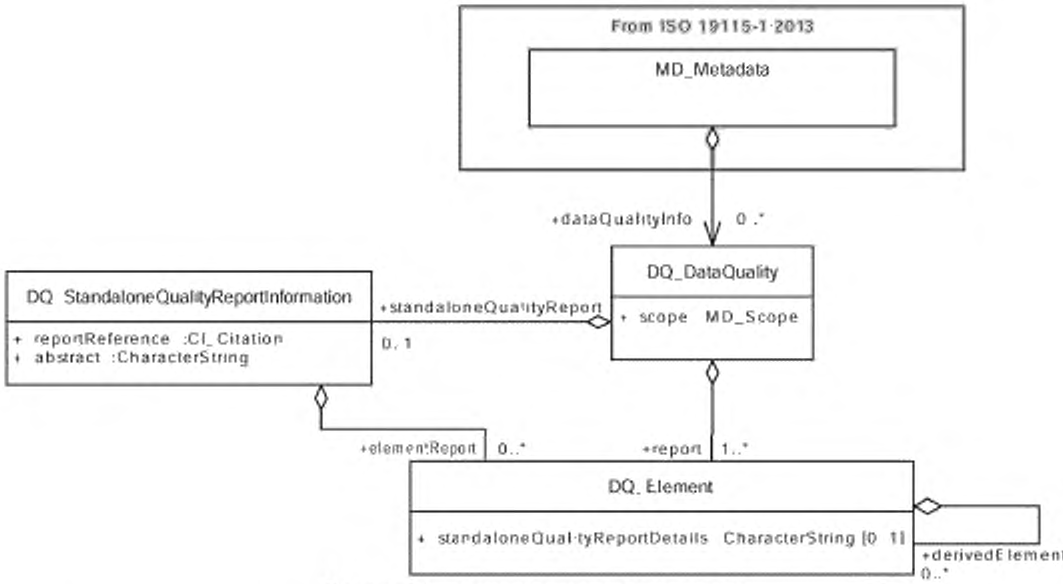


Рисунок 14 — Отчетность о качестве данных

10.2 Частные случаи

10.2.1 Отчетность по обобщению (обобщенные результаты)

При обобщении результатов должен предоставляться самостоятельный отчет о качестве (standalone quality report) для дополнения информации, представленной в метаданных. В рамках этого самостоятельного отчета о качестве должна быть представлена детальная информация по исходному результату с мерой (мерами) и процедурой (процедурами) оценки, обобщенному результату и методу обобщения.

В метаданных:

- когда несколько показателей качества одного элемента качества объединяются в один показатель этого элемента, то этот показатель должен быть представлен в метаданных в виде результата этого элемента качества данных. Примеры представлены в Е.4.1.2 и Е.4.1.3 приложения Е.
- когда несколько показателей качества различных элементов объединяются в один показатель, то это должно быть отражено в метаданных в виде результата элемента применения (DQ_UsabilityElement). Пример представлен в Е.4.1.4 приложения Е.

В обоих случаях в метаданных должна даваться по крайней мере ссылка на первоначальный результат проверки качества данных, также может быть представлена информация о мере обобщения и методе обобщения.

10.2.2 Отчетность по преобразованию (производные результаты)

Когда только производные результаты представляются в метаданных, то необходимо сгенерировать самостоятельный отчет по оценке качества, чтобы предоставить исходные результаты по качеству данных, из которых был получен производный результат. В метаданные должна быть включена ссылка на самостоятельный отчет по оценке качества и исходный результат качества данных.

Пример — Результат соответствия часто является производным из количественного результата. Если в метаданных указывается только результат соответствия, то тогда количественные результаты должны быть представлены в самостоятельном отчете по оценке качества.

10.2.3 Ссылка на результат проверки качества исходных данных

Когда производный или обобщенный результат(ы) представляются в метаданных, то ссылка на первоначальный результат проверки качества данных может быть дана с использованием двух атрибутов:

- атрибут derivedElement ссылается на элемент качества [и его результат(ы)], описанные в метаданных;
- атрибут standaloneQualityReportDetails ссылается на часть самостоятельного отчета о качестве, где описан исходный результат(ы).

Приложение А
(обязательное)

Комплекс проверок

А.1 Наименование проверки: Процесс оценки качества

a) Цель проверки: убедиться в правильности процесса оценки качества данных.

b) Метод проверки: проверить, включает ли процесс оценки качества все пункты, указанные в 9.1.3. Что предполагает:

1) определение положений спецификации на информационный продукт или требований пользователя, имеющих отношение к качеству данных, и их использование для идентификации применимых элементов качества данных и их соответствующей области. Сравнение применяемых элементов качества данных с оцениваемыми элементами качества данных для удостоверения в том, что все соответствующие элементы качества данных были идентифицированы и оценены по соответствующей области определения;

2) проверку того, что мера качества данных, применяемая для каждой оценки качества, уместна в соответствии с положениями спецификации на информационный продукт или требованиями пользователя;

3) проверка того, что процедура оценки качества, примененная для каждой оценки качества данных, уместна в соответствии с положениями спецификации информационного продукта или требованиями пользователя.

c) Ссылка: 9.1.

d) Тип проверки: предварительная.

А.2 Наименование проверки: Метаданные о качестве данных

a) Цель проверки: убедиться в том, что метаданные о качестве данных моделируются в соответствии с моделями UML и словарем данных.

b) Метод проверки: проверить, что метаданные содержат соответствующие компоненты качества данных и следуют правилам для каждого компонента.

c) Ссылка: раздел 7, раздел 10 и приложение С.

d) Тип проверки: предварительная.

А.3 Наименование проверки: Соответствие метаданных

a) Цель проверки: убедиться в том, что метаданные о качестве данных представлены в соответствии с *ГОСТ Р 57668* и ГОСТ Р *57656*.

b) Метод проверки: проверить тесты, представленные в *ГОСТ Р 57668*, в А.2.1, А.2.2, А.2.3, А.2.4, А.2.5 приложения А.

c) Ссылка: *ГОСТ Р 57668*, в А.2.1, А.2.2, А.2.3, А.2.4, А.2.5 приложения А.

d) Тип проверки: предварительная.

А.4 Наименование проверки: Независимый отчет по оценке качества

a) Цель проверки: убедиться в том, что самостоятельный отчет по оценке качества включает в себя разделы по всем соответствующим аспектам качества и что описание всех компонентов качества данных следует правилам, определенным в настоящем стандарте.

b) Метод проверки: проверить, содержит ли самостоятельный отчет по оценке качества все соответствующие компоненты.

c) Ссылка: раздел 7 и раздел 10.

d) Тип проверки: предварительная.

А.5 Наименование проверки: Меры качества данных

a) Цель проверки: убедиться в том, что мера качества данных структурно и семантически точно определена.

b) Метод проверки: проверить, описаны ли использованные меры качества, как это указано в разделе 8, и смоделированы ли в соответствии с моделью UML и словарем данных.

c) Ссылка: раздел 8 и приложение С.

d) Тип проверки: предварительная.

Приложение В
(справочное)

Концепции оценки качества данных и их применение

В.1 Концептуальные основы качества данных

Набор данных может быть создан для конкретного приложения или для нескольких возможных приложений. Качество набора данных может быть определено только при наличии информации об оценке элементов качества данных и в некоторых случаях косвенно — на основе неколичественной качественной информации по использованию, происхождению и назначению (см. *ГОСТ Р 57668*). Элементы качества данных оценивают расхождение между набором данных и предметной областью (т. е. правильным набором данных, который соответствует спецификации продукта). Неколичественная информация о качестве предоставляет общую информацию, из которой можно получить связанные с качеством знания.

Концепции оценки качества данных предоставляют важную основу для разработчиков, а также пользователей данных. Разработчик данных получает средства для проверки того, насколько набор данных отражает его предметную область согласно спецификации продукта. Пользователи данных могут оценить качество набора данных для удостоверения в том, что набор данных удовлетворяет требованиям пользовательского применения (см. рисунок В.1).

Заявленные результаты качества являются действительными в отношении спецификации информационного продукта или требований пользователей. Если они изменились, то оценка качества должна быть повторена в отношении измененной спецификации или требований. Следует проявлять осторожность при сравнении результатов качества там, где предметная область отличается. Типичным примером этого является преобразование модели в инфраструктурах пространственных данных или при генерализации. Например, при изменении геометрии типа объекта также изменяются результаты позиционной точности.



Рисунок В.1 — Концептуальные основы оценки качества данных

В.2 Структура наборов данных и компонентов для описания качества

Набор данных может принадлежать комплекту наборов данных, что означает, что все наборы данных комплекта основаны на одной и той же спецификации информационного продукта. Качество всех наборов данных, принадлежащих одному комплекту набора данных, может быть одинаковым.

Набор данных может рассматриваться как содержащий большое, хотя и конечное, число поднаборов данных. Поднаборы данных, объединенные принадлежностью к одному и тому же типу объекта, атрибуту объекта или отношению, критерию сбора или к одной и той же географической и временной протяженности, часто имеют одинаковое качество. Поднабор данных минимально может состоять из экземпляра объекта, значения атрибута или вида отношения. Концепция оценки качества данных позволяет каждому экземпляру объекта, значению атрибута и виду отношения объекта иметь собственное качество. Качество поднабора данных в наборе данных может не совпадать с качеством остальной части набора данных, к которому они принадлежат. Концепции оценки качества данных допускают выдачу информации о качестве набора данных и, кроме того, другой информации о качестве поднабора данных, идентифицируя эти группы как данные, которые определяются спецификой области определения качества данных. Информация о качестве, представленная для нескольких областей определения качества данных, меньше, чем та, что выводится для набора данных, но представляет более полную и подробную картину качества, чем общее качество полного набора данных.

Примечание — Для разработчика данных спецификация информационного продукта описывает предметную область и содержит правила для построения набора данных. Для пользователя данных требования описывают предметную область, которая может совпадать или не совпадать с предметной областью набора данных. Качество набора данных определяется тем, насколько хорошо он представляет предметную область. Качество одного и того же набора данных может отличаться в зависимости от предметной области, относительно которой он оценивается.

Качество набора данных описывается элементами качества данных и их дескрипторами. Некоторая информация, относящаяся к качеству, также может быть получена с помощью неколичественных элементов использования, происхождения и назначения.

Метакачество обеспечивает информацию по качеству в отношении оценки качества.

Элементы качества данных позволяют оценить, насколько набор данных соответствует критериям, изложенным в спецификации информационного продукта, или требованиям пользователя. Элементы качества данных оценивают различными способами и на различных этапах жизненного цикла набора данных. Концепция оценки качества данных подразумевает, что не все элементы качества данных применимы ко всем типам наборов данных. Помимо этого некоторые элементы качества данных являются применимыми к большим наборам данных, в то время как другие больше применимы к подмножеству данных в более крупном наборе данных. Некоторые элементы качества данных применимы как для единичных экземпляров данных, так и для большего количества, а некоторые применимы только для множества экземпляров.

Настоящий стандарт определяет элементы качества данных в первую очередь как средства идентификации и выдачи различных видов информации о качестве, также он подразумевает, что элементы качества данных часто взаимосвязаны. Например, ошибка в координатах может повлечь за собой по меньшей мере два типа ошибок — позиционную и топологическую (см. приложение I). Значение элементов качества данных с точки зрения продукта и способа, которым обрабатываются элементы качества данных, находится в компетенции оценщика качества.

В.3 Когда применять процедуры оценки качества

Процедуры оценки качества могут применяться на различных этапах жизненного цикла продукта. Оценка качества может быть применена на следующих стадиях жизненного цикла продукта:

- разработка спецификации информационного продукта или требований пользователя: при разработке спецификации или определении требований пользователя процедуры оценки качества могут быть использованы для установления уровней соответствия качества, которым должен отвечать конечный продукт. Спецификация информационного продукта или требования пользователя могут включать уровни соответствия качества для данных и процедуры оценки качества, которые должны применяться в процессе производства и обновления;
- контроль качества при создании набора данных: На стадии производства разработчик может применять процедуры оценки качества, которые явно установлены или же не включены в спецификацию информационного продукта, как часть процесса по контролю качества. Описание процедур оценки качества, применяемых для контроля качества продукции, рекомендуется представлять как метаданные происхождения, включая, в частности, оценку качества применяемых процедур, установленные уровни качества соответствия и результаты;
- проверка на соответствие спецификации информационного продукта: по завершении производства продукта рекомендуется использовать процесс оценки качества для формирования и представления результатов качества данных. По этим результатам можно определить, насколько набор данных соответствует или не соответствует спецификации информационного продукта. Если набор данных проходит проверку (состоящую из серии процедур оценки качества), то его считают готовым к использованию. Результаты по операции проверки следует

выдавать в соответствии с разделом 10. Пример описания оценки и выдачи информации о качестве данных см. в приложении Е. Итогом проверки будет принятие либо непринятие набора данных. Если набор данных будет отклонен, то после исправления данных необходимо будет провести новую проверку, и так до тех пор, пока продукт не будет соответствовать спецификации информационного продукта;

- оценка соответствия набора данных требованиям пользователя: процедуры оценки качества могут быть использованы для удостоверения в том, что набор данных отвечает уровням соответствия качества, установленным требованиями пользователя. При анализе соответствия набора данных требованиям пользователя могут применяться как косвенные, так и прямые методы;

- контроль качества при обновлении набора данных: процедуры оценки качества применяются к операциям обновления набора данных как для используемых элементов обновления, так и для сравнения качества набора данных до и после обновления.

В.4 Выдача информации о качестве

В.4.1 Зачем выдавать информацию о качестве

Необходимость выдавать информацию о качестве данных существует по ряду причин, включая следующие:

- способствовать обнаружению и поддерживать использование набора данных;
- показывать соответствие спецификации информационного продукта или требованиям пользователя;
- как часть инициатив менеджмента поставщика;
- уменьшить поток суждений о качестве информации, полученной из набора данных;
- обеспечить рациональное (оптимальное) принятие решения, когда известно, что все данные содержат дефекты.

В.4.2 Когда выдавать информацию о качестве

Наборы данных постоянно создают, обновляют и объединяют, в результате чего качество или компонент качества набора данных может изменяться. На информацию о качестве набора данных могут повлиять три условия:

- когда некоторый объем данных удаляется из набора данных, изменяется или добавляется к набору данных;
- когда изменяется спецификация набора данных или устанавливаются требования нового пользователя к качеству данных;
- когда изменяется реальный мир.

Первое условие — изменение набора данных — может происходить довольно часто. Многие наборы данных не являются статическими. Происходят увеличение обмена информацией, использование наборов данных в различных целях и обусловленное этим обновление и уточнение наборов данных для удовлетворения различных целей. Если информация о качестве некоторого набора данных изменяется в результате его модификации, то качество этого набора данных должно быть заново оценено и уточнено.

При создании набора данных необходимо обеспечить полное представление обо всех применяемых элементах качества данных. Первоначально могут сообщаться сведения об использовании набора данных его разработчиком (в предположении, что разработчик данных действительно использует такой набор данных). В дальнейшем пользователи данных должны предоставлять информацию об использовании набора данных, если оно отличается от его прямого предназначения. Благодаря таким сообщениям появится возможность постоянного обновления этого описательного элемента качества данных для отражения в нем непредвиденных вариантов использования.

Второе условие — изменение спецификации набора данных — чаще всего происходит перед созданием исходного набора данных и выпуском информации о качестве. Вполне вероятно, что по мере использования некоторого набора данных его спецификация будет обновляться для того, чтобы будущие изменения этого набора данных лучше отвечали фактическим потребностям. При изменении спецификации информационного продукта изменяется также и качество текущего набора данных. Информация о качестве набора данных всегда должна отражать текущий набор данных с учетом текущей спецификации информационного продукта.

Третье условие — изменение реального мира — происходит постоянно. Изменения могут быть вызваны природными явлениями, такими как движения земной коры или эрозия почвы, но чаще всего они являются результатом деятельности человека. Изменения, как правило, наступают очень быстро и внезапно. По этой причине время сбора данных, также как и время оценки качества, играют важную роль при оценке качества набора данных. В некоторых случаях даже скорость изменения данных представляет интерес. А в некоторых случаях интерес может также представлять и частота изменений набора данных. Тем не менее настоящий стандарт признает, что выпуск новой информации о качестве данных при каждых изменениях реального мира не представляется возможным.

В.4.3 Как следует выдавать информацию о качестве

В.4.3.1 Иерархический принцип

Данный стандарт устанавливает принцип иерархического уровня:

Качество данных, определяемое на верхнем уровне (например, комплект наборов данных), применимо и на более низком уровне (например, набор данных), см. таблицу В.1. Если качество данных отличается на верхнем и нижнем уровнях, то необходимо получить дополнительную информацию на более низком уровне.

Таблица В.1 — Иерархические уровни

Верхний уровень ↑	Комплекты	
	Набор данных	
↓	Поднабор	
Нижний уровень	Тип объекта	Тип атрибута
	Экземпляр объекта	Экземпляр атрибута

Примечание — Информация о качестве экземпляра объекта, атрибута объекта или отношений между объектами может быть выдана как атрибут этого экземпляра согласно [1].

В.4.3.2 Метаданные и самостоятельный отчет по оценке качества

В.4.3.2.1 Общие положения

Информацию о качестве можно выпускать в виде метаданных и автономного отчета о качестве. Эти два механизма дополняют друг друга, позволяя выдавать информацию об оценке качества данных с различными уровнями детализации:

- метаданные направлены на предоставление краткой, синтетической и общеструктурированной информации для обеспечения операционной совместимости метаданных и использования веб-сервисов;
- самостоятельный отчет по оценке качества может быть использован для получения более подробной информации об оценке качества данных. Он должен прилагаться к набору данных или продукту для возможности непосредственного ознакомления с ним.

Например, в случае обобщения различных результатов качества самостоятельный отчет по оценке качества обеспечит полную информацию по исходным результатам (с применимыми процедурами оценки и мерами), результату обобщения и методу обобщения, в то время как метаданные могут только описать результат обобщения со ссылкой на исходные результаты, описанные в самостоятельном отчете по оценке качества.

В.4.3.2.2 Выдача информации о качестве в виде метаданных

Класс MD_Metadata, согласно *ГОСТ Р 57668*, обобщает ноль, одну или несколько единиц качества данных (экземпляры класса DQ_DataQuality, согласно настоящему стандарту) (см. рисунок В.2).

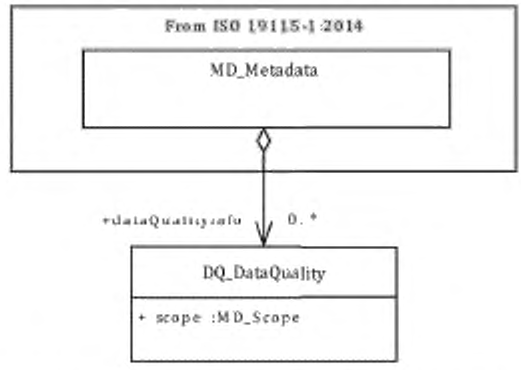


Рисунок В.2 — Информация о качестве данных

В.4.3.2.3 Выдача информации о качестве в самостоятельном отчете

Стандартизация терминологии элементов качества данных и структуры представления основной информации о качестве данных обеспечивают лучшее понимание и сравнение результатов оценки качества.

В самостоятельный отчет по оценке качества следует включать область применения для однозначного определения пространственно-временной протяженности оцениваемого набора данных.

Каждый отчет должен содержать достаточный объем информации для полного описания соответствующих аспектов качества данных и их результатов. Описание может быть представлено в форме ссылки на документы, такие как спецификация информационного продукта или каталог мер.

Полная структура такого самостоятельного отчета по оценке качества не была стандартизирована, чтобы каждая конкретная организация смогла адаптировать его для своих собственных нужд, методик и процедур оценки. Он может представлять собой свободный текст. Однако информация о качестве должна быть представлена в краткой, понятной и легкодоступной форме. Пример самостоятельного отчета по оценке качества приведен в приложении Е.

Приложение С
(обязательное)

Словарь данных для оценки качества

С.1 Обзор словаря данных

С.1.1 Введение

Словарь данных описывает характеристики модели качества данных, определенной в разделах 7, 8, 9 и 10. Словарь состоит из таблиц с графами и строками в иерархии для установления отношений и организации информации.

Затемненные строки таблицы представляют классы. Незатемненные строки таблицы представляют атрибуты класса и ассоциации. Классы и атрибуты классов в таблицах словаря данных определяются шестью графами таблицы, описанной в С.1.2—С.1.7.

С.1.2 Имя/ролевое имя

Имя роли — это условное обозначение, относящееся к классу или атрибуту класса. Имена классов начинаются с прописной буквы. Пробелы не используются в имени класса. Вместо этого несколько слов соединяются, и каждое новое подслово начинается с заглавной буквы (например: XnnnYmmm). Имена классов уникальны в пределах всего словаря данных настоящего стандарта. Имена атрибутов класса уникальны в пределах класса, а не всего словаря данных настоящего стандарта. Имена атрибутов класса устанавливаются уникальными в пределах приложения посредством сочетания имени класса и имени атрибута класса. Ролевые имена используются для идентификации абстрактных моделей ассоциаций и начинаются с префикса «Role name», чтобы отличить их от других атрибутов класса. Имена и имена ролей могут быть на любом другом языке, отличающемся от используемого в настоящем стандарте.

С.1.3 Определение

Определение — это описание класса или атрибута класса.

С.1.4 Признак обязательности

С.1.4.1 Общие положения

Признак обязательности — это дескриптор, указывающий, подлежит ли класс или атрибут класса обязательному документированию или указывается (т. е. содержит значения) только иногда. Этот дескриптор может иметь следующие значения: О (обязательный), У (условный) или Н (необязательный).

С.1.4.2 Обязательный (О)

Признак «О» означает, что класс или атрибут класса должны быть указаны.

С.1.4.3 Условный (У)

Признак «У» определяет управляемое электронное условие, при котором хотя бы один класс, атрибут класса или ассоциация являются обязательными. Признак «У» используется в следующих трех ситуациях:

- представление выбора между двумя или несколькими параметрами. По крайней мере один из параметров является обязательным и должен быть задокументирован;
- документирование класса, атрибута класса или ассоциации, если другой класс был задокументирован;
- документирование атрибута класса или ассоциации, если конкретное значение другого атрибута класса уже задокументировано. Для облегчения чтения людьми конкретное значение используется в обычном тексте. Однако для проверки условия в электронном интерфейсе пользователя должен использоваться код.

Если ответ на условие положительный, то класс, атрибуты класса или ассоциация должны быть обязательными.

С.1.4.4 Необязательный (Н)

Признак «Н» означает, что класс, атрибут класса или ассоциация метаданных являются необязательными и могут как присутствовать, так и отсутствовать. В настоящем стандарте для обеспечения интероперабельности между пользователями пространственных данных и производителями определены необязательные классы метаданных и необязательные атрибуты метаданных. Если необязательный класс не используется, то элементы, содержащиеся в этом классе (включая обязательные элементы), тоже не используются.

Необязательные классы могут иметь обязательные элементы; эти элементы становятся обязательными, только если необязательный класс используется.

С.1.5 Максимум вхождений (МВ)

Признак максимума вхождений определяет максимально допустимое количество экземпляров класса, атрибута класса или ассоциации. Единичное вхождение показано как «1»; повторяющееся, без ограничений, вхождение обозначено «N». Допускается фиксированное число вхождений, отличающееся от одного, которое будет обозначаться соответствующим числом (т. е. «2», «3» и т. д.).

С.1.6 Тип данных

Тип данных определяет множество различных значений для представления атрибутов класса: например, integer, real, string, DateTime и Boolean. Атрибут типа данных также используется для определения классов, стереотипов и ассоциаций классов.

Примечание — Типы данных определены в [3].

С.1.7 Область допустимых значений (домен)

Для класса (затемненные строки) домен указывает номера строк, относящихся к атрибутам классов и ассоциациям этого класса.

Для атрибута класса или ассоциации домен определяет допустимые значения или использование произвольного текста. Понятие «произвольный текст» (free text) означает, что никаких ограничений на содержание этого поля не накладывается. Целочисленные коды должны применяться для представления значений доменов, содержащих кодовые списки.

С.2 Словарь данных для пакетов качества данных

С.2.1 Качество данных

С.2.1.1 Общие положения

Глобальная модель UML для всего пакета качества данных представлена на рисунке 2. Модель UML приведена на рисунках 3 и 15.

Таблица С.1 — Качество данных

	Имя/ролевое имя	Определение	Признак обязательности	Максимум вхождений (MB)	Тип данных	Домен
1	DQ_DataQuality	Информация о качестве данных, заданных областью определения качества данных	Признак обязательности ссылочного объекта	MB ссылочного объекта	Aggregated Class (MD_ Metadata)	Строки 2—4
2	Scope	Конкретные данные, к которым относится информация о качестве	O	1	Class	MD_Scope < < DataType > > (ГОСТ Р ИСО 19115-1:2014)
3	*Role name:* report		O	N	Association	DQ_Element (Abstract class) (C.2.1.2)
4	*Role name:* standalone-QualityReport		H	1	Association	DQ_Standalone-QualityReportInfor-mation (C.2.1.6)

С.2.1.2 Элементы качества данных

Модель UML представлена на рисунках 4, 5, 10 и 14.

Таблица С.2 — Элементы качества данных

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (MB)	Тип данных	Домен
5	*DQ_Element*	Количественная информация о качестве данных	Признак обязательности ссылочного объекта	MB ссылочного объекта	Aggregated Class (DQ_Data-Quality) (Abstract class)	Строки 6—10
6	standalone-QualityReport-Details	Раздел в standalone-QualityReport, в котором этот элемент качества данных или любой связанный элемент качества данных (исходные результаты в случае производного или обобщения) описываются	H	1	Character string	Текст

Продолжение таблицы С.2

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
7	*Role name:* measure	Ссылка на применяемую меру	Н	1	Association	DQ_Measure-Reference (С.2.1.3)
8	*Role name:* evaluation-Method	Информация по оценке	Н	1	Association	DQ_Evaluation-Method (С.2.1.4)
9	*Role name:* result	Значение (или набор значений), полученное в результате применения мер качества данных или результате оценки полученного значения (или набора значений) в сравнении с заданным приемлемым уровнем соответствия качества	О	N	Association	DQ_Result (Abstract class) (С.2.1.5)
10	*Role name:* derived-Element	В случае агрегирования или вывода указывает исходный элемент	Н	N	Association	DQ_Element (Abstract class) (С.2.1.2)
11	*DQ_Completeness*	Присутствие или отсутствие данных в наборе, их атрибутов и отношений между ними	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_Element) (Abstract class)	Строки 6—10
12	DQ_CompletenessCommission	Избыточные данные, присутствующие в наборе	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_Completeness)	Строки 6—10
13	DQ_CompletenessOmission	Данные, отсутствующие в наборе данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_Completeness)	Строки 6—10
14	*DQ_LogicalConsistency*	Различные аспекты соблюдения логических правил структур данных, атрибутов и отношений (структуры данных могут быть концептуальными, логическими или физическими)	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_Element) (Abstract class)	Строки 6—10
15	DQ_ Conceptual-Consistency	Соблюдение правил концептуальной схемы	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_Logical Consistency)	Строки 6—10
16	DQ_Domain-Consistency	Соответствие значений доменам допустимых значений	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_Logical Consistency)	Строки 6—10

Продолжение таблицы С.2

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
17	DQ_Format-Consistency	Порядок хранения данных в соответствии с физической структурой набора данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Logical Consist-ency)	Строки 6—10
18	DQ_TopologicalConsistency	Правильность представления закодированных топологических характеристик набора данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Logical Consist-ency)	Строки 6—10
19	*DQ_Positional-Accuracy*	Точность позиционирования данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Element) (Abstract class)	Строки 6—10
20	DQ_Absolute-External-Positional-Accuracy	Соответствие описанных значений положения объектов из набора данных их значениям, принятым в качестве или являющимся истинными	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Positional Accuracy)	Строки 6—10
21	DQ_Rela-tiveInternal-Positional-Accuracy	Соответствие относительных положений объектов в области определения их относительному положению, принятому в качестве или являющемуся истинным	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Positional Accuracy)	Строки 6—10
22	DQ_Gridded-DataPositional-Accuracy	Соответствие значений положения сеточных данных их реальным значениям положения	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Positional Accuracy)	Строки 6—10
23	*DQ_Temporal-Quality*	Точность временных атрибутов и временных отношений объектов	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Element) (Abstract class)	Строки 6—10
24	*DQ_Accuracy-OfATime-Measurement*	Правильность временных измерений элемента (вывод ошибок при временном измерении)	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Temporal Quality)	Строки 6—10
25	DQ_Temporal-Consistency	Правильность упорядоченных событий или последовательно-стей, если заявлено	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Temporal Quality)	Строки 6—10
26	DQ_Temporal-Validity	Актуальность данных, заданных областью определения, по отношению ко времени	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Temporal Quality)	Строки 6—10

Окончание таблицы С.2

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
27	DQ_Thematic-Accuracy	Точность количественных атрибутов и правильность неколичественных атрибутов, а также правильность классификации объектов и их отношений	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Element) (Abstract class)	Строки 6—10
28	DQ_Thematic-Classification-Correctness	Правильность выбора классов для объектов набора и их атрибутов в рамках предметной области	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Thematic Accuracy)	Строки 6—10
29	DQ_NonQuantitativeAttributeCorrectness	Правильность неколичественных атрибутов	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Thematic Accuracy)	Строки 6—10
30	DQ_QuantitativeAttribute-Accuracy	Правильность количественных атрибутов	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Thematic Accuracy)	Строки 6—10
31	DQ_Usability-Element	Степень соблюдения набором данных определенных требований	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Element)	Строки 6—10
32	*DQ_Metaquality*	Информация о достоверности результатов оценки качества данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Element) (Abstract class)	Строки 33 и 6—10
33	*Role name*: derivedElement	Производный элемент	О	1	Association	DQ_Element (Abstract class) (С.2.1.2)
34	DQ_Confidence	Достоверность результата оценки качества данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Metaquality)	Строки 33 и 6—10
35	DQ_Representativity	Степень того, насколько результат примененной выборки представляет данные в области определения качества данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Metaquality)	Строки 33 и 6—10
36	DQ_Homogeneity	Прогнозируемая или тестируемая однородность результатов, полученных для оценки качества данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Metaquality)	Строки 33 и 6—10

С.2.1.3 Указатель меры

Модель UML представлена на рисунке 6.

Таблица С.3 — Указатель меры

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (MB)	Тип данных	Домен
37	DQ_Measure-Reference	Ссылка на применяемую меру	Признак обязательности ссылочного объекта	MB ссылочного объекта	Aggregated Class (DQ_ Element)	Строки 38—40
38	measureIden-tification	Идентификатор меры, значение, уникально идентифицирующее меру в пространстве имен	H	1	Class	MD_Identifier < < Data Type > > *(см. ГОСТ Р 57668, В.3.3.3)*
39	nameOfMea-sure	Наименование процедуры оценки, применяемой к данным	У/если «measureI-dentification» не задокумен-тирован	N	Character string	Текст
40	measure-Description	Описание меры	H	1	Character string	Текст

С.2.1.4 Информация об оценке качества данных

Модель UML представлена на рисунках 7 и 13.

Таблица С.4 — Информация об оценке качества данных

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (MB)	Тип данных	Домен
41	DQ_Evalua-tionMethod	Описание примененного метода и процедуры оценки	Признак обязательности ссылочного объекта	MB ссылочного объекта	Aggregated Class (DQ_ Element)	Строки 42—46
42	evaluation-MethodType	Тип метода, используемого для оценки качества данных	H	1	Class	DQ_Evaluation-MethodTypeCode < < CodeList > > (С.3.2)
43	evaluation-MethodDe-scription	Описание метода оценки	H	1	Character-String	Текст
44	evaluation-Procedure	Ссылка на информацию о процедуре оценки	H	1	Class	CI_Citation < < Data Type > > *(см. ГОСТ Р 57668, В.3.2.1)*
45	referenceDoc	Информация о документах, на которые ссылают-ся при разработке и применении методов оценки качества данных	H	N	Class	CI_Citation < < Data Type > > *(см. ГОСТ Р 57668, В.3.2.1)*

Окончание таблицы С.4

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
46	dateTime	Дата или временной интервал, в котором применялась мера	Н	N	Class	DateTime (см. [3])
47	DQ_Data-Evaluation	Метод оценки данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Evaluation-Method) (Abstract class)	Строки 42—46
48	DQ_FullIn-spection	Полная проверка	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ DataEvalu-ation)	Строки 42—46
49	DQ_Indirect-Evaluation	Непрямая оценка	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ DataEvalu-ation)	Строки 42—46 и 50
50	deductive-Source	Информация, согласно которой данные используются в качестве источников дедуктивного метода оценки	О	1	Character-String	Текст
51	DQ_Sample-BasedIn-spection	Контроль на основе выборки	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ DataEvalua-tion)	Строки 42—46 и 52—54
52	sampling-Scheme	Информация о типе выборочной схемы и описании процедуры выбороч-ного контроля	О	1	Character-String	Текст
53	lotDescription	Информация по составлению партии	О	1	Character-String	Текст
54	samplingRatio	Информация о том, сколько в среднем выборок извлекается из каждой партии генеральной совокупности	О	1	Character-String	Текст
55	DQ_Aggrega-tionDerivation	Метод агрегирования или деривации	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Evaluation)	Строки 42—46

С.2.1.5 Результат оценки качества данных

Модель UML представлена на рисунке 8.

Таблица С.5 — Результат оценки качества данных

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
56	*DQ_Result*	Обобщение нескольких специфических классов результатов	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Aggregated Class (DQ_ Element) (Abstract class)	Строки 57—58
57	resultScope	Область определения результата	H	1	Class	MD_Scope *(ГОСТ Р 57668)*
58	dateTime	Дата генерирования результата	H	1	Class	DateTime (см. [6])
59	DQ_Conform-anceResult	Информация о результате сравнения полученного значения (или набора значений) с заданным допустимым уровнем качества	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Result)	Строки 60—62 и 57—58
60	specification	Ссылка на спецификацию продукта или требования пользователя, с которыми сравнивается результат	O	1	Class	CI_Citation < < Data Type > > *(см. ГОСТ Р 57668, таблица В.3.2.1)*
61	explanation	Разъяснение смысла соответствия для этого результата	H	1	Character-String	Текст
62	pass	Признак соответствия результата, где 0 означает несоответствие, а 1 — соответствие	O	1	Boolean	1 = да 0 = нет
63	DQ_Quantita-tiveResult	Значения или информация о значениях (или набора значений), полученных в результате применения меры качества данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Result)	Строки 64—66 и 57—58
64	value	Количественное значение или значения, определенные в результате выполнения процедуры оценки и в соответствии с типом значения и valueStructure, определенных для меры	O	N	Class	Record (см. [3])

Окончание таблицы С.5

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
65	valueUnit	Единица измерения результата оценки качества	Н	1	Class	UnitOfMeasure (см. [3])
66	value-RecordType	Тип значения результата оценки качества в зависимости от реализации	Н	1	Class	RecordType << Meta class >> (см. [3])
67	DQ_Descriptive-Result	Описательный результат оценки качества	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Specified Class (DQ_ Result)	Строки 68 и 57—58
68	statement	Текстовое выражение описательного результата	О	1	Character-String	Текст

С.2.1.6 Информация о самостоятельном отчете по оценке качества

Модель UML представлена на рисунке 14.

Таблица С.6 — Информация о самостоятельном отчете по качеству

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
69	DQ_Stan-daloneQuality-ReportInfor-mation	Ссылка на внешний самостоятельный отчет об оценке качества	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Class	Строки 70—71
70	reportRefe-rence	Ссылка на сводный самостоятельный отчет об оценке качества	О	1	Class	CI_Citation << Data Type >> (*см. ГОСТ Р 57668, таблица В.3.2.1)*
71	abstract	Краткое описание сводного самостоятельного отчета об оценке качества	О	1	Character-String	Текст

С.2.2 Мера качества данных

С.2.2.1 Общее представление

Модель UML мер представлена на рисунке 11.

С.2.2.2 Меры качества данных

Модель UML представлена на рисунке 11.

Таблица С.7 — Меры качества данных

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
72	DQM_ Measure	Мера качества данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Class	Строки 73—84

Продолжение таблицы С.7

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
73	measureIdentifier	Значение, уникально идентифицирующее меру в пространстве имен	O	1	Class	MD_Identifier < < Data Type > > *(см. ГОСТ Р 57668, таблица В.3.3.3)*
74	Name	Наименование меры качества, применяемой к данным	O	1	CharacterString	Текст
75	alias	Другое общепринятое наименование, сокращенное или краткое название той же меры качества данных	H	N	CharacterString	Текст
76	elementName	Наименование элемента качества данных для оценки качества	O	N	Class	TypeName < < type > > (см. [3])
77	definition	Определение фундаментальной концепции меры качества данных	O	1	CharacterString	Текст
78	description	Описание меры качества данных, включая все формулы и/или иллюстрации, необходимые для получения результата при применении меры	У/если описания недостаточно для понимания концепции меры качества	1	Class	DQM_Description < < Datatype > > (С.2.2.5)
79	valueType	Тип значения для вывода информации о результате оценки качества данных (должны быть одним из типов данных, определенных в [3])	O	1	Class	TypeName < < type > > (см. [3])
80	valueStructure	Структура для представления информации о комплексных результатах оценки качества	H	1	Class	DQM_ValueStructure < < CodeList > > (С.3.3)
81	Пример	Пример использования меры качества данных	H	N	Class	DQM_Description (С.2.2.5)
82	*Role name:* basicMeasure	Наименование базовой меры качества данных, из которой мера качества данных происходит	У/если происходит из базовой меры	1	Association	DQM_BasicMeasure (С.2.2.3)

Окончание таблицы С.7

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
83	*Role name:* sourceReference	Ссылка на источник элемента, который был взят из внешнего источника	У/если внешний источник существует	N	Association	DQM_SourceReference (С.2.2.6)
84	*Role name:* parameter	Вспомогательная переменная, используемая мерой качества данных, включая ее наименование, определение и в некоторых случаях ее описание	У/если потребуется	N	Association	DQM_Parameter (С.2.2.4)

С.2.2.3 Базовая мера качества данных

Модель UML представлена на рисунке 11.

Таблица С.8 — Базовая мера качества данных

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
85	DQM_BasicMeasure	Базовая мера качества данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Class	Строки 86—89
86	name	Наименование базовой меры качества, применяемой к данным	O	1	CharacterString	Текст
87	definition	Описание базовой меры качества данных	O	1	CharacterString	Текст
88	Пример	Пример использования меры качества данных	H	1	Class	DQM_Description < < Datatype > > (С.2.2.5)
89	valueType	Тип значения для вывода информации о результате оценки качества данных (должны быть одним из типов данных, определенных в [3])	O	1	Class	TypeName < < type > > (см. [3])

С.2.2.4 Параметр оценки качества данных

Модель UML представлена на рисунке 11.

Таблица С.9 — Параметр оценки качества данных

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
90	DQM_Parameter	Параметр оценки качества данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Class	Строки 91—95

Окончание таблицы С.9

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
91	name	Наименование параметра оценки качества данных	О	1	Character-String	Текст
92	definition	Определение параметра оценки качества данных	О	1	Character-String	Текст
93	description	Описание параметра оценки качества данных	Н	1	Class	DQM_Description < < Datatype > > (С.2.2.5)
94	valueType	Тип значения для вывода информации о результате оценки качества данных (должны быть одним из типов данных, определенных в [3])	О	1	Class	TypeName < < type > > (см. [3])
95	valueStructure	Структура параметра оценки качества данных	Н	1	Class	DQM_ValueStructure < < CodeList > > (С.3.3)

С.2.2.5 Описание меры качества данных

Модель UML представлена на рисунке 11.

Таблица С.10 — Дескриптор качества данных

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
96	DQM_Description	Описание меры качества данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Class	Строки 97—98
97	textDescription	Текстовое описание	О	1	Character-String	Текст
98	extendedDescription	Пример	Н	1	Class	MD_BrowseGraphic (см. *ГОСТ Р 57668, В.2.3.2)*

С.2.2.6 Указатель источника меры качества данных

Модель UML представлена на рисунке 11.

Таблица С.11 — Указание источника меры качества данных

	Имя/ролевое имя	Описание	Признак обязательности	Максимум вхождений (МВ)	Тип данных	Домен
99	DQM_ SourceReference	Ссылка на источник меры качества данных	Признак обязательности ссылочного объекта	МВ ссылочного объекта	Class	Строка 100
100	Citation	Ссылка на источник	О	1	Class	CI_Citation < < Data Type > > (см. *ГОСТ Р 57668)*

С.3 Списки значений

С.3.1 Введение

Классы со стереотипом <<CodeList>> представлены ниже. Они не содержат графы «признак обязательности», «максимум вхождений», «тип данных» и «домен».

С.3.2 Тип метода для оценки

Таблица С.12 — Тип метода для оценки

	Наименование	Домен	Определение
1	DQ_EvaluationMethodTypeCode	EvalMethTypeCd	Тип метода, используемого для оценки качества набора данных
2	directInternal	001	Метод оценки качества набора данных, базирующийся на исследовании элементов внутри набора, где вся информация, необходимая для оценки, уже содержится в наборе данных
3	directExternal	002	Метод оценки качества набора данных, базирующийся на исследовании элементов внутри набора, где требуется дополнительная информация, не содержащаяся в наборе данных
4	indirect	003	Метод оценки качества набора данных, базирующийся на внешних знаниях

С.3.3 Структура значений

Таблица С.13 — Структура значений

	Наименование	Код домена	Определение
1	DQM_ValueStructure	ValueStructureCd	
2	bag	001	Конечная, неупорядоченная совокупность связанных элементов (объектов или значений), в которой допускается их повторение [7]
3	set	002	Неупорядоченная совокупность взаимосвязанных неповторяющихся элементов [7]
4	sequence	003	Конечная упорядоченная совокупность взаимосвязанных элементов (объектов или значений), в которой допускается их повторение [7]
5	table	004	Упорядоченная структура данных, в которой каждый элемент может быть идентифицирован с помощью аргументов или ключей [8]
6	matrix	005	Прямоугольный массив чисел [9]
7	coverage	006	Модель объекта, представляющая собой функцию, возвращающую значения из диапазона допустимых значений для каждой точки в пространственной, временной или пространственно-временной области определения [10]

Приложение D
(обязательное)

Перечень стандартизированных мер качества данных

D.1 Введение

В данном приложении устанавливается и определяется перечень стандартизированных мер качества данных. Настоящее приложение определяет меры качества данных. С целью получения четко определенной и сравнимой информации о качестве настоятельно рекомендуется проводить оценку и отчетность по качеству данных с использованием таких мер качества.

D.2 Полнота

D.2.1 Присутствие

Меры качества данных для присутствия элементов качества данных представлены в таблицах D.1—D.4.

Таблица D.1 — Избыточный элемент

№	Компонент	Описание
1	Название	Избыточный элемент
2	Краткое название	—
3	Название элемента	Присутствие
4	Базовая мера	Индикатор ошибок
5	Определение	Признак того, что элемент неверно представлен в данных
6	Описание	—
7	Параметр	—
8	Тип значения	Булева переменная (значение «true» указывает на то, что элемент является избыточным)
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	True (в наборе данных в качестве зданий классифицировано больше элементов, чем в предметной области)
12	Идентификатор	1

Таблица D.2 — Количество избыточных элементов

№	Компонент	Описание
1	Название	Количество избыточных элементов
2	Краткое название	—
3	Название элемента	Присутствие
4	Базовая мера	Подсчет ошибок
5	Определение	Количество элементов в наборе данных или выборке, которые не должны быть представлены
6	Описание	—
7	Параметр	—
8	Тип значения	Integer (целое)
9	Структура значения	—

Окончание таблицы D.2

№	Компонент	Описание
10	Ссылка на источник	—
11	Пример	2 (в наборе данных содержится 12 зданий, в то время как в предметной области только 10)
12	Идентификатор	2

Таблица D.3 — Коэффициент избыточных элементов

№	Компонент	Описание
1	Название	Коэффициент избыточных элементов
2	Краткое название	—
3	Название элемента	Присутствие
4	Базовая мера	Частота ошибок
5	Определение	Число избыточных элементов в наборе данных или выборке по отношению к числу элементов, которые должны быть представлены
6	Описание	—
7	Параметр	—
8	Тип значения	Real (вещественное)
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	10 % (в наборе данных представлено на 10 % больше зданий, чем в предметной области)
12	Идентификатор	3

Таблица D.4 — Количество повторяющихся экземпляров объекта

№	Компонент	Описание
1	Название	Количество повторяющихся экземпляров объекта
2	Краткое название	—
3	Название элемента	Присутствие
4	Базовая мера	Подсчет ошибок
5	Определение	Общее количество точных дубликатов экземпляров объектов в данных
6	Описание	Количество всех элементов в наборе данных с дублированной геометрией, которые определяются некорректно
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	Объекты с идентичными атрибутами и координатами: ▲ две (или более) точки, расположенные поверх друг друга; — две (или более) кривые, расположенные поверх друг друга; □ две (или более) поверхности, расположенные поверх друг друга
12	Идентификатор	4

D.2.2 Отсутствие

Меры качества данных для элемента качества данных «отсутствие» представлены в таблицах D.5—D.7.

Таблица D.5 — Отсутствующий элемент

№	Компонент	Описание
1	Название	Отсутствующий элемент
2	Краткое название	—
3	Название элемента	Отсутствие
4	Базовая мера	Индикатор ошибок
5	Определение	Признак того, что элемент отсутствует в данных
6	Описание	—
7	Параметр	—
8	Тип значения	Булева переменная (значение «true» указывает на то, что элемент отсутствует)
9	Структура значения	—
10	Ссылка на источник	
11	Пример	Спецификацией информационного продукта задается отбор всех башен выше 300 м. Мера качества данных «отсутствующий элемент» позволяет оценщику данных или их пользователю выдать отчет о том, что конкретный элемент, в данном случае объект типа «башня» (название зависит от схемы приложения), отсутствует. Область определения качества данных: все башни высотой > 300. Пример результата оценки полноты определенного набора данных: Отсутствующий элемент = true для r Имя башни = Эйфелева башня
12	Идентификатор	5

Таблица D.6 — Число отсутствующих элементов

№	Компонент	Описание
1	Название	Количество отсутствующих элементов
2	Краткое название	—
3	Название элемента	Отсутствие
4	Базовая мера	Подсчет ошибок
5	Определение	Количество всех элементов, которые должны присутствовать в наборе данных или выборке, но отсутствуют
6	Описание	—
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	2 (в наборе данных присутствуют 10 домов, тогда как в предметной области — 12)
12	Идентификатор	6

Таблица D.7 — Коэффициент отсутствующих элементов

№	Компонент	Описание
1	Название	Коэффициент избыточных элементов
2	Краткое название	—
3	Название элемента	Отсутствие
4	Базовая мера	Частота ошибок
5	Определение	Число отсутствующих элементов в наборе данных или выборке по отношению к числу элементов, которые должны быть представлены
6	Описание	—
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	10 % (в наборе данных представлено на 10 % меньше домов, чем в предметной области)
12	Идентификатор	7

D.3 Логическая согласованность

D.3.1 Концептуальная согласованность

Меры качества данных для подэлемента качества данных «концептуальная согласованность» даны в таблицах D.8—D.13.

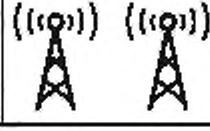
Таблица D.8 — Несогласованность с концептуальной схемой

№	Компонент	Описание
1	Название	Несогласованность с концептуальной схемой
2	Краткое название	—
3	Название элемента	Концептуальная согласованность
4	Базовая мера	Индикатор ошибки
5	Определение	Показатель того, что объект не отвечает правилам соответствующей концептуальной схемы
6	Описание	—
7	Параметр	—
8	Тип значения	Булева переменная (значение «true» указывает на то, что элемент не отвечает правилам соответствующей концептуальной схемы)
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	True (существует связь одного объекта, которая не определена в концептуальной схеме)
12	Идентификатор	8

Таблица D.9 — Согласованность с концептуальной схемой

№	Компонент	Описание
1	Название	Согласованность с концептуальной схемой
2	Краткое название	—
3	Название элемента	Концептуальная согласованность
4	Базовая мера	Индикатор корректности
5	Определение	Показатель того, что элемент отвечает правилам соответствующей концептуальной схемы
6	Описание	—
7	Параметр	—
8	Тип значения	Булева переменная (значение «true» указывает на то, что элемент отвечает правилам соответствующей концептуальной схемы)
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	9

Таблица D.10 — Число объектов, не согласованных с правилами концептуальной схемы

№	Компонент	Описание
1	Название	Число объектов, не согласованных с правилами концептуальной схемы
2	Краткое название	—
3	Название элемента	Концептуальная согласованность
4	Базовая мера	Подсчет ошибок
5	Определение	Количество всех объектов в наборе данных, которые не отвечают правилам концептуальной схемы
6	Описание	Если концептуальная схема явно или неявно описывает правила, то эти правила должны быть соблюдены. Нарушением таких правил может быть, например, неверное расположение объектов в рамках заданной допустимой погрешности, дублирование объектов и неверное перекрытие объектов
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	**Пример 1:** Башни с одинаковыми атрибутами и в рамках допустимости поиска (допустимый поиск = 10 м) 

Окончание таблицы D.10

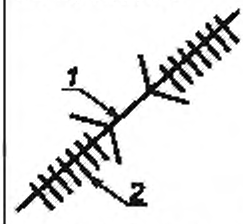
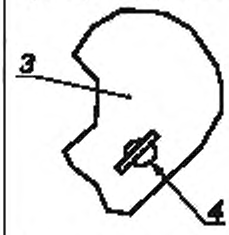
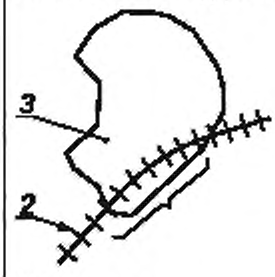
№	Компонент	Описание
11	Пример	**Пример 2:** Мост имеет недопустимое транспортное сообщение. Применение категории дороги  **Пример 3:** Недопустимое расположение аэропорта в озере  **Пример 4:** Недопустимое перекрытие области объекта «озеро» и линейного объекта «железная дорога»  Обозначения: 1 — мост; 3 — озеро; 2 — железная дорога; 4 — аэропорт
12	Идентификатор	10

Таблица D.11 — Число неверных перекрытий поверхностей

№	Компонент	Описание
1	Название	Число неверных перекрытий поверхностей
2	Краткое название	Перекрытие поверхностей
3	Название элемента	Концептуальная согласованность
4	Базовая мера	Подсчет ошибок
5	Определение	Общее количество некорректных перекрытий в данных
6	Описание	Приложение определяет, какие поверхности могут пересекаться, а какие не должны. Не все перекрывающиеся поверхности обязательно являются некорректными. При представлении отчета об этой мере качества данных должно быть указано также, какие типы объектов не должны пересекаться

Окончание таблицы D.11

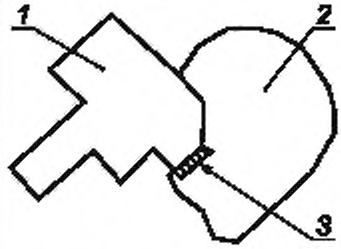
№	Компонент	Описание
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	 Обозначения: 1 — поверхность 1; 2 — поверхность 2; 3 — область перекрытия
12	Идентификатор	11

Таблица D.12 — Уровень несогласованности в отношении правил концептуальной схемы

№	Компонент	Описание
1	Название	Уровень несогласованности в отношении правил концептуальной схемы
2	Краткое название	—
3	Название элемента	Концептуальная согласованность
4	Базовая мера	Частота ошибок
5	Определение	Количество объектов в наборе данных, которые не соответствуют правилам концептуальной схемы по отношению к общему числу этих объектов, которые должны быть в наборе данных
6	Описание	—
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	2 %
12	Идентификатор	12

Таблица D.13 — Уровень согласованности с правилами концептуальной схемы

№	Компонент	Описание
1	Название	Уровень согласованности с правилами концептуальной схемы
2	Краткое название	—
3	Название элемента	Концептуальная согласованность
4	Базовая мера	Уровень корректных объектов

Окончание таблицы D.13

№	Компонент	Описание
5	Определение	Количество объектов в наборе данных в соответствии с правилами концептуальной схемы по отношению к общему числу объектов
6	Описание	—
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	90 %
12	Идентификатор	13

D.3.2 Доменная согласованность

Меры качества данных для подэлемента качества данных «доменная согласованность» представлены в таблицах D.14—D.18.

Таблица D.14 — Несогласованность области значений

№	Компонент	Описание
1	Название	Несогласованность области значений
2	Краткое название	—
3	Название элемента	Доменная согласованность
4	Базовая мера	Индикатор ошибок
5	Определение	Показатель несогласованности объекта с его областью значения
6	Описание	
7	Параметр	—
8	Тип значения	Булева переменная (значение «true» указывает на то, что объект не согласуется с областью значений)
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	14

Таблица D.15 — Согласованность области значений

№	Компонент	Описание
1	Название	Согласованность области значений
2	Краткое название	—
3	Название элемента	Доменная согласованность
4	Базовая мера	Индикатор корректности
5	Определение	Показатель согласованности объекта с областью значений
6	Описание	—
7	Параметр	—

Окончание таблицы D.15

№	Компонент	Описание
8	Тип значения	Булева переменная (значение «true» указывает на то, что объект согласуется с областью значений)
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	15

Таблица D.16 — Количество несогласованных объектов со своей областью значений

№	Компонент	Описание
1	Название	Количество несогласованных объектов со своей областью значений
2	Краткое название	—
3	Название элемента	Доменная согласованность
4	Базовая мера	Подсчет ошибок
5	Определение	Количество всех объектов в наборе данных, которые не согласуются с областью значений
6	Описание	—
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	
10	Указатель источника	—
11	Пример	—
12	Идентификатор	16

Таблица D.17 — Степень согласованности области значений

№	Компонент	Описание
1	Название	Степень согласованности области значений
2	Краткое название	—
3	Название элемента	Доменная согласованность
4	Базовая мера	Частота корректных объектов
5	Определение	Количество объектов в наборе данных, которые соответствуют области значений по отношению к общему количеству объектов в наборе данных
6	Описание	—
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	17

Таблица D.18 — Степень несогласованности области значений

№	Компонент	Описание
1	Название	Степень несогласованности области значений
2	Краткое название	—
3	Название элемента	Доменная согласованность
4	Базовая мера	Частота ошибок
5	Определение	Количество объектов в наборе данных, которые не соответствуют области значений по отношению к общему количеству объектов в наборе данных
6	Описание	—
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	18

D.3.3 Согласованность по формату

Меры качества данных для подэлемента качества данных «согласованность по формату» даны в таблицах D.19—D.21.

Таблица D.19 — Конфликты физической структуры

№	Компонент	Описание
1	Название	Конфликты физической структуры
2	Краткое название	—
3	Название элемента	Согласованность по формату
4	Базовая мера	Индикатор ошибок
5	Определение	Показатель того, что объекты конфликтуют с физической структурой набора данных
6	Описание	—
7	Параметр	—
8	Тип значения	Булева переменная (значение «true» указывает на конфликт физической структуры)
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	True (набор данных хранится в неправильном формате и шейп-файле вместо gml)
12	Идентификатор	19

Таблица D.20 — Количество конфликтов физической структуры

№	Компонент	Описание
1	Название	Количество конфликтов физической структуры
2	Краткое название	—

Окончание таблицы D.20

№	Компонент	Описание
3	Название элемента	Согласованность по формату
4	Базовая мера	Подсчет ошибок
5	Определение	Количество всех объектов, которые хранятся в конфликте с физической структурой набора данных
6	Описание	—
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	5 (применен код 5 для типа «жилое помещение», хотя по требованию в спецификации информационного продукта должно быть не более чем 3)
12	Идентификатор	19

Таблица D.21 — Частота конфликтов физической структуры

№	Компонент	Описание
1	Название	Частота конфликтов физической структуры
2	Краткое название	—
3	Название элемента	Согласованность по формату
4	Базовая мера	Частота ошибок
5	Определение	Количество объектов в наборе данных, которые хранятся в конфликте с физической структурой набора данных, деленное на общее количество элементов
6	Описание	—
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	20

D.3.4 Топологическая согласованность

Меры качества данных в таблицах D.22—D.28 предназначены для проверки топологической согласованности геометрических представлений объектов. Они не могут служить в качестве мер согласованности заданных описаний топологии с использованием топологических объектов, определенных в [7].

Таблица D.22 — Количество некорректных соединений точки и кривой

№	Компонент	Описание
1	Название	Количество некорректных соединений точки и кривой
2	Краткое название	Посторонние узлы
3	Название элемента	Топологическая согласованность
4	Базовая мера	Подсчет ошибок

Окончание таблицы D.22

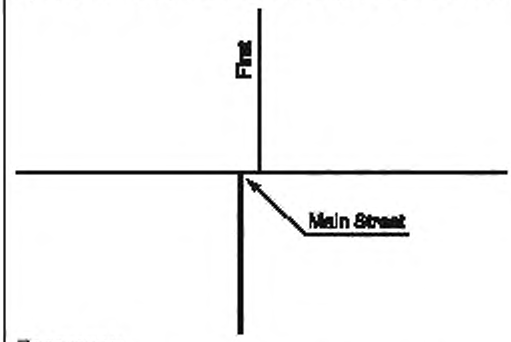
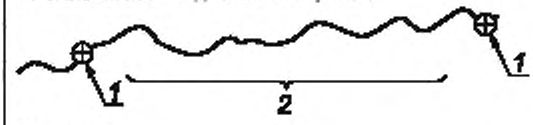
№	Компонент	Описание
5	Определение	Количество некорректных соединений «точка — кривая» в наборе данных
6	Описание	Соединение между точкой и кривой образуется в месте соприкосновения кривых. Эти кривые имеют внутренние топологические отношения, которые должны отражать истинную комбинацию. Если соединение точки и кривой противоречит предметной области, то такое соединение некорректно в отношении этой меры качества данных. Мера качества данных подсчитывает количество ошибок такого рода
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	**Пример 1:** Соединение кривой в двух точках вместо одной.  Пояснение: Перекресток двух дорог должен образовать пересечение «+». **Пример 2:** Система автоматически размещает точки на кривой через определенное число вершин, встроенное в программный код, где нет пространственного согласования для точки и кривой.  Пояснения: 1 — узел соединения; 2 — допустимый установленный предел
12	Идентификатор	21

Таблица D.23 — Степень некорректных соединений точки и кривой

№	Компонент	Описание
1	Название	Степень некорректных соединений точки и кривой
2	Краткое название	—
3	Название элемента	Топологическая согласованность
4	Базовая мера	Частота ошибок
5	Определение	Количество неверных узловых соединений по отношению к числу предполагаемых соединений

Окончание таблицы D.23

№	Компонент	Описание
6	Описание	Соединение между точкой и кривой образуется в месте соприкосновения кривых. Эти кривые имеют внутренние топологические отношения, которые должны отражать истинную комбинацию. Если соединение точки и кривой противоречит предметной области, то такое соединение некорректно в отношении этой меры качества данных. Эта мера качества данных выдает некорректные соединения «точка — кривая» по отношению к общему числу таких соединений
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	22

Таблица D.24 — Количество пропущенных соединений из-за «недохода»

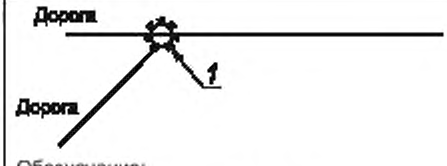
№	Компонент	Описание
1	Название	Количество пропущенных соединений из-за «недохода»
2	Краткое название	«Недоход»
3	Название элемента	Топологическая согласованность
4	Базовая мера	Подсчет ошибок
5	Определение	Количество объектов в наборе данных в рамках допустимой погрешности параметров, не совпавших друг с другом из-за «недохода»
6	Описание	—
7	Параметр	Радиус поиска от конца недоведенной (висячей) линии
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	 Обозначение: 1 — устойчивость поиска = 3 м
12	Идентификатор	23

Таблица D.25 — Количество пропущенных соединений из-за «перехода»

№	Компонент	Описание
1	Название	Количество пропущенных соединений из-за «перехода»
2	Краткое название	«Переход»
3	Название элемента	Топологическая согласованность
4	Базовая мера	Подсчет ошибок

Окончание таблицы D.25

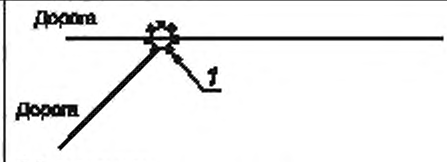
№	Компонент	Описание
5	Определение	Количество объектов в наборе данных, не согласованных между собой из-за перехода за границу в пределах параметра допуска
6	Описание	—
7	Параметр	Поисковый допуск минимальной длины перехода за границу в наборе данных
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	 Обозначение: 1 — устойчивость поиска = 3 м
12	Идентификатор	24

Таблица D.26 — Количество недопустимых «разрывов»

№	Компонент	Описание
1	Название	Количество недопустимых «разрывов»
2	Краткое название	«Разрывы»
3	Название элемента	Топологическая согласованность
4	Базовая мера	Подсчет ошибок
5	Определение	Количество всех объектов в наборе данных, которые представляют недопустимые разрывы поверхностей
6	Описание	Разрыв — это непреднамеренно созданная область, которая возникает из-за неверного цифрового преобразования соседних поверхностей. Между границами смежных поверхностей могут случайно образовываться в небольших количествах «пробелы» или «перекрытия», что приводит к топологической ошибке
7	Параметр	Данная мера качества включает два параметра. Параметр 1 Название: максимальная площадь области разрыва. Описание: максимальная площадь, которая определяет верхний предел разрыва. Это предотвращает от ошибочного распознавания поверхности с извилистыми границами и большой областью в качестве разрыва. Тип значения: Real Параметр 2 Название: показатель толщины. Описание: показатель толщины должен быть действительным числом от 0 до 1. Данный показатель определяется по следующей формуле: T — показатель толщины. T = 4π [площадь]/[периметр]2. T = 1 значение соответствует окружности, которая имеет самое большое значение [площадь]/[периметр]2. T = 0 значение соответствует линии, которая имеет наименьшее значение [площадь]/[периметр]2. Описание: показатель толщины не зависит от размера поверхности, и чем ближе значение к 0, тем тоньше будет выбранная поверхность разрыва. Тип значения: Real

Окончание таблицы D.26

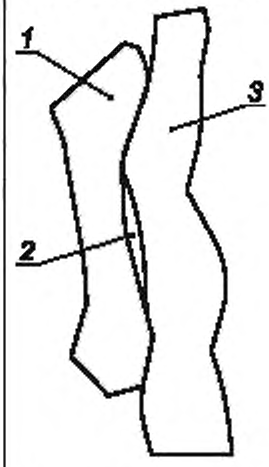
№	Компонент	Описание
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	[11]
11	Пример	 Обозначения: 1 — одинарная дренажная линия; 2 — двойная дренажная линия a) параметр максимальной площади предотвращает верное изображение двойной дренажной линии от пометки ее как ошибки.  Обозначения: 1 — песок; 2 — разрыв; 3 — двойной линейный дренаж b) разрыв меньше, чем максимальный параметр, и проверяется на возможную ошибку
12	Идентификатор	24

Таблица D.27 — Количество некорректных самопересечений

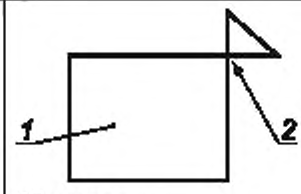
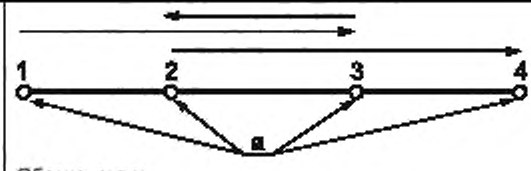
№	Компонент	Описание
1	Название	Количество некорректных самопересечений
2	Краткое название	Петли
3	Название элемента	Топологическая согласованность
4	Базовая мера	Подсчет ошибок
5	Определение	Общее количество некорректных пересечений объектов
6	Описание	—
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	 Обозначения: 1 — здание 1; 2 — некорректное пересечение (петля)
12	Идентификатор	26

Таблица D.28 — Количество некорректных самоперекрытий

№	Компонент	Описание
1	Название	Количество некорректных самоперекрытий
2	Краткое название	«Отдача»
3	Название элемента	Топологическая согласованность
4	Базовая мера	Подсчет ошибок
5	Определение	Общее количество некорректных самопересечений объектов
6	Описание	—
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	 Обозначение: а) вершины
12	Идентификатор	27

D.4 Позиционная точность

D.4.1 Абсолютная или внешняя точность

D.4.1.1 Основные меры для позиционной неопределенности

Основные меры качества данных для позиционной неопределенности для элемента качества «абсолютная/внешняя точность» даны в таблицах D.29—D.34.

Таблица D.29 — Среднее значение позиционной неопределенности

№	Компонент	Описание
1	Название	Среднее значение позиционной неопределенности (1D, 2D и 3D)
2	Краткое название	—
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	—
6	Описание	Среднее значение позиционной неопределенности набора позиций, где позиционные неопределенности определяются как расстояние между измеряемым положением и тем, которое взято в качестве соответствующего истинного положения 1D: $e_i = \left\|x_{mi} - x_{ti}\right\|$; 2D: $e_i = \sqrt{\left(x_{mi} - x_{ti}\right)^2 + \left(y_{mi} - y_{ti}\right)^2}$; 3D: $e_i = \sqrt{\left(x_{mi} - x_{ti}\right)^2 + \left(y_{mi} - y_{ti}\right)^2 + \left(z_{mi} - z_{ti}\right)^2}$. Средняя позиционная неопределенность горизонтальной абсолютной или внешней позиции вычисляется по формуле $\bar{e} = \frac{1}{N}\sum_{i=1}^{N} e_i$. Также следует указывать критерий для установления соответствия (например, позволяющий устанавливать согласованность к ближайшей позиции, на вершинах или вдоль линий). Критерий/критерии поиска соответствующих точек должны представляться вместе с результатом оценки качества данных. Эта мера качества данных отлична от стандартного отклонения
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	28

Таблица D.30 — Смещение позиции

№	Компонент	Описание
1	Название	Смещение позиций (1D, 2D и 3D)
2	Краткое название	—
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется

Окончание таблицы D.30

№	Компонент	Описание
5	Определение	Для множества точек, где расстояние не превышает определенную предельную величину, среднее арифметическое расстояний между их измеренными позициями и той, что взята в качестве истинной позиции
6	Описание	Для числа точек (N) измеренные позиции представлены как координаты x_{mi}, y_{mi} и z_{mi} в зависимости от измерения, в котором измеряется положение точки. Соответствующий набор координат x_{ti}, y_{ti} и z_{ti} представляет истинную позицию. Отклонение и смещение рассчитываются следующим образом: Единичное отклонение: $e_{xi} = x_{mi} - x_{ti}$, $e_{yi} = y_{mi} - y_{ti}$, $e_{zi} = z_{mi} - z_{ti}$. Смещение: $a_x = \frac{\sum e_{xi}}{N_x}$, $a_y = \frac{\sum e_{yi}}{N_y}$, $a_z = \frac{\sum e_{zi}}{N_z}$, $a_p = \sqrt{a_x^2 + a_y^2}$, $a_{3D} = \sqrt{a_x^2 + a_y^2 + a_z^2}$. Критерий для установления согласования также следует указывать (например, с учетом согласования к ближайшей позиции, на узлах или вдоль линий). Критерий/критерии для поиска соответствующих точек должны представляться вместе с результатом оценки качества данных
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	128

Таблица D.31 — Среднее значение позиционной неопределенности без учета выбросов

№	Компонент	Описание
1	Название	Среднее значение позиционной неопределенности без учета выбросов (2D)
2	Краткое название	—
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	Для множества точек, расстояние которых не превышает определенный порог, среднее арифметическое расстояние между их измеренными позициями и той, что взята в качестве истинной позиции

Окончание таблицы D.31

№	Компонент	Описание
6	Описание	Для числа точек (N) измеренные позиции представлены как координаты x_{mi}, y_{mi} и z_{mi} в зависимости от измерения, в котором измеряется положение точки. Соответствующий набор координат x_{ti}, y_{ti} и z_{ti} представляет истинную позицию. Все позиционные неопределенности выше определенного порога e_{max} затем исключаются из набора. Позиционная неопределенность рассчитывается следующим образом: $e'_i = \begin{cases} e_i, & \text{если } e_i \le e_{max} \\ 0, & \text{если } e_i > e_{max} \end{cases}$. Расчет e_i определяется мерой качества данных «среднее значение позиционной неопределенности» в одно-, двух- и трехмерных измерениях. Для оставшегося числа ошибок (N_R) среднее значение горизонтального абсолютного положения вычисляется следующим образом: $\bar{e}$, исключая выбросы, $= \frac{1}{N}\sum_{i=1}^{N} e'_i$. Критерий для установления согласования также следует указывать (например, с учетом согласования к ближайшей позиции, на узлах или вдоль линий). Критерий/критерии для поиска соответствующих точек должны представляться вместе с результатом оценки качества данных
7	Параметр	Название: e_{max}. Определение: порог для принятой позиционной неопределенности. Тип значения: число
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	29

Таблица D.32 — Позиционная неопределенность вышезаданной предельной величины

№	Компонент	Описание
1	Название	Число позиционных неопределенностей вышезаданной предельной величины
2	Краткое название	
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Подсчет ошибок
5	Определение	Число позиционных неопределенностей выше заданного порогового значения для набора позиций. Ошибки определяются как расстояние между измеренным положением и тем, которое заявлено в качестве истинного
6	Описание	Для числа точек (N) измеренные позиции представлены как координаты x_{mi}, y_{mi} и z_{mi} в зависимости от измерения, в котором измеряется положение точки. Соответствующий набор координат x_{ti}, y_{ti} и z_{ti} представляет истинную позицию. Расчет e_i определяется мерой качества данных «среднее значение позиционной неопределенности» в одно-, двух- и трехмерных измерениях. Всякая позиционная неопределенность, превышающая установленный порог e_{max} ($e_i > e_{max}$), считается ошибкой. Критерий для установления согласования также следует указывать (например, с учетом согласования к ближайшей позиции, на узлах или вдоль линий). Критерий/критерии для поиска соответствующих точек должны представляться вместе с результатом оценки качества данных.

Окончание таблицы D.32

№	Компонент	Описание
7	Параметр	Название: e_{max}. Определение: порог для принятой позиционной неопределенности. Тип значения: число
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	30

Таблица D.33 — Уровень позиционных ошибок выше заданного порога

№	Компонент	Описание
1	Название	Уровень позиционных ошибок выше заданного порога
2	Краткое название	—
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	Число позиционных неопределенностей выше заданного порогового значения для набора позиций по отношению к общему количеству измеренных положений. Ошибки определяются как расстояние между измеренным положением и тем, которое заявлено в качестве истинного
6	Описание	Для числа точек (N) измеренные позиции представлены как координаты x_{mi}, y_{mi} и z_{mi} в зависимости от измерения, в котором измеряется положение точки. Соответствующий набор координат x_{ti}, y_{ti} и z_{ti} представляет истинную позицию. Расчет e_i определяется мерой качества данных «среднее значение позиционной неопределенности» в одно-, двух- и трехмерных измерениях. Всякая позиционная неопределенность, превышающая установленный порог e_{max} ($e_i > e_{max}$), считается ошибкой. Количество ошибок установлено в зависимости от общего количества измеренных точек. Критерий для установления согласования также следует указывать (например, с учетом согласования к ближайшей позиции, на узлах или вдоль линий). Критерий/критерии для поиска соответствующих точек должны представляться вместе с результатом оценки качества данных
7	Параметр	Название: e_{max}. Определение: порог, выше которого подсчитывается позиционная неопределенность. Тип значения: число
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	25 % узлов в пределах области определения качества данных имеют погрешность расстояния больше 1 м
12	Идентификатор	31

Таблица D.34 — Ковариационная матрица

№	Компонент	Описание
1	Название	Ковариационная матрица
2	Краткое название	Вариационно-ковариационная матрица
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	Симметричная квадратная матрица с дисперсией координат точек на главной диагонали и ковариацией между этими координатами по недиагональным элементам
6	Описание	Ковариационная матрица генерализует концепцию дисперсии от одномерного до *N*-го измерения, т. е. от скалярных случайных величин до векторных случайных величин (кортежи скалярных случайных величин). (1) 1D-координаты (например, данные высоты) Векторная случайная величина: $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Ковариационная матрица: $\sum_{xx} = \begin{bmatrix} \sigma^2_{x_1} & \cdots & \sigma_{x_1 x_n} \\ \vdots & \ddots & \vdots \\ \sigma_{x_n x_1} & \cdots & \sigma^2_{x_n} \end{bmatrix}$, где $\sigma_{x_1 x_n} = \sigma_{x_n x_1}$. $\sigma^2_{x_1}$ определяет объект дисперсии x_1, ее квадратный корень задает стандартное отклонение этого элемента $\sigma_{x_1} = \sqrt{\sigma^2_{x_1}}$. Корреляция между двумя элементами может быть вычислена по формуле $\rho_{x_i x_j} = \frac{\sigma_{x_i x_j}}{\sigma_{x_i} \sigma_{x_j}}$. Если координаты некоррелированы, то недиагональные объекты имеют значение 0. (2) 2D-координаты Векторная случайная величина: $x = \begin{bmatrix} x_1 \\ y_1 \\ \vdots \\ x_n \\ y_n \end{bmatrix}$. Ковариационная матрица: $\sum_{xx} = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1 y_1} & \cdots & \sigma_{x_1 y_n} \\ \sigma_{y_1 x_1} & \sigma^2_{y_1} & \cdots & \sigma_{y_1 y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{y_n x_1} & \sigma_{y_n y_1} & \cdots & \sigma^2_{y_n} \end{bmatrix}$.

Окончание таблицы D.34

№	Компонент	Описание
6	Описание	(3) 3D-координаты Векторная случайная величина: $X = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ \vdots \\ y_n \\ z_n \end{bmatrix}$. Ковариационная матрица: $\sum_{xx} = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1 y_1} & \sigma_{x_1 z_1} & \cdots & \sigma_{x_1 y_n} & \sigma_{x_1 z_n} \\ \sigma_{x_1 y_1} & \sigma^2_{y_1} & \sigma_{y_1 z_1} & \cdots & \sigma_{y_1 y_n} & \sigma_{y_1 z_n} \\ \sigma_{x_1 z_1} & \sigma_{y_1 z_1} & \sigma^2_{z_1} & \cdots & \sigma_{z_1 y_n} & \sigma_{z_1 z_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{x_1 y_n} & \sigma_{y_1 y_n} & \sigma_{z_1 y_n} & \cdots & \sigma^2_{y_n} & \sigma_{y_n z_n} \\ \sigma_{x_1 z_n} & \sigma_{y_1 y_n} & \sigma_{z_1 z_n} & \cdots & \sigma_{y_n z_n} & \sigma^2_{y_n} \end{bmatrix}$. (4) произвольно выбранные координаты Векторная случайная величина: $X = \begin{bmatrix} a \\ b \\ \vdots \\ z \end{bmatrix}$. Ковариационная матрица: $\Sigma_{xx} = \begin{bmatrix} \sigma\frac{2}{a} & \sigma_{b_a} & \cdots & \sigma_{z_a} \\ \sigma_{a_b} = \sigma_{b_a} & \sigma\frac{2}{b} & \cdots & \sigma_{z_b} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{a_z} = \sigma_{z_a} & \sigma_{b_z} = \sigma_{z_b} & \cdots & \sigma\frac{2}{z} \end{bmatrix}$.
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	Матрица
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	32

D.4.1.2 Вертикальная позиционная неопределенность

Измерения высоты представляют собой позиционные наблюдения в одном измерении, а значит, высоту можно рассматривать как одномерную случайную величину. Поэтому меры качества данных для позиционной неопределенности основаны на базовой мере качества данных «одномерная случайная величина».

Меры качества данных для вертикальной позиционной неопределенности подэлемента качества данных «абсолютная/внешняя точность» представлены в таблицах D.35—D.43.

Таблица D.35 — Линейное вероятное отклонение

№	Компонент	Описание
1	Название	Линейное вероятное отклонение
2	Краткое название	LEP
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	LE50 или LE50(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение находится с вероятностью 50 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	33

Таблица D.36 — Стандартное линейное отклонение

№	Компонент	Описание
1	Название	Стандартное линейное отклонение
2	Краткое название	SLE
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	LE68.3 или LE68.3(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение находится с вероятностью 68,3 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	34

Таблица D.37 — Линейная картографическая точность при 90-процентном уровне значимости

№	Компонент	Описание
1	Название	Линейная картографическая точность при 90-процентном уровне значимости
2	Краткое название	LMAS 90 %
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	LE90 или LE90(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение находится с вероятностью 90 %
6	Описание	См. G.3.2

Окончание таблицы D.37

№	Компонент	Описание
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	35

Таблица D.38 — Линейная картографическая точность при 95-процентном уровне значимости

№	Компонент	Описание
1	Название	Линейная картографическая точность при 95-процентном уровне значимости
2	Краткое название	LMAS 95 %
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	LE95 или LE95(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение находится с вероятностью 95 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	36

Таблица D.39 — Линейная картографическая точность при 99-процентном уровне значимости

№	Компонент	Описание
1	Название	Линейная картографическая точность при 99-процентном уровне значимости
2	Краткое название	LMAS 99 %
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	LE99 или LE99(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение находится с вероятностью 99 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	37

Таблица D.40 — Уровень высокой линейной достоверности

№	Компонент	Описание
1	Название	Уровень высокой линейной достоверности
2	Краткое название	—
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	LE99.8 или LE99.8(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение находится с вероятностью 99,8 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	38

Таблица D.41 — Средняя квадратическая ошибка

№	Компонент	Описание
1	Название	Среднеквадратическая ошибка
2	Краткое название	RMSE
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	—
6	Описание	Истинное значение измеряемой величины Z известно как z_t. Исходя из этого, оценка $\sigma_z = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_{mi} - z_t)^2}$ приводит к среднеквадратической ошибке RMSE = α_z
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	39

Таблица D.42 — Абсолютная линейная ошибка при 90-процентном уровне значимости смещенных вертикальных данных

№	Компонент	Описание
1	Название	Абсолютная линейная ошибка при 90-процентном уровне значимости смещенных вертикальных данных (Вариант 1)
2	Краткое название	LMAS

Окончание таблицы D.42

№	Компонент	Описание
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	Абсолютная вертикальная точность координат данных, выраженная в терминах линейной ошибки при 90-процентной вероятности наличия смещения
6	Описание	Сравнение данных (источник) и контроль (эталон) вычисляется следующим образом: 1. Вычисление абсолютной погрешности в вертикальном измерении в каждой точке: δV_i = источник V_i – эталон V_i при $i = 1 \ldots N$. 2. Вычисление абсолютного значения смещения: $\lvert\overline{\delta V}\rvert = \left\lvert\frac{1}{N}\sum_{i=1}^{N}\delta V_i\right\rvert$. 3. Вычисление линейного стандартного отклонения измеренных различий между тестируемым продуктом и эталонным источником: $\sigma_{dv} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta V_i^2}$. 4. Вычисление линейного стандартного отклонения ошибок в эталонном источнике: σ_R. 5. Вычисление линейного стандартного отклонения ошибок в тестируемом продукте: $\sigma_V = \sqrt{\sigma_M^2 + \sigma_R^2}$. 6. Вычисление соотношения (ratio) абсолютной величины средней ошибки к стандартному отклонению: $\text{ratio} = \frac{\lvert\overline{\delta V}\rvert}{\sigma_V}$. 7. Если ratio > 1,4, то LMAS = σ_V [1,282 + ratio]. 8. Если ratio ≤ 1,4, то LMAS = σ_V [1,6435 + 0,92 · ratio^2 – 0,28 · ratio^3]
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	[12]
11	Пример	—
12	Идентификатор	40

Таблица D.43 — Абсолютная линейная ошибка при 90-процентном уровне значимости смещенных вертикальных данных

№	Компонент	Описание
1	Название	Абсолютная линейная ошибка при 90-процентном уровне значимости смещенных вертикальных данных (Вариант 2)
2	Краткое название	ALE
3	Название элемента	Абсолютная или внешняя точность

Окончание таблицы D.43

№	Компонент	Описание
4	Базовая мера	Не применяется
5	Определение	Абсолютная вертикальная точность координат данных, выраженная в терминах линейной ошибки при 90-процентной вероятности наличия смещения
6	Описание	Сравнение данных (источника) с эталоном вычисляется следующим образом: 1. Вычисление абсолютной погрешности в вертикальном измерении в каждой точке: δV_i = источник V_i – эталон V_i при $i = 1 \ldots N$. 2. Вычисление средней ошибки по вертикали: $\lvert\overline{\delta V}\rvert = \left\lvert\frac{1}{N}\sum_{i=1}^{N}\delta V_i\right\rvert$. 3. Вычисление стандартного отклонения ошибок по вертикали: $\sigma_{ce} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta V_i^2}$. **4. Вычисление соотношения (ratio) абсолютной величины средней ошибки к** стандартному отклонению: ratio = $\lvert\overline{\delta V}\rvert / \sigma_V$. 5. Если ratio > 1,4, тогда k = 1,2815. 6. Если ratio ≤ 1,4, тогда вычисление k проводится на основе соотношения вертикального смещения к стандартному отклонению высот с помощью кубического полинома в соответствии с табличными значениями, определенными в Справочнике таблиц по теории вероятностей и статистики (Handbook of Tables for Probability and Statistics) [12]: $k = 1{,}6435 - (0{,}999556 \cdot \text{ratio}) + (0{,}923237 \cdot \text{ratio}^2) - (0{,}282533 \cdot \text{ratio}^3)$. 7. Вычисление LE90 для источника: $\text{LE90}_{\text{источник}} = \lvert\overline{\delta V}\rvert + (k \cdot \sigma_V)$. 8. Вычисление абсолютного LE90: $\text{LE90}_{\text{abc}} = \sqrt{\text{LE90}^2_{\text{эталон}} + \text{LE90}^2_{\text{источник}}}$
7	Параметр	Название: объем выборки. Определение: используется, как правило, минимум 30 точек, что не всегда возможно в зависимости от идентифицируемых контрольных точек. Для 10-процентной выборки уровня атрибуции объектов от генеральной совокупности объектов. Тип значения: Real
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	[13] [4] [12]
11	Пример	—
12	Идентификатор	41

D.4.1.3 Горизонтальная позиционная неопределенность

Горизонтальные координаты точки определяются 2D-координатами. Неопределенность координаты любой точки можно описать с помощью базовых мер качества данных для двухмерной случайной величины, описанной в G.3.3. Меры качества данных для горизонтальной позиционной неопределенности элемента качества данных «абсолютная/внешняя точность» представлены в таблицах D.44—D.53.

Таблица D.44 — Стандартное круговое отклонение

№	Компонент	Описание
1	Название	Стандартное круговое отклонение
2	Краткое название	Стандартная круговая ошибка, точечная ошибка по Гельмерту, CSE
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	CE39.4
5	Определение	Радиус описывающей окружности, в которой истинное положение точек расположено с вероятностью 39,4 %
6	Описание	См. G.3.3
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	42

Таблица D.45 — Вероятная круговая ошибка

№	Компонент	Описание
1	Название	Вероятная круговая ошибка
2	Краткое название	CEP
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	CE50
5	Определение	Радиус описывающей окружности, в которой истинное положение точек расположено с вероятностью 50 %
6	Описание	См. G.3.3
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	43

Таблица D.46 — Стандарт круговой картографической точности

№	Компонент	Описание
1	Название	Круговая ошибка при 90-процентном уровне значимости
2	Краткое название	Стандарт круговой картографической точности (CMAS)
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	CE90
5	Определение	Радиус описывающей окружности, в которой истинное положение точек расположено с вероятностью 90 %

Окончание таблицы D.46

№	Компонент	Описание
6	Описание	См. G.3.3
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	44

Таблица D.47 — Круговая ошибка при 95-процентном уровне значимости

№	Компонент	Описание
1	Название	Круговая ошибка при 95-процентном уровне значимости
2	Краткое название	Точность навигации
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	CE95
5	Определение	Радиус описывающей окружности, в которой истинное положение точек расположено с вероятностью 95 %
6	Описание	См. G.3.3
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	45

Таблица D.48 — Почти достоверная круговая ошибка

№	Компонент	Описание
1	Название	Почти достоверная круговая ошибка
2	Краткое название	CNCE
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	CE99.8
5	Определение	Радиус описывающей окружности, в которой истинное положение точек расположено с вероятностью 99,8 %
6	Описание	См. G.3.3
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	46

Таблица D.49 — Среднеквадратическая ошибка планиметрии

№	Компонент	Описание
1	Название	Среднеквадратическая ошибка планиметрии
2	Краткое название	RMSEP
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	Радиус окружности вокруг заданной точки, в которой истинное значение расположено с вероятностью P
6	Описание	Истинные значения наблюдаемых координат X и Y известны как x_t и y_t. Исходя из этого, статистическая оценка $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\left(x_{mi} - x_t\right)^2 + \left(y_{mi} - y_t\right)^2\right]}$ приводит к среднеквадратической ошибке планиметрии RMSEP = σ
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	47

Таблица D.50 — Абсолютная круговая ошибка при 90-процентном уровне значимости смещенных данных

№	Компонент	Описание
1	Название	Абсолютная круговая ошибка при 90-процентном уровне значимости смещенных данных (Вариант 2)
2	Краткое название	Мера абсолютной горизонтальной точности при 90-процентном уровне значимости смещенных данных (CMAS)
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	Абсолютная горизонтальная точность координат данных, выраженная в терминах круговой ошибки при вероятности 90 %, учитывая наличие смещения
6	Описание	Сравнение данных (источник) и контроль (эталон) вычисляется следующим образом: 1. Вычисление абсолютной погрешности в горизонтальной плоскости в каждой точке и каждой координате X_i и Y_i: δX_i = (источник X_i – эталон X_i) и δY_i = (источник Y_i – эталон Y_i) при i = 1 … N. 2.Вычисление средней горизонтальной ошибки каждой координаты: $\overline{\delta X} = \frac{1}{N}\sum_{1}^{N}\delta X_i$ и $\overline{\delta Y} = \frac{1}{N}\sum_{1}^{N}\delta Y_i$.

Окончание таблицы D.50

№	Компонент	Описание
6	Описание	3. Вычисление линейного стандартного отклонения измеренных различий между тестируемым продуктом и эталонным источником: $\sigma_{CM} = \sqrt{\frac{1}{2(N-1)}\left\{\sum_{i=1}^{N}\left(\delta X_i - \overline{\delta X}\right)^2 + \sum_{i=1}^{N}\left(\delta X_i - \overline{\delta X}\right)^2\right\}}$. 4. Вычисление линейного стандартного отклонения ошибок в эталонном источнике: σ_{CR}. 5. Вычисление линейного стандартного отклонения ошибок в тестируемом продукте: $\sigma_C = \sqrt{\sigma_{CM}^2 + \sigma_{CR}^2}$. 6. Вычисление абсолютной круговой ошибки при 90-процентном уровне достоверности смещенных данных (CMAS): $\mathrm{CMAS} = \sigma_C \cdot \left[1{,}2943 + \sqrt{\left(\frac{\overline{\delta X}^2 + \overline{\delta Y}^2}{\sigma_C}\right)} \cdot 0{,}7254\right]$
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	[12]
11	Пример	—
12	Идентификатор	48

Таблица D.51 — Абсолютная круговая ошибка при 90-процентном уровне значимости неравномерно распределенных данных

№	Компонент	Описание
1	Название	Абсолютная круговая ошибка при 90-процентном уровне значимости смещенных данных
2	Краткое название	ACE
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	Абсолютная горизонтальная точность координат данных, выраженная в терминах круговой ошибки при вероятности 90 %, учитывая наличие смещения
6	Описание	Сравнение данных (источник) с эталоном вычисляется следующим образом: 1. Вычисление абсолютной погрешности в горизонтальной плоскости в каждой точке: $\Delta H_i = \sqrt{(\text{источник } X_i - \text{эталон } X_i)^2 + (\text{источник } Y_i - \text{эталон } Y_i)^2}$ при $i = 1 \ldots N$. 2. Вычисление средней горизонтальной ошибки: $\mu_H = \left(\sum \Delta H_i\right) / N$.

Окончание таблицы D.51

№	Компонент	Описание
6	Описание	3. Вычисление стандартного отклонения горизонтальных ошибок: $\sigma_H = \sqrt{\sum (\Delta H_i - \mu_H)^2 / (N-1)}$. 4. Вычисление соотношения (ratio) абсолютной величины средней ошибки к стандартному отклонению: ratio = $\lvert\mu_H\rvert / \sigma_H$. 5. Если ratio > 1,4, то k = 1,2815. 6. Если ratio ≤ 1,4, то вычисление k проводится на основе соотношения вертикального смещения к стандартному отклонению высот с помощью кубического полинома в соответствии с табличными значениями, определенными в CRC (Справочнике таблиц по теории вероятностей и статистики) [12]. $k = 1{,}6435 - (0{,}999556 \cdot \text{ratio}) + (0{,}923237 \cdot \text{ratio}^2) - (0{,}282533 \cdot \text{ratio}^3)$. 7. Вычисление CE90 для источника: $CE90_{source} = \lvert\mu_H\rvert + (k \cdot \sigma_H)$. 8. Вычисление абсолютного CE90: $CE90_{abs} = \sqrt{CE90^2_{эталон} + CE90^2_{источник}}$
7	Параметр	Название: объем выборки. Определение: используется, как правило, минимум 30 точек, что не всегда возможно в зависимости от идентифицируемых контрольных точек. Для 10-процентной выборки уровня атрибуции объектов от генеральной совокупности объектов. Тип значения: Real
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	[13] [4]
11	Пример	
12	Идентификатор	49

Таблица D.52 — Эллипс неопределенности

№	Компонент	Описание
1	Название	Эллипс неопределенности
2	Краткое название	Стандартный эллипс ошибок точек
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	2D-эллипс с двумя основными осями с указанием направления и значения самой высокой и самой низкой неопределенности 2D-точки
6	Описание	Из заданной ковариационной матрицы (меры качества данных, таблица D.34) 2D-координат точек элементы, описывающие эллипс неопределенности, могут быть определены на основе их собственных значений. Для одной точки k ковариационная матрица имеет вид $\Sigma^k_{xx} = \begin{bmatrix} \sigma^2_{x_k} & \sigma_{x_k y_k} \\ \sigma_{y_k x_k} & \sigma^2_{y_k} \end{bmatrix}$, где $\sigma_{x_k y_k} = \sigma_{y_k x_k}$.

Окончание таблицы D.52

№	Компонент	Описание
6	Описание	Направление α (bearing) основной полуоси эллипса неопределенности может быть вычислено по формуле $\varnothing = \frac{1}{2}\arctan\frac{2\sigma_{x_k y_k}}{\sigma^2_{x_k} - \sigma^2_{y_k}}$ и $a = \sqrt{\frac{1}{2}\chi^2_{1-\alpha}(2)\left(\sigma^2_{x_k} + \sigma^2_{y_k} + \sqrt{\left(\sigma^2_{x_k} - \sigma^2_{y_k}\right)^2 + 4\sigma^2_{x_k y_k}}\right)}$ $b = \sqrt{\frac{1}{2}\chi^2_{1-\alpha}(2)\left(\sigma^2_{x_k} - \sigma^2_{y_k} - \sqrt{\left(\sigma^2_{x_k} - \sigma^2_{y_k}\right)^2 + 4\sigma^2_{x_k y_k}}\right)}$
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	Последовательность (a, b, φ)
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	50

Таблица D.53 — Доверительный эллипс

№	Компонент	Описание
1	Название	Эллипс достоверности
2	Краткое название	Доверительный эллипс ошибок точек
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	2D-эллипс с двумя основными осями с указанием направления и значения самой высокой и самой низкой неопределенности 2D-точки
6	Описание	Из заданной ковариационной матрицы (меры качества данных, таблица D.34) элементы, описывающие эллипс неопределенности, могут быть определены на основе их собственных значений. Для одной точки k ковариационная матрица имеет вид $\Sigma^k_{xx} = \begin{bmatrix} \sigma^2_{x_k} & \sigma_{x_k y_k} \\ \sigma_{y_k x_k} & \sigma^2_{y_k} \end{bmatrix}$ для $\sigma_{x_k y_k} = \sigma_{y_k x_k}$. Направление α (bearing) основной полуоси эллипса неопределенности может быть вычислено по формуле $\varnothing = \frac{1}{2}\arctan\frac{2\sigma_{x_k y_k}}{\sigma^2_{x_k} - \sigma^2_{y_k}}$ и $a = \sqrt{\frac{1}{2}\chi^2_{1-\alpha}(2)\left(\sigma^2_{x_k} + \sigma^2_{y_k} + \sqrt{\left(\sigma^2_{x_k} - \sigma^2_{y_k}\right)^2 + 4\sigma^2_{x_k y_k}}\right)}$ $b = \sqrt{\frac{1}{2}\chi^2_{1-\alpha}(2)\left(\sigma^2_{x_k} - \sigma^2_{y_k} - \sqrt{\left(\sigma^2_{x_k} - \sigma^2_{y_k}\right)^2 + 4\sigma^2_{x_k y_k}}\right)}$.

Окончание таблицы D.53

№	Компонент	Описание
6	Описание	Со значениями для $\chi^2_{1-\alpha}(2)$-распределения 2D доверительного эллипса $\chi^2_{1-\alpha}(2)$ $P = 1-\alpha = 95\%$ 5,99 $P = 1-\alpha = 99\%$ 9,21
7	Параметр	Название: значительный уровень. Определение: $1-\alpha$. Тип значения: число
8	Тип значения	Мера
9	Структура значения	Последовательность (a, b, φ)
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	51

D.4.1.4 Относительная или внутренняя точность

Данный элемент качества предусматривает использование тех же мер качества данных, что и элемент «абсолютная или внешняя точность». Разница заключается лишь в методе оценки.

Относительная точность между объектами может быть выражена с помощью мер качества данных «относительная вертикальная ошибка» и «относительная горизонтальная ошибка», которые представлены в таблицах D.54 и D.55.

Таблица D.54 — Относительная вертикальная ошибка

№	Компонент	Описание
1	Название	Относительная вертикальная ошибка
2	Краткое название	Rel LE90
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	Оценка случайных ошибок между объектами рельефа в одном наборе данных или карте/графике. Это функция случайной ошибки в двух возвышениях по отношению к общей высотной (вертикальной) дате
6	Описание	Сравнение данных (источник) с эталоном вычисляется следующим образом: 1. Определение всех возможных комбинаций пар точек. Комбинации Пар точек = $m = n(n1)/2$. 2. Вычисление абсолютной вертикальной ошибки в каждой точке: ΔZ_i = Измеренная высота$_i$ – Эталонная высота$_i$ при $i = 1 \ldots n$. 3. Вычисление относительной вертикальной ошибки для всех комбинаций пар точек: $\Delta Z_{\text{rel } kj} = \Delta Z_k \quad \Delta Z_j$ при $k = 1 \ldots m1$, $j = k + 1 \ldots m$. 4. Вычисление относительного вертикального стандартного отклонения $\sigma_{Z\,\text{rel}} = \sqrt{\frac{\sum Z^2_{\text{rel}}}{m-1}}$. 5. Вычисление относительной LE путем приведения сигмы к 90-процентному уровню значимости: Rel LE90 =1,645$\sigma_{Z\,\text{rel}}$

Окончание таблицы D.54

№	Компонент	Описание
7	Параметр	Название: *n*. Определение: объем выборки. Тип значения: Integer
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	[13]
11	Пример	—
12	Идентификатор	52

Таблица D.55 — Относительная горизонтальная ошибка

№	Компонент	Описание
1	Название	Относительная горизонтальная ошибка
2	Краткое название	Rel CE90
3	Название элемента	Абсолютная или внешняя точность
4	Базовая мера	Не применяется
5	Определение	Оценка случайных ошибок в горизонтальной позиции между объектами в наборе данных или карте/графике
6	Описание	Сравнение данных (источник) с эталоном вычисляется следующим образом: 1. Определение всех возможных комбинаций пар точек: Комбинации пар точек = $m = n(n1)/2$. 2. Вычисление абсолютной погрешности в X и Y плоскостях в каждой точке: ΔX_i = Измеренное X_i – Эталонное X_i при $i = 1 \ldots n$, ΔY_j = Измеренное Y_j – Эталонное Y_j при $i = 1 \ldots n$. 3. Вычисление относительной погрешности в X и Y плоскостях для всех комбинаций пар точки: $\Delta X_{\text{rel } kj} = \Delta X_k - \Delta X_j$ при $k = 1 \ldots m1$, $j = k + 1 \ldots m$, $\Delta Y_{\text{rel } kj} = \Delta Y_k - \Delta Y_j$ при k = $1 \ldots m1$, $j = k + 1 \ldots m$. 4. Вычисление относительного стандартного отклонения на каждой оси: $\sigma_{X\ \text{rel}} = \sqrt{\frac{\sum \Delta X^2_{\text{rel}}}{m-1}}$, $\sigma_{Y\ \text{rel}} = \sqrt{\frac{\sum \Delta Y^2_{\text{rel}}}{m-1}}$. 5. Вычисление относительного горизонтального стандартного отклонения: $\sigma_{H\ \text{rel}} = \sqrt{\frac{\sigma^2_{X_{\text{rel}}} + \sigma^2_{Y_{\text{rel}}}}{2}}$. 6. Вычисление относительной LE путем приведения сигмы к 90-процентному уровню значимости: Rel CE90 = $2{,}146\sigma_{H\ \text{rel}}$
7	Параметр	Название: *n*. Определение: объем выборки. Тип значения: Integer

Окончание таблицы D.55

№	Компонент	Описание
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	[13]
11	Пример	—
12	Идентификатор	53

D.4.2 Точность позиционирования данных относительно координатной сетки

Для описания сеточных данных могут использоваться те же меры качества, что и для горизонтальной позициональной неопределенности в соответствии с D.4.1.3. Для описания групповых значений в растрах может применяться точность количественных атрибутов в соответствии с D.6.3.

D.5 Точность временных измерений

Временные измерения можно рассматривать как одномерные случайные величины. Использование базовых мер качества данных согласно G.3.2 приводит к мерам качества данных в соответствии с таблицами D.56—D.61.

Таблица D.56 — Временная точность при 68,3-процентном уровне значимости

№	Компонент	Описание
1	Название	Временная точность при 68,3-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность измерения времени
4	Базовая мера	LE68.3 или LE68.3(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение экземпляра времени лежит с вероятностью 68,3 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	54

Таблица D.57 — Временная точность при 50-процентном уровне значимости

№	Компонент	Описание
1	Название	Временная точность при 50-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность измерения времени
4	Базовая мера	LE50 или LE50(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение экземпляра времени лежит с вероятностью 50 %
6	Описание	См. G.3.2
7	Параметр	—

Окончание таблицы D.57

№	Компонент	Описание
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	55

Таблица D.58 — Временная точность при 90-процентном уровне значимости

№	Компонент	Описание
1	Название	Временная точность при 90-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность измерения времени
4	Базовая мера	LE90 или LE90(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение экземпляра времени лежит с вероятностью 90 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	
11	Пример	—
12	Идентификатор	56

Таблица D.59 — Временная точность при 95-процентном уровне значимости

№	Компонент	Описание
1	Название	Временная точность при 95-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность измерения времени
4	Базовая мера	LE95 или LE95(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение экземпляра времени лежит с вероятностью 95 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	
12	Идентификатор	57

Таблица D.60 — Временная точность при 99-процентном уровне значимости

№	Компонент	Описание
1	Название	Временная точность при 99-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность измерения времени
4	Базовая мера	LE99 или LE99(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение экземпляра времени лежит с вероятностью 99 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	58

Таблица D.61 — Временная точность при 99,8-процентном уровне значимости

№	Компонент	Описание
1	Название	Временная точность при 99,8-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность измерения времени
4	Базовая мера	LE99.8 или LE99.8(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение экземпляра времени лежит с вероятностью 99,8 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	59

D.5.1 Временная согласованность

В таблице D.62 дана одна мера качества данных для подэлемента качества «временная согласованность».

Таблица D.62 — Хронологическая последовательность

№	Компонент	Описание
1	Название	Хронологическая последовательность
2	Краткое название	—
3	Название элемента	Временная согласованность
4	Базовая мера	Индикатор ошибок

Окончание таблицы D.62

№	Компонент	Описание
5	Определение	Признак того, что событие неверно упорядочено в сравнении с другими событиями
6	Описание	—
7	Параметр	—
8	Тип значения	Булева переменная (значение «true» указывает на то, что событие упорядочено неверно)
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	True (в наборе данных присутствуют 5 исторических событий, но они упорядочены неправильно)
12	Идентификатор	159

D.5.2 Временное соответствие

Для описания временного соответствия можно использовать те же меры качества данных, что и для допустимых значений атрибутов (см. меры качества данных в таблицах D.14—D.18 подэлемента качества «доменнная согласованность»).

D.6 Тематическая точность

D.6.1 Корректность классификации

Отнесение элемента к определенному классу может быть правильным либо неправильным. В таблицах D.63—D.67 дано несколько мер качества данных в зависимости от классифицируемого элемента.

Таблица D.63 — Количество некорректно классифицированных объектов

№	Компонент	Описание
1	Название	Количество некорректно классифицированных объектов
2	Краткое название	—
3	Название элемента	Корректность классификации
4	Базовая мера	Подсчет ошибок
5	Определение	Количество неправильно классифицированных объектов
6	Описание	—
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	60

Таблица D.64 — Степень некорректности классификации

№	Компонент	Описание
1	Название	Степень некорректности классификации
2	Краткое название	—
3	Название элемента	Корректность классификации

Окончание таблицы D.64

№	Компонент	Описание
4	Базовая мера	Частота ошибок
5	Определение	Количество неправильно классифицированных объектов по отношению к количеству объектов, которые должны присутствовать в классификации
6	Описание	—
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	61

Таблица D.65 — Матрица неправильной классификации

№	Компонент	Описание
1	Название	Матрица неправильной классификации
2	Краткое название	Матрица неточностей
3	Название элемента	Корректность классификации
4	Базовая мера	—
5	Определение	Матрица, которая указывает количество объектов класса (i), отнесенных к классу (j)
6	Описание	Матрица неправильной классификации (МСМ) — это квадратичная матрица со столбцами n и строками n. n обозначает число рассматриваемых классов. МСМ (i, j) = [# объекты класса (i), отнесенные к классу (j)]. Диагональные объекты матрицы неправильной классификации содержат правильно классифицированные объекты, а недиагональные объекты содержат количество ошибок неправильной классификации
7	Параметр	Название: n. Определение: количество рассматриваемых классов. Тип значения: Integer
8	Тип значения	Integer
9	Структура значения	Матрица ($n \times n$)
10	Ссылка на источник	—
11	Пример	
12	Идентификатор	62

		Dataset class			
True class		A	B	C	Count
	A	7	2	1	10
	B	1	2	2	5
	C	1	1	3	5
	Count	9	5	6	20

Таблица D.66 — Относительная матрица неправильной классификации

№	Компонент	Описание
1	Название	Относительная матрица неправильной классификации
2	Краткое название	—
3	Название элемента	Корректность классификации
4	Базовая мера	—
5	Определение	Матрица, которая указывает количество объектов класса (i), отнесенных к классу (j) и деленных на количество объектов класса (i)
6	Описание	Относительная матрица неправильной классификации (RMCM) — это квадратичная матрица с n столбцами и n строками. n обозначает число рассматриваемых классов RMCM (i, j) = [# объекты класса (i), отнесенные к классу (j)] / (# объекты класса (i)] · 100 %
7	Параметр	Название: n. Определение: количество рассматриваемых классов. Тип значения: Integer
8	Тип значения	Real
9	Структура значения	Матрица ($n \times n$)
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	63

Таблица D.67 — Коэффициент каппа

№	Компонент	Описание
1	Название	Коэффициент каппа
2	Краткое название	
3	Название элемента	Корректность классификации
4	Базовая мера	—
5	Определение	Коэффициент для определения соотношения согласия присвоения классов путем удаления некорректной классификации
6	Описание	С объектами матрицы неправильной классификации матрицы MCM (I, J), заданной в качестве меры качества данных в таблице D.65, коэффициент каппа (K) может быть вычислен по формуле $$K = \frac{N \cdot \sum_{i=1}^{r} MCM(i,i) - \sum_{i=1}^{r}\left(\sum_{j=1}^{r} MCM(i,j) \cdot \sum_{j=1}^{r} MCM(j,i)\right)}{N^2 - \sum_{i=1}^{r}\left(\sum_{j=1}^{r} MCM(i,j) \cdot \sum_{j=1}^{r} MCM(j,i)\right)}$$ N — это количество классифицированных объектов
7	Параметр	Название: n. Определение: количество рассматриваемых классов. Тип значения: Integer
8	Тип значения	Real
9	Структура значения	—

Окончание таблицы D.67

№	Компонент	Описание
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	64

D.6.2 Корректность неколичественных атрибутов

Меры качества данных для подэлемента качества «корректность неколичественных атрибутов» даны в таблицах D.68—D.70.

Таблица D.68 — Число некорректных значений атрибутов

№	Компонент	Описание
1	Название	Число некорректных значений
2	Краткое название	—
3	Название элементов	Корректность неколичественных атрибутов
4	Базовая мера	Подсчет ошибок
5	Определение	Общее количество ошибочных значений атрибутов в соответствующей части набора данных
6	Описание	Количество всех значений атрибутов с неверным значением
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	5 (5 географических названий написаны с ошибками)
12	Идентификатор	65

Таблица D.69 — Степень корректных значений атрибутов

№	Компонент	Описание
1	Название	Степень корректных значений атрибутов
2	Краткое название	—
3	Название элемента	Корректность неколичественных атрибутов
4	Базовая мера	Степень корректных объектов
5	Определение	Количество корректных значений атрибутов по отношению к общему числу значений атрибутов
6	Описание	—
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	66

Таблица D.70 — Степень некорректных значений атрибутов

№	Компонент	Описание
1	Название	Степень некорректных значений атрибутов
2	Краткое название	—
3	Название элемента	Корректность неколичественных атрибутов
4	Базовая мера	Частота ошибок
5	Определение	Количество значений атрибутов с некорректными значениями по отношению к общему числу значений атрибутов
6	Описание	
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	67

D.6.3 Точность количественных атрибутов

Меры качества данных для подэлемента качества «точность количественных атрибутов» представлены в таблицах D.71—D.76.

Таблица D.71 — Неопределенность значения атрибута при 68,3-процентном уровне значимости

№	Компонент	Описание
1	Название	Неопределенность значения атрибута при 68,3-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность количественных атрибутов
4	Базовая мера	LE68.3 или LE68.3(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала определяется верхним и нижним пределом, в котором истинное значение для количественного атрибута лежит с вероятностью 68,3 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	68

Таблица D.72 — Неопределенность значения атрибута при 50-процентном уровне значимости

№	Компонент	Описание
1	Название	Неопределенность значения атрибута при 50-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность количественных атрибутов

Окончание таблицы D.72

№	Компонент	Описание
4	Базовая мера	LE50 или LE50(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала определяется верхним и нижним пределом, в котором истинное значение для количественного атрибута лежит с вероятностью 50 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	69

Таблица D.73 — Неопределенность значения атрибута при 90-процентном уровне значимости

№	Компонент	Описание
1	Название	Неопределенность значения атрибута при 90-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность количественных атрибутов
4	Базовая мера	LE90 или LE90(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала определяется верхним и нижним пределом, в котором истинное значение для количественного атрибута лежит с вероятностью 90 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	
12	Идентификатор	70

Таблица D.74 — Неопределенность значения атрибута при 95-процентном уровне значимости

№	Компонент	Описание
1	Название	Неопределенность значения атрибута при 95-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность количественных атрибутов
4	Базовая мера	LE95 или LE95(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала определяется верхним и нижним пределом, в котором истинное значение для количественного атрибута лежит с вероятностью 95 %
6	Описание	См. G.3.2
7	Параметр	

Окончание таблицы D.74

№	Компонент	Описание
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	71

Таблица D.75 — Неопределенность значения атрибута при 99-процентном уровне значимости

№	Компонент	Описание
1	Название	Неопределенность значения атрибута при 99-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность количественных атрибутов
4	Базовая мера	LE99 или LE99(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение для количественного атрибута лежит с вероятностью 99 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	72

Таблица D.76 — Неопределенность значения атрибута при 99,8-процентном уровне значимости

№	Компонент	Описание
1	Название	Неопределенность значения атрибута при 99,8-процентном уровне значимости
2	Краткое название	—
3	Название элемента	Точность количественных атрибутов
4	Базовая мера	LE99.8 или LE99.8(r) в зависимости от процедуры оценки
5	Определение	Половина длины интервала, определяемого верхним и нижним пределом, в котором истинное значение для количественного атрибута лежит с вероятностью 99,8 %
6	Описание	См. G.3.2
7	Параметр	—
8	Тип значения	Мера
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	73

D.7 Меры обобщения

В спецификации информационного продукта устанавливают некоторые требования к продукту для его соответствия данной спецификации. В таблицах D.77—D.81 для этого элемента приведены меры качества данных.

Таблица D.77 — Соответствие спецификации информационного продукта

№	Компонент	Описание
1	Название	Соответствие спецификации информационного продукта
2	Краткое название	—
3	Название элемента	Элемент применимости (юзабилити)
4	Базовая мера	Индикатор корректности
5	Определение	Признак того, что все требования, упомянутые в спецификации информационного продукта выполнены
6	Описание	—
7	Параметр	—
8	Тип значения	Булева переменная («true», если все требования, упомянутые в спецификации информационного продукта, выполнены)
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	101

Таблица D.78 — Количество несоответствий спецификации информационного продукта

№	Компонент	Описание
1	Название	Количество несоответствий спецификации информационного продукта
2	Краткое название	—
3	Название элемента	Элемент применимости (юзабилити)
4	Базовая мера	Подсчет ошибок
5	Определение	Число требований спецификации информационного продукта, которые не выполняются текущим продуктом/набором данных
6	Описание	
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	102

Таблица D.79 — Количество соответствий спецификации информационного продукта

№	Компонент	Описание
1	Название	Количество соответствий спецификации информационного продукта
2	Краткое название	—
3	Название элемента	Элемент применимости (юзабилити)

Окончание таблицы D.79

№	Компонент	Описание
4	Базовая мера	Подсчет корректных элементов
5	Определение	Число требований спецификации информационного продукта, которые выполняются текущим продуктом/набором данных
6	Описание	—
7	Параметр	—
8	Тип значения	Integer
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	103

Таблица D.80 — Степень несоответствия спецификации информационного продукта

№	Компонент	Описание
1	Название	Степень несоответствия спецификации информационного продукта
2	Краткое название	
3	Название элемента	Элемент практичности (юзабилити)
4	Базовая мера	Частота ошибок
5	Определение	Число требований спецификации информационного продукта, которые не выполняются текущим продуктом/набором данных по отношению к общему числу требований спецификации информационного продукта
6	Описание	—
7	Параметр	—
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	104

Таблица D.81 — Степень соответствия спецификации информационного продукта

№	Компонент	Описание
1	Название	Степень соответствия спецификации информационного продукта
2	Краткое название	—
3	Название элемента	Элемент практичности (юзабилити)
4	Базовое название	Частота корректных элементов
5	Определение	Число требований спецификации информационного продукта, которые выполняются текущим продуктом/набором данных по отношению к общему числу требований спецификации информационного продукта
6	Описание	—
7	Параметр	—

Окончание таблицы D.81

№	Компонент	Описание
8	Тип значения	Real
9	Структура значения	—
10	Ссылка на источник	—
11	Пример	—
12	Идентификатор	105

Приложение Е
(справочное)

Оценка и выдача информации о качестве данных

Е.1 Введение

В данном приложении приведен основной пример оценки и выдачи информации о качестве данных.

В разделе D.5 приложения D приведены дополнительные примеры выдачи информации в виде метаданных иного характера, таких как описательный результат, метакачество и оценка выборки.

Е. 2 Описание набора данных

Е.2.1 Спецификация информационного продукта

Е.2.1.1 Общие положения

Для оценки качества данных применяют меры качества.

Спецификация информационного продукта, приведенная ниже, описывает область применения. Спецификация определяет те объекты, атрибуты и отношения, которые считаются важными и должны содержаться в наборе данных.

Примечание — Полное описание спецификации информационного продукта см. ГОСТ Р 57657.

Данный продукт включает транспортную сеть (пути и дороги), здания (жилые и промышленные) и деревья.

Е.2.1.2 Типы объектов

Каждый тип объекта, с пустым или непустым набором атрибутов, указан в таблице Е.1. После каждого имени атрибута следует тип значения (строка или целочисленная переменная), а также область допустимых значений.

Таблица Е.1 — Типы объектов

	Тип объекта	Наименование атрибута	Тип значения	Область допустимых значений
Здания	Промышленное здание			
	Дом	Фамилия	String	
		Количество жильцов	Integer	
Транспортная сеть	Тропа			
	Дорога	Условие	String	С твердым покрытием, грунтовая
	Дерево	Высота	String	A: от 1 до 3 м; B: от 3 до 5 м; C: от 5 до 10 м; D: выше 10 м

Е.2.1.3 Правила

Типы объектов в таблице Е.1 должны придерживаться следующих правил:

- деревья высотой менее 1 м не должны учитываться;
- атрибут «Условие» дороги может не иметь значения («неопределенное значение»);
- атрибуты «Имя» и «Количество жильцов» в доме могут не иметь значения («неопределенное значение»).

Е.2.1.4 Требования к качеству

Общее требование к качеству должно быть совместимым с требованиями к качеству данных, набор данных должен отвечать всем нижеописанным требованиям к качеству данных.

а) В наборе данных могут присутствовать только те типы объектов и атрибуты, которые определены в данной спецификации к информационному продукту.

Транспортная сеть:

б) Максимально два объекта могут быть пропущены для каждого типа объекта.

в) Максимум два объекта могут быть в избытке для каждого типа объекта.

г) Максимальное число экземпляров объектов, которые могут быть неправильно классифицированы в качестве другого типа объекта транспортной сети, — два; в качестве других типов объектов — ноль.

Здания:

д) Максимально два объекта могут быть пропущены для каждого типа объекта.

е) Максимальное число экземпляров объектов, которые могут быть неправильно классифицированы в качестве остальных типов объектов зданий, — два; в качестве других типов объектов — ноль.

Деревья:

ж) Максимально пропущено 10 %.

з) Максимально избыточны 10 %.

и) Максимальный процент деревьев, которые могут иметь неправильную высоту, — 20 %.

к) Процент экземпляров, которые могут быть неправильно классифицированы в качестве других типов объектов, — 0.

Е.2.2 Представление реального мира, предметной области и набора данных

Отношение между тремя понятиями выглядит следующим образом:

- рисунок Е.1 представляет «реальный мир», который обычно содержит больше объектов, чем будет содержаться в наборе данных;
- рисунок Е.2 представляет «предметную область», определенную спецификацией данных; это та часть реального мира, которая должна быть включена в набор данных, если набор данных создан полностью и точно;
- рисунок Е.3 представляет набор данных как он есть.

На всех рисунках:

- цифра или буква, обозначающая диапазон цифр под символом дерева, — это высота дерева в метрах,
- цифра внутри символа дома — это количество жильцов дома;
- имена обитателей дома расположены рядом с символом дома.

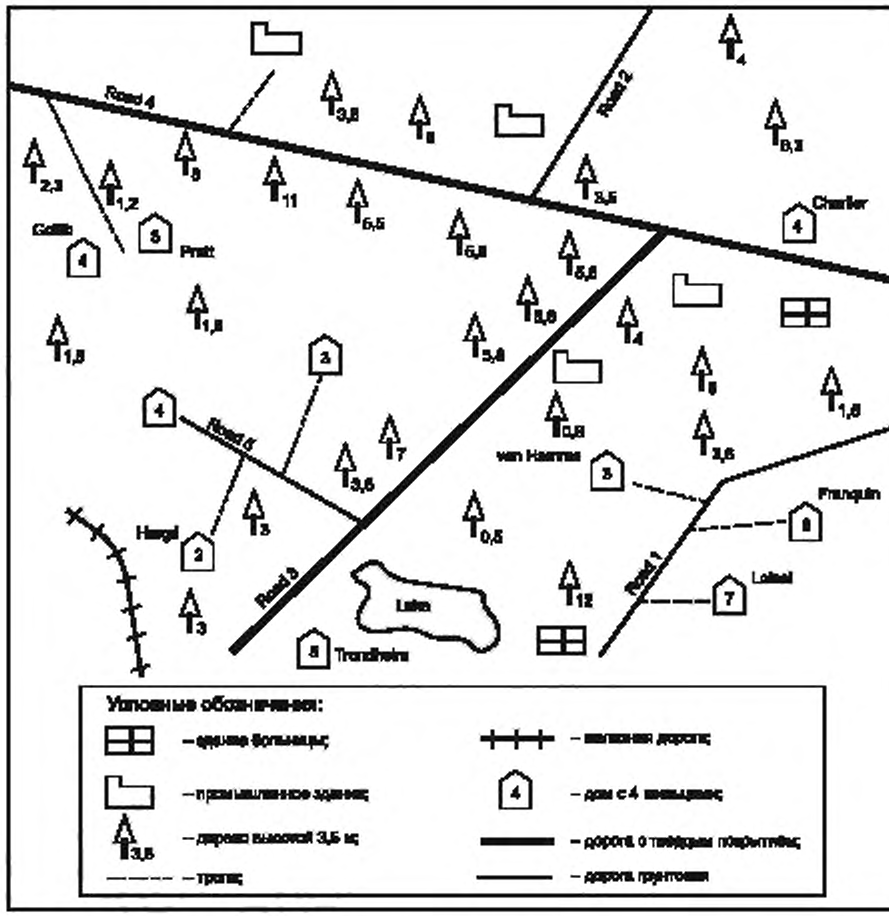


Рисунок Е.1 — Графическое представление «реального мира»

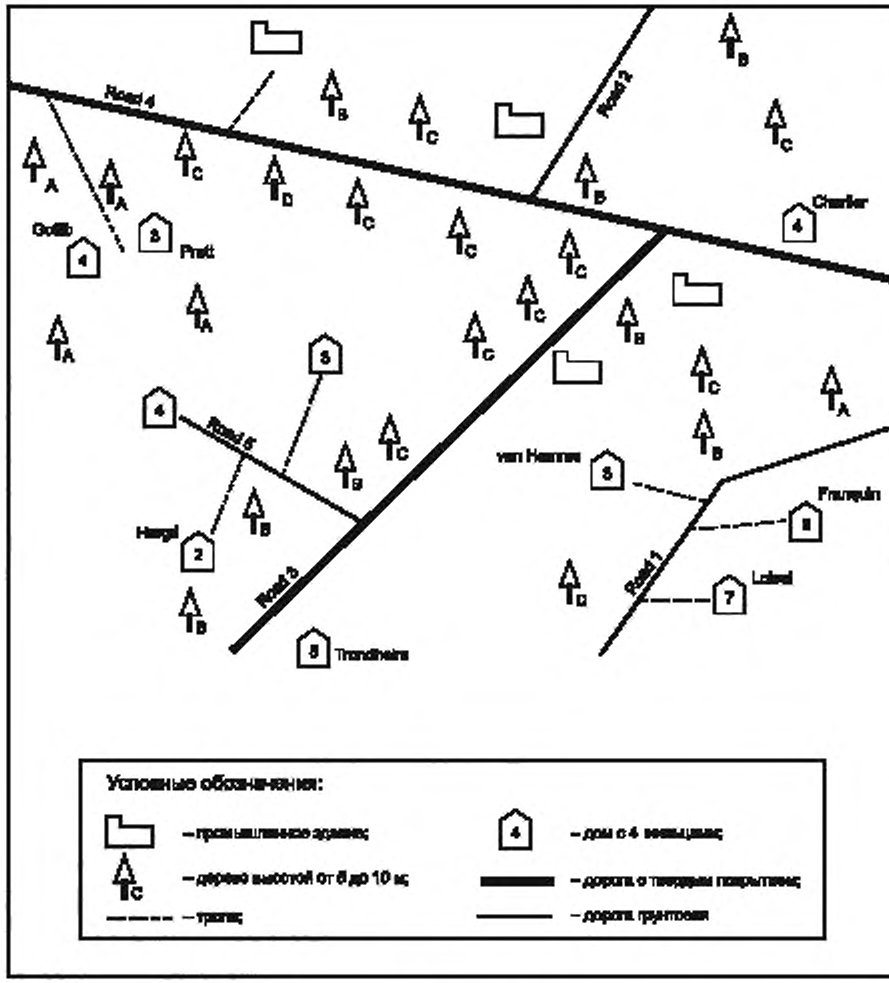


Рисунок Е.2 — Графическое представление предметной области

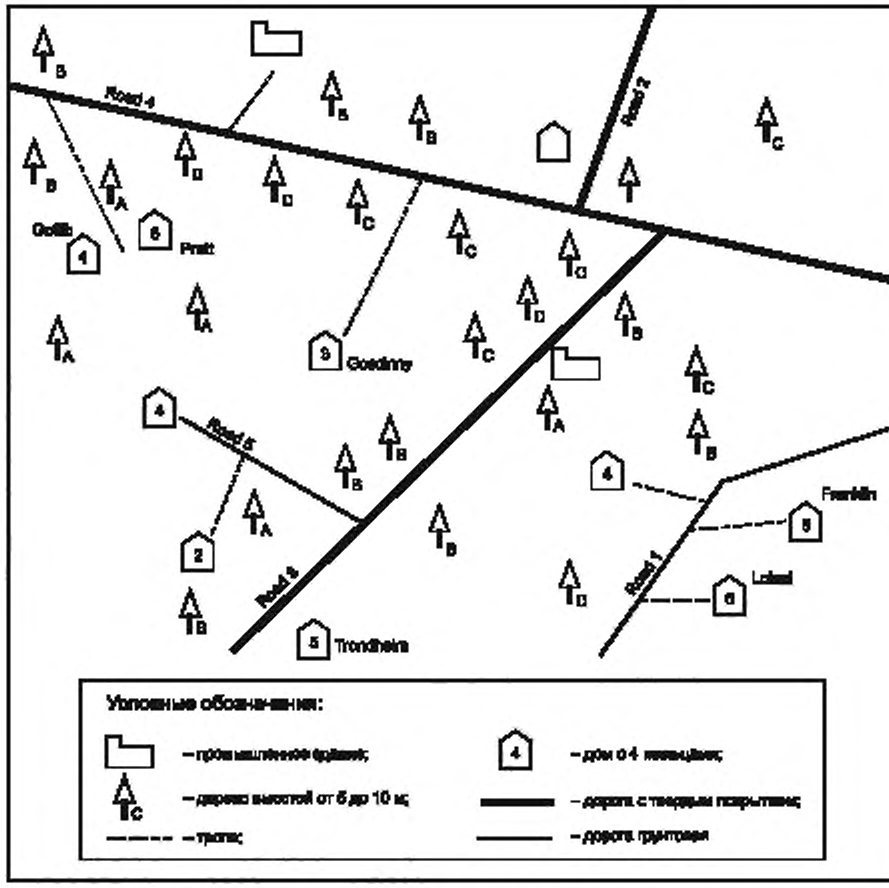


Рисунок Е.3 — Графическое представление набора данных

Е.3 Процесс оценки качества

Е.3.1 Определение единиц(ы) качества данных

Единица качества данных включает область определения и элемент(ы) качества. В этом примере полнота и тематическая точность оцениваются на соответствие спецификации информационного продукта:

- первая единица качества включает концептуальную согласованность, полноту (присутствие и отсутствие) и точность тематической классификации, оценка которых проводится во всем наборе данных;
- две другие единицы качества включают обобщенную концептуальную согласованность, полноту (присутствие или отсутствие) и точность тематической классификации по транспортным сетям и зданиям;
- одна единица качества включает точность количественного атрибута, оцениваемую по типу объекта (дерево);
- последняя единица качества включает элементы применимости (общее соответствие требованиям спецификации информационного продукта), оцениваемые по всему набору данных.

Рекомендации по выбору подходящих элементов качества данных приведены в приложении I.

Е.3.2 Определение мер качества данных

Меры, которые используются в этом примере, взяты из списка зарегистрированных мер из приложения D.

Для описания логической согласованности используют следующие меры:

- Мера 9, «соответствие концептуальной схемы».

Для описания полноты:

- Мера 1, «избыточный элемент»;
- Мера 2, «число избыточных элементов»;
- Мера 3, «показатель избыточных элементов»;

- Мера 5, «отсутствующий элемент»;
- Мера 6, «число отсутствующих элементов»;
- Мера 7, «показатель отсутствующих элементов».

Для описания тематической точности используют:

- Мера 62, «матрица ошибок классификации».

Для описания применимости используют:

- Мера 101, «принятая спецификация информационного продукта».

Е.3.3 Определение процедур оценки качества данных

В этом примере используется процедура прямой внешней оценки.

В данном примере осуществляется полная проверка.

Примечание — Пример процедуры выборочного контроля дан в Е.5.4.

Е.3.4 Определение оценки качества данных на выходе (Результат)

Е.3.4.1 Выявление ошибок

Сравнивая набор данных (рисунок Е.3) с предметной областью (рисунок Е.2), можно получить перечень ошибок в данном примере набора данных (рисунок Е.4).

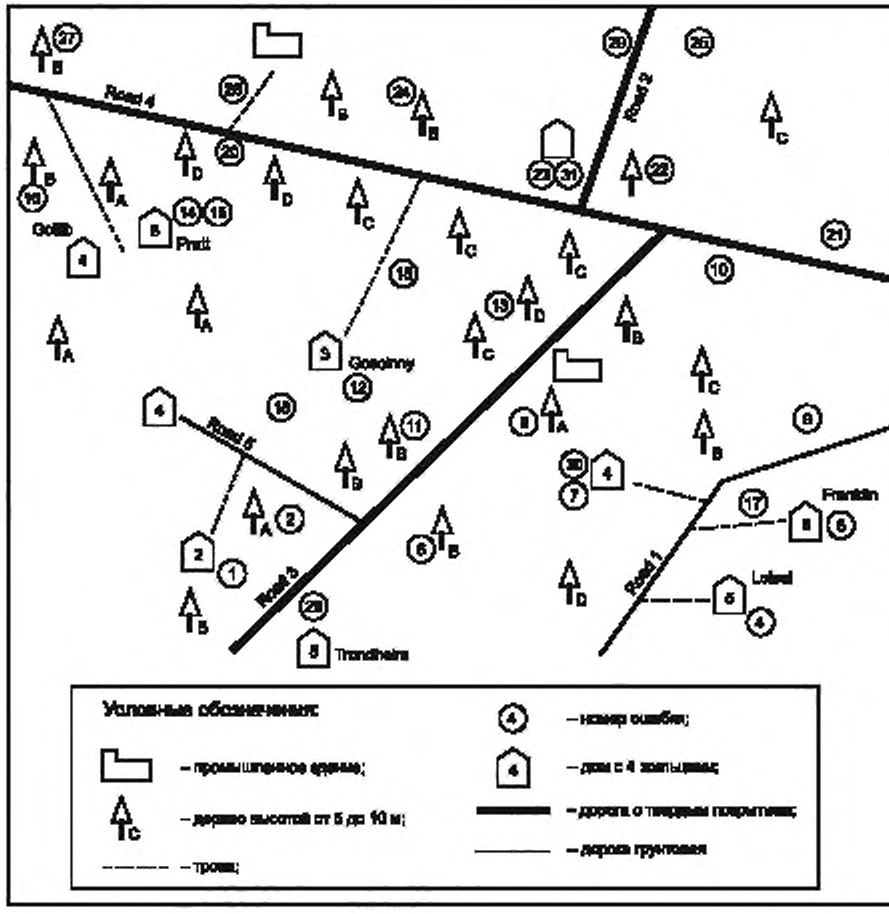


Рисунок Е.4 — Графическое представление местоположения ошибок в наборе данных

Ниже приведен список обнаруженных ошибок с цифрами ошибок, указанных для сравнения.

- Ошибки присутствия или отсутствия при регистрации деревьев. Три дерева (№ 6, № 8, № 27) — в избытке, и два дерева отсутствуют (№ 9, № 25).
- Ошибки присутствия или отсутствия при регистрации троп. Одна тропа отсутствует (№ 18) и одна находится в избытке (№ 19).

- Дом заменяет промышленное здание (№ 23).
- Две тропы ошибочно обозначены в виде дороги (№ 17, № 26).
- Отсутствует дом (№ 21).
- Ошибка атрибута дорог. Две дороги имеют неправильное «состояние» (№ 29, № 28).
- Два дерева высотой менее 1 м представлены в наборе данных (№ 6, № 8).
- Отсутствует код класса атрибута высоты дерева. У дерева отсутствует код класса, в то время как в предметной области стоит В (№ 22).
- Атрибут высоты дерева неправильно классифицирован. У шести деревьев указан неправильный класс высоты.
- Ошибка атрибута наименования дома «family name» («фамилия»). Дома «van Hamme» (№ 7) и «Hergé» (№ 1) в предметной области не имеют имени в наборе данных. Дом под названием «Goscinny» в наборе данных (№ 12) не имеет наименования в предметной области.
- Ошибка атрибута наименования дома «family name» («фамилия»). Дома «Franquin» (№ 5) и «Pratt» (№ 15) в предметной области идут под именами «Franklin» и «Prat» соответственно в наборе данных.
- Ошибка атрибута числа жильцов. В одном доме отсутствует атрибут количества жильцов (№ 31), а в трех домах оно указано неверно (№ 4, № 14, № 30).
- Ошибка типа «пропуск» промышленного здания. Одно промышленное здание отсутствует (№ 10).

Примечание — Классификация ошибок по отсутствию/присутствию, полноте и тематической точности выполнена субъективно.

Например, неправильная классификация дома в качестве промышленного здания также может рассматриваться как ошибка отсутствия одного элемента и присутствие другого.

E.3.4.2 Логическая согласованность

В наборе данных присутствуют только те типы объектов и атрибуты, которые определены в спецификации к информационному продукту.

Результат соответствия концептуальной согласованности см. в таблице E.2.

Таблица E.2 — Результат соответствия логической согласованности

Область определения	Элемент качества	Требования к качеству данных	Число оценок	Значение «да/нет»	Прохождение
Набор данных	Концептуальная согласованность	В наборе данных могут присутствовать только типы объектов и атрибуты, определенные в схеме приложения	1 (ошибок не обнаружено)	1/0	Да

E.3.4.3 Полнота

E.3.4.3.1 Общие положения

В этом примере полнота классифицируется в соответствии с классом объекта. Типы мер проверены на присутствие и отсутствие. Результаты представлены в таблицах E.3—E.5.

E.3.4.3.2 Количественный результат

Таблица E.3 изображает способ классификации полноты посредством количественных значений.

Таблица E.3 — Полнота в соответствии с классом объекта

Класс объекта	Число экземпляров в предметной области	Количество избыточных элементов	Процентное отношение избыточных элементов[a]	Количество отсутствующих элементов	Процентное отношение отсутствующих элементов[b]
Тропа	7	1	14	3	43
Дорога	5	2	40	0	0
Дерево	25	3	12	2	8
Промышленное здание	4	0	0	2	50
Дом	10	1	10	1	10

[a] Процент присутствия = количество присутствующих элементов/количество элементов в предметной области · 100.

[b] Процент отсутствия (пропусков) = количество отсутствующих элементов/количество элементов в предметной области · 100.

Е.3.4.3.3 Производный результат соответствия

В таблице Е.4 представлены результаты соответствия, выведенные из количественных результатов.

Таблица Е.4 — Соответствие полноты

Идентификация оценки	Элемент качества	Мера и идентификация меры	Тип объекта	Необходимое количество	AQL	Ошибки	Совокупность	Прохождение
1	Присутствие	Избыточный элемент (1)	Тропа	3	2	1	7	Да
2	Отсутствие	Отсутствующий элемент (5)	Тропа	2	2	3	7	Нет
3	Присутствие	Избыточный элемент (1)	Дорога	3	2	2	5	Да
4	Отсутствие	Отсутствующий элемент (5)	Дорога	2	2	0	5	Да
5	Присутствие	Избыточный элемент (1)	Дерево	9	10 %	3	25	Нет
6	Отсутствие	Отсутствующий элемент (5)	Дерево	8	10 %	2	25	Да
7	Присутствие	Избыточный элемент (1)	Промышленное здание	6	2	0	4	Да
8	Отсутствие	Отсутствующий элемент (5)	Промышленное здание	5	2	2	4	Да
9	Присутствие	Избыточный элемент (1)	Дом	6	2	1	10	Да
10	Отсутствие	Отсутствующий элемент (5)	Дом	5	2	1	10	Да

Е.3.4.3.4 Обобщенный результат соответствия

Результаты соответствия в отношении транспортных сетей (троп и дорог) и зданий (промышленных и жилых) объединены в таблице Е.5 с использованием следующего правила: если один из исходных результатов «непригоден», то обобщенный результат будет «непригоден» (однозначная оценка пригодности/непригодности представлена в приложении J).

Таблица Е.5 — Обобщенное соответствие полноты

Область определения	Элемент качества	Требования к качеству данных	Количество оценок и идентификация (см. таблицу Е.4)	Значение «да/нет»	Прохождение
Транспортная сеть	Отсутствие	Максимальное отсутствие для каждого типа объекта — два	2 (оценка № 2 и 4)	1/1	Нет
Транспортная сеть	Присутствие	Максимальная избыточность для каждого типа объекта — два	2 (оценка № 1 и 3)	2/0	Нет
Здания	Отсутствие	Максимальное отсутствие для каждого типа объекта — два	2 (оценка № 8 и 10)	2/0	Да
Здания	Присутствие	Максимальная избыточность для каждого типа объекта — два	2 (оценка № 7 и 9)	2/0	Да

Е.3.4.4 Тематическая точность — правильность классификации

Е.3.4.4.1 Общие положения

Информация о тематической точности может внести дальнейшую ясность в информацию о полноте. Например, две из трех пропущенных троп на самом деле классифицируются как дороги (см. таблицу Е.6). Результаты представлены в таблицах Е.6—Е.8.

Е.3.4.4.2 Количественный результат

Одним из способов отображения ошибок, связанных с тематической точностью, является использование меры «матрица неправильной классификации».

В таблице Е.6 представлена матрица неправильной классификации, показывающая ошибки по классам объектов. Она показывает, насколько верно классифицируются экземпляры в наборе данных. Различные проценты должны всегда относиться к генеральной совокупности в наборе данных.

Примечание — Матрица неправильной классификации представляет собой квадратную матрицу, где (i, j) элемент соответствует количеству экземпляров, отнесенных к классу j, когда на самом деле они принадлежат классу i.

Таблица Е.6 — Матрица неправильной классификации объектов

Предметная область	Набор данных					
	Тропа	Дорога	Дерево	Промышленное здание	Дом	Итого
Тропа	4	2	0	0	0	6
Дорога	0	5	0	0	0	5
Дерево	0	0	23	0	0	23
Промышленное здание	0	0	0	2	1	3
Дом	0	0	0	0	9	9
Итого	4	7	23	2	10	46

Расхождение между суммой и числом элементов в предметной области и наборе данных связано с отсутствием и избытком элементов.

Е.3.4.4.3 Производный результат соответствия

В таблице Е.7 представлены результаты соответствия, полученные из количественных результатов.

Таблица Е.7 — Соответствие тематической точности

Идентификация оценки	Элемент качества	Мера	Тип объекта	Необходимое количество	AQL	Число неверных классификаций	Прохождение
11	Корректность тематической классификации	Количество некорректно классифицированных объектов	Тропа	4	2	2	Да
12	Корректность тематической классификации	Количество некорректно классифицированных объектов	Дорога	4	2	0	Да
13	Корректность тематической классификации	Количество некорректно классифицированных объектов	Промышленное здание	7	2	1	Да
14	Корректность тематической классификации	Количество некорректно классифицированных объектов	Дом	7	2	0	Да
15	Корректность тематической классификации	Количество некорректно классифицированных объектов	Дерево	11	0	0	Да

Е.3.4.4.4 Агрегированный результат соответствия

Результаты соответствия в отношении транспортных сетей (троп и дорог) и зданий (промышленных и жилых) объединены в таблице Е.8 на основе следующего метода: если один из исходных результатов имеет значение со-

ответствия «непригоден», то обобщенный результат будет «непригоден» (однозначная оценка пригодности/непригодности представлена в приложении J).

Таблица Е.8 — Обобщенное соответствие правильности классификации

Область определения	Элемент качества	Требования к качеству данных	Количество оценок и идентификация (см. таблицу Е.7)	Значение «да/нет»	Прохождение
Транспортная сеть	Корректность тематической классификации	Максимум два экземпляра объекта неправильно классифицированы по каждому типу объекта в виде другого типа объекта Транспортной сети	2 (оценка № 11 и 12)	2/0	Да
Здания	Корректность тематической классификации	Максимум два экземпляра объекта неправильно классифицированы по каждому типу объекта в виде другого типа объекта Здания	2 (оценка № 13 и 14)	2/0	Да

Е.3.4.5 Тематическая точность — точность количественных атрибутов

Е.3.4.5.1 Общие положения

В этом примере проверяется точность количественных атрибутов. В таблице Е.9 учитываются только те объекты, которые имеют общего предка в том же типе объекта («класс»). Результаты представлены в таблицах Е.9 и Е.10.

Е.3.4.5.2 Количественный результат

Атрибут «высота деревьев» представлен в таблице Е.9.

Таблица Е.9 — Матрица неверной классификации атрибута высоты объектов «высота деревьев»

Предметная область	Набор данных				
	Класс А 1—3 м	Класс В 3—5 м	Класс С 5—10 м	Класс D > 10 м	Итого
Класс А	3	1	0	0	4
Класс В	1	5	0	0	6
Класс С	0	2	6	2	10
Класс D	0	0	0	2	2
Итого	4	8	6	4	22

У одного дерева отсутствует код класса, и, следовательно, оно не учитывается в матрице ошибок классификации. Данная ошибка может происходить из-за ошибки доменной согласованности.

Е.3.4.5.3 Производный результат соответствия

В таблице Е.10 представлен результат соответствия, выведенный из количественных результатов.

Таблица Е.10 — Соответствие тематической точности

Элемент качества	Мера и идентификация меры	Тип/атрибут объекта	Необходимое количество	AQL	Число неверных классификаций	Совокупность	Прохождение
Точность количественного атрибута	Матрица неверной классификации (62)	Класс дерево/высота	10	20 %	6	22	Нет

Е.3.4.6 Применимость — обобщенное соответствие спецификации информационного продукта

В таблице Е.11 все результаты соответствия зданий, транспортных сетей и деревьев объединены вместе в соответствии с концептуальной схемой, чтобы обеспечить соответствие спецификации информационного продукта после зарегистрированной меры «принятая спецификация информационного продукта», идентификатор 101 (см. таблицу D.77).

Таблица Е.11 — Применимость — соответствие спецификации продукта

Область определения	Элемент качества	Требования к качеству данных	Число оценок	Значение «да/нет»	Совместимость
Набор данных	Элемент применимости	Общее требование к качеству данных: набор данных должен отвечать всем требованиям к качеству данных в схеме приложения, чтобы быть признанным соответствующим требованиям к качеству данных	11 требований	8/3 (Не прошел требования 2, 9 и 10)	Набор данных не соответствует

Е.4 Выдача информации о качестве данных

Е.4.1 Выдача информации в виде метаданных

Е.4.1.1 Общие положения

В разделах Е.4.1.2—Е.4.1.4 приведены примеры того, как выдавать информацию о качестве в виде метаданных согласно настоящему стандарту (раздел 10 и приложение С) и ИСО 19115-1:2013. Так, один экземпляр MD_Metadata включает один или более экземпляров DQ_DataQuality.

В данных примерах некоторые экземпляры классов (DQ_Quality и DQ_Elements) получили идентификатор (ID) в соответствии с принципами XML. Эти идентификаторы используются при ссылке к таким экземплярам в других классах.

Е.4.1.2 Создание отчета о присутствии

В таблице Е.12 показан пример выдачи информации о количественных результатах, производном результате соответствия и агрегированном результате соответствия для типов объектов транспортной сети.

Механизм вывода этих результатов похож на тот, что применяется для других типов объектов из набора данных.

Таблица Е.12 — Создание отчета о присутствии в виде метаданных

XML-элемент	Пример	Комментарий
DQ_DataQuality		
scope: MD_Scope		
level: MD_ScopeCode	Набор данных	Область определения этой единицы качества данных
standaloneQualityReport: DQ_StandaloneQualityREportInformation		
reportReference: CI_Citation		Ссылка и описание к самостоятельному отчету по качеству
title: CharacterString	Выдача информации в виде самостоятельного отчета по оценке качества (см. Е.4.2)	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
abstract: CharacterString	Самостоятельный отчет по качеству, который прилагается к этой оценке качества, представляет более подробную информацию о методах деривации и агрегирования	
report: DQ_Commission *id = quantitative_commission*		В этом экземпляре избыточности количественный результат представлен для каждого типа объектов меры 2 (число избыточных объектов)

Продолжение таблицы Е.12

XML-элемент	Пример	Комментарий
measure: DQ_MeasureReference		
nameOfMeasure: CharacterString	Количество избыточных элементов	
measureIdentification: MD_Identifier		
code: CharacterString	2	
measureDescription: CharacterString	Количество элементов в наборе данных, которые не должны были в нем находиться	
evaluation: DQ_FullInspection		
evaluationMethodType: DQ_EvaluationMethodTypeCode	directExternal	
evaluationMethodDescription: CharacterString	Сравнение количества элементов в наборе данных с числом элементов в предметной области	
result: DQ_QuantitativeResult		Для повышения читаемости здесь представляется только присутствие троп и дорог, но если областью определения качества данных является набор данных, то о каждом типе объекта должна быть представлена информация
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Тропа	
value: Record	0	
valueUnit: UnitOfMeasure	Нет	
result: DQ_QuantitativeResult		
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Дорога	
value: Record	2	
valueUnit: UnitOfMeasure	Нет	
report: DQ_Commission *id = conformance_commission*		В этом экземпляре избыточности производный результат соответствия предусмотрен для каждого типа объекта для меры 1 (избыточный элемент)
measure: DQ_MeasureReference		
nameOfMeasure: CharacterString	Избыточный элемент	
measureIdentification: MD_Identifier		
code: CharacterString	1	

Продолжение таблицы Е.12

XML-элемент	Пример	Комментарий
measureDescription: CharacterString	Показатель того, что элемент неправильно присутствует в данных	
evaluation: DQ_AggregationDerivation		
evaluationMethodType: DQ_EvaluationMethodTypeCode	Прямой	
evaluationMethodDescription: CharacterString	Производная от количественного результата	
derivedElement: DQ_Element	*quantitative_commission*	Ссылка на исходные результаты
result: DQ_ConformanceResult		Производный результат соответствия для избыточности троп. Для улучшения читаемости здесь представлена только избыточность троп и дорог, но если областью определения качества данных является набор данных, то о каждом типе объекта должна быть представлена информация
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Тропа	
specification: CI_Citation		
title: CharacterString	Спецификация информационного продукта (см. Е.2.1) требование 2	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
pass: Boolean	True	
result: DQ_ConformanceResult		Производный результат соответствия для избыточности дорог. Для улучшения читаемости здесь представлена только избыточность троп и дорог, но если областью определения качества данных является набор данных, то о каждом типе объекта должна быть представлена информация
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Дорога	
specification: CI_Citation		
title: CharacterString	Спецификация информационного продукта (см. Е.2.1), требование 2	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
pass: Boolean	true	
DQ_DataQuality *id = agg_commission1*		Обобщенный результат соответствия для транспортной сети

Окончание таблицы Е.12

XML-элемент	Пример	Комментарий
scope: MD_Scope		Областью определения теперь является тип объекта транспортной сети => изменилась единица качества данных. Поэтому был создан новый экземпляр DQ_DataQuality
level: MD_ScopeCode	FeatureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	TransportNetwork (дорога и тропа)	
report: DQ_Commission		
evaluation: DQ_AggregationDerivation		Метод агрегирования
evaluationMethodType: DQ_EvaluationMethodTypeCode	Косвенный	
evaluationMethodDescription: CharacterString	100-процентное агрегирование значений pass fail результата соответствия по присутствию дорог и троп	
evaluationProcedure: CI_Citation		
title: CharacterString	Приложение J	
Date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
derivedElement: DQ_Element	*conformance_commission*	Ссылка на исходные результаты
result: DQ_ConformanceResult		
specification: CI_Citation		
title: CharacterString	Спецификация информационного продукта (см. Е.2.1), требование 2	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
Pass: Boolean	true	

Е.4.1.3 Создание отчета о правильности классификации

В таблице Е.13 приведен пример выдачи информации о производных результатах соответствия и обобщенных результатов соответствия для типа объекта здания.

Механизм вывода этих результатов похож на тот, что применяется для других типов объектов набора данных.

Таблица Е.13 — Создание отчета о правильности классификации в виде метаданных

XML-элемент	Пример	Комментарий
DQ_DataQuality		
scope: MD_Scope		
level: MD_ScopeCode	Набор данных	Область определения этой единицы качества данных
standaloneQualityReport: DQ_StandaloneQualityReportInformation		

Продолжение таблицы Е.13

XML-элемент	Пример	Комментарий
reportReference: CI_Citation		Ссылка и описание к самостоятельному отчету по качеству
title: CharacterString	Выдача информации в виде самостоятельного отчета по оценке качества (см. Е.4.2)	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
abstract: CharacterString	Самостоятельный отчет по качеству, который прилагается к этой оценке качества, представляет более подробную информацию о методах деривации и агрегирования	
report: DQ_ThematicClassificationCorrectness *id = conformance_classification*		В этом экземпляре корректности классификации производный результат соответствия представлен для каждого типа объектов меры 60 (количество некорректно классифицированных объектов)
measure: DQ_MeasureReference		
nameOfMeasure: CharacterString	Количество некорректно классифицируемых объектов	
measureIdentification: MD_Identifier		
code: CharacterString	60	
evaluation: DQ_AggregationDerivation		
evaluationMethodType: DQ_EvaluationMethodTypeCode	Косвенный	
evaluationMethodDescription: CharacterString	Производная от количественных результатов, представленных в самостоятельном отчете о качестве	
standaloneQualityReportDetails: CharacterString	Исходные количественные результаты описаны в Е.3.4.4.2 самостоятельного отчета о качестве	Ссылка на исходные результаты
result: DQ_ConformanceResult		Производный результат соответствия для классификации промышленных зданий. Исходный количественный результат намеренно не предусмотрен в метаданных. Он описан в самостоятельном отчете по качеству. Атрибут standaloneQualityReportDetails дает точную ссылку на исходный результат в самостоятельном отчете по качеству
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Промышленное здание	
specification: CI_Citation		
title: CharacterString	Спецификация информационного продукта (см. Е.2.1), требование 7	

Продолжение таблицы Е.13

XML-элемент	Пример	Комментарий
date: CI_Date		Для улучшения читаемости здесь выдается информация только о присутствии промышленных зданий и домов, но если областью определения качества данных является набор данных, то о каждом типе объекта должна быть представлена информация
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
explanation: CharacterString	Исходный количественный результат представлен в Е.3.4.4.2 самостоятельного отчета по качеству	
pass: Boolean	True	
result: DQ_ConformanceResult		Производный результат соответствия для присутствия промышленных зданий. Исходный количественный результат намеренно не предусмотрен в метаданных. Он описан в самостоятельном отчете по качеству. Атрибут standaloneQualityReportDetails дает точную ссылку на исходный результат в самостоятельном отчете по качеству. Для улучшения читаемости здесь выдается информация только о присутствии промышленных зданий и домов, но если областью определения качества данных является набор данных, то о каждом типе объекта должна быть представлена информация
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Дом	
specification: CI_Citation		
title: CharacterString	Спецификация информационного продукта (см. Е.2.1), требование 7	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
explanation: CharacterString	Исходный количественный результат представлен в самостоятельном отчете по качеству	
pass: Boolean	True	
DQ_DataQuality *id = agg_classification2*		Обобщенный результат корректности классификации зданий
Scope: MD_Scope		Областью определения теперь является тип объекта здание => изменилась единица качества данных. Поэтому был создан новый экземпляр DQ_DataQuality
level: MD_ScopeCode	FeatureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Здания (промышленное здание и дом)	
report: DQ_ThematicClassificationCorrectness		
evaluation: DQ_AggregationDerivation		Метод агрегирования
evaluationMethodType: DQ_EvaluationMethodTypeCode	Косвенный	

Окончание таблицы Е.13

XML-элемент	Пример	Комментарий
evaluationMethodDescription: CharacterString	100-процентное агрегирование значений pass fail результата соответствия по присутствию промышленных зданий и домов	Метод агрегирования
evaluationProcedure: CI_Citation		
title: CharacterString	Приложение J	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	создание	
derivedElement: DQ_Element	*conformance_classification*	Ссылка на исходные результаты
result: DQ_ConformanceResult		
specification: CI_Citation		
title: CharacterString	Спецификация информационного продукта (см. Е.2.1), требование 7	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
pass: Boolean	True	

Е.4.1.4 Создание отчета о соответствии спецификации информационного продукта с использованием элемента применимости

В таблице Е.14 приведен пример вывода информации о соответствии спецификации информационного продукта путем обобщения результатов для различных потребностей. Используемый элемент качества — применимость.

Таблица Е.14 — Создание отчета о применимости в виде метаданных

XML-элемент	Пример	Комментарий
DQ_DataQuality		
scope: MD_Scope		
level: MD_ScopeCode	Набор данных	
standaloneQualityReport: DQ_StandaloneQuali tyReportInformation		Ссылка и описание к прилагаемому самостоятельному отчету по качеству
reportReference: CI_Citation		
title: CharacterString	Выдача информации в виде самостоятельного отчета по оценке качества (см. Е.4.2)	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
abstract: CharacterString	Самостоятельный отчет по качеству, который прилагается к этой оценке качества, представляет более подробную информацию о применяемой оценке и полученных результатов	

Продолжение таблицы Е.14

XML-элемент	Пример	Комментарий
report: DQ_UsabilityElement		Данный элемент используется для вывода информации о соответствии набора данных спецификации информационного продукта
measure: DQ_MeasureReference		
nameOfMeasure: CharacterString	Прошедшая спецификация информационного продукта	
measureIdentification: MD_Identifier		
code: CharacterString	101	
measureDescription: CharacterString	Показатель того, что все требования, упомянутые в спецификации продукта данных, выполнены	
evaluation: DQ_AggregationDerivation		
evaluationMethodType: DQ_EvaluationMethodTypeCode	Косвенный	
evaluationMethodDescription: CharacterString	100-процентное агрегирование значений pass fail результата соответствия для требования, обозначенного в спецификации информационного продукта	
evaluationProcedure: CI_Citation		
title: CharacterString	Приложение J	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
standaloneQualityReportDetails: CharacterString	Исходные результаты описаны в Е.3.4.2, Е.3.4.3.4, Е.3.4.4.4 и Е.3.4.5.3 самостоятельного отчета по оценке качества	Ссылка на исходные результаты в самостоятельном отчете качества (результат соответствия концептуальной согласованности, результат соответствия точности количественных атрибутов для высоты деревьев…)
derivedElement: DQ_Element	*agg_commission1*	Ссылка на обобщенный результат соответствия присутствия транспортной сети, ранее описанный в метаданных
derivedElement: DQ_Element	*(id)*	Ссылка на обобщенный результат соответствия присутствия зданий, ранее описанный в метаданных

Окончание таблицы Е.14

XML-элемент	Пример	Комментарий
derivedElement: DQ_Element	*(id)*	Ссылка на обобщенный результат соответствия присутствия деревьев, ранее описанный в метаданных
derivedElement: DQ_Element	*(id)*	Ссылка на обобщенный результат соответствия отсутствия транспортной сети, ранее описанный в метаданных
derivedElement: DQ_Element	*(id)*	Ссылка на обобщенный результат соответствия отсутствия зданий, ранее описанный в метаданных
derivedElement: DQ_Element	*(id)*	Ссылка на обобщенный результат соответствия отсутствия деревьев, ранее описанный в метаданных
derivedElement: DQ_Element	*(id)*	Ссылка на обобщенный результат соответствия корректности классификации транспортной сети, ранее описанный в метаданных
derivedElement: DQ_Element	*agg_classification2*	Ссылка на обобщенный результат соответствия корректности классификации зданий, ранее описанный в метаданных
derivedElement: DQ_Element	*(id)*	Ссылка на обобщенный результат соответствия корректности классификации деревьев, ранее описанный в метаданных
result: DQ_ConformanceResult		
specification: CI_Citation		
title: CharacterString	Спецификация информационного продукта (см. Е.2.1)	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Создание	
explanation: CharacterString	Не выполняются 3 из 11 требований: набор данных не соответствует	
pass: Boolean	False	

Е.4.2 Создание самостоятельного отчета о качестве

Структура самостоятельного отчета о качестве имеет свободную форму. В Е.2 и Е.3 даны примеры самостоятельного отчета о качестве.

Е.5 Дополнительные примеры

Е.5.1 Общие положения

Некоторые понятия не были описаны в предыдущем примере. В разделах Е.5.2—Е.5.4 даны дополнительные примеры вывода информации об описательном результате, метакачестве и процедурах выборочной оценки. Некоторые понятия не были описаны в примерах Е.4.

Е.5.2 Выдача информации об описательных результатах в виде метаданных

Иногда невозможно представить оценку элемента качества данных в количественном виде. В таких случаях используется описательный результат. Пример описательных результатов в виде метаданных приведен в таблице Е.15.

Таблица Е.15 — Создание отчета об описательных результатах в виде метаданных

XML-элемент	Пример	Комментарий
DQ_DataQuality		
scope: MD_Scope		
level: MD_ScopeCode	Набор данных	Набор данных описывает археологические объекты
report: DQ_RelativeInternalPositionalAccuracy		
evaluation: DQ_IndirectEvaluation		
evaluationMethodType: DQ_EvaluationMethodTypeCode	Косвенный	
evaluationMethodDescription: CharacterString	Сравнение абсолютной позиционной точности археологических объектов и абсолютной позиционной точности рек	
deductiveSource: CharacterString	Позиционная точность рек вблизи археологического лагеря	
result: DQ_DescriptiveResult		
statement: CharacterString	Относительная позиционная точность между археологическими объектами и реками выше, чем абсолютная позиционная точность археологических объектов (5 м)	

Е.5.3 Создание отчета о метакачестве в виде метаданных

При оценке абсолютной позиционной точности в топологическом исследовании на месте археологических раскопок результат точности составляет 5 м.

Определение качества оценки затем проводится с помощью элемента доверительного метакачества, для которого применяется мера «Фактор безопасности».

В таблице Е.16 представлена выдача информации о метакачестве в виде метаданных.

Таблица Е.16 — Создание отчета о метакачестве в виде метаданных

XML-элемент	Пример	Комментарий
DQ_DataQuality		
scope: MD_Scope		
level: MD_ScopeCode	Набор данных	

Продолжение таблицы Е.16

XML-элемент	Пример	Комментарий
report: DQ_AbsolutExternalPositionalAccuracy *id = positionalaccuracy1*		Отчет об абсолютной позиционной точности. Идентификатор (id) присваивается элементу качества данных для возможности ссылаться на него в следующем элементе метакачества. Все дополнительные атрибуты не были здесь заполнены
measure: DQ_MeasureReference		
nameOfMeasure: CharacterString	Средняя квадратическая ошибка	
measureIdentification: MD_Identifier		
code: CharacterString	39	
measureDescription: CharacterString	Стандартное отклонение, где истинное значение не оценивается по наблюдениям, а известно априори	
evaluation: DQ_FullInspection		
evaluationMethodType: DQ_Evaluation MethodTypeCode	directExternal	
evaluationProcedure: CI_Citation		
title: CharacterString	Процедура оценки качества данных IGN	
date: CI_Date		
date: Date	1995-02-09	
dateType: CI_DateTypeCode	Создание	
result: DQ_QuantitativeResult		
value: Record	5	
valueUnit: UnitOfMeasure	Метр	
report: DQ_Confidence		Отчет о метакачестве (confidence), связанный с предыдущим отчетом о качестве
relatedElement: DQ_Element	*positionalaccuracy1*	
measure: DQ_MeasureReference		
nameOfMeasure: CharacterString	Фактор безопасности	
measureIdentification: MD_Identifier		
code: CharacterString	1	
authority: CI_Citation		
title: CharacterString	Меры IGN	
date: CI_Date		
date: Date	1995-01-01	
dateType: CI_DateTypeCode	Создание	
measureDescription: CharacterString	Соотношение между классом точности элементов оценки и классом точности, которое должно быть получено в наборе данных	
evaluation: DQ_FullInspection		

Окончание таблицы Е.16

XML-элемент	Пример	Комментарий
evaluationMethodType: DQ_EvaluationMethodTypeCode	directExternal	
evaluationMethodDescription	Чем выше «фактор безопасности», тем достовернее оценка. Для проверки оценки «фактор безопасности» должен быть выше 2	
evaluationProcedure: CI_Citation		
title: CharacterString	Arrêté 2003 (Законодательство Франции)	
date: CI_Date		
date: Date	2003	
dateType: CI_DateTypeCode	Публикация	
result: DQ_QuantitativeResult		
value: Record	2.4	
valueUnit: UnitOfMeasure		

Е.5.4 Как выдавать информацию о процедуре выборочного контроля

За основу примера взята топографическая БД (ТБД) европейского Национального управления земельной съемки.

Уровень соответствия качества установлен в спецификации к информационному продукту.

В данном примере тип объекта «дорога» оценивается посредством выборочной оценки.

Процедура выборки проходит с использованием принципов [14], как описано в таблице Е.17.

Таблица Е.17 — Процедура выборочного контроля

Этап процесса	Пример
Определение метода выборки	Многоступенчатый отбор. Отбор достаточного количества единиц для выполнения выборочного отношения. Отбор осуществляется на основе взвешенных объектов
Определение элементов	Все объекты
Разделение области определения качества данных (совокупности) на партии	Некоторое количество наборов данных
Разделение партий на единицы выборки	N-число квадратов 1 × 1 км
Определение доли выборки или объема выборки	Объем выборки зависит от значения AQL для этой партии
Отбор единиц выборки	Отбор необходимого количества единиц выборки, чтобы было выполнено выборочное отношение или достигнут требуемый объем выборки в элементах
Контроль элементов в единицах выборки	Контроль каждого элемента в единицах выборки

Если требования к качеству объекта — 1 несоответствие на 100 единиц (AQL = 1), то все собранные объекты проверяются из источника данных. Проверка выборочным контролем производится при AQL в диапазоне от 4 или 15.

Контролируемая партия для тестирования должна состоять из наборов данных, которые созданы по возможности в одно время и с использованием одинаковых методов. Из партии единицы выборки, состоящей из N-количества квадратов размером 1 × 1 км^2, выбираются таким образом, чтобы число объектов в выборке было достаточным для AQL = 4. В таблице Е.18 показан пример того, как выдавать информацию о процедуре выборочного контроля в виде метаданных.

Таблица Е.18 — Выдача информации о процедуре выборочного контроля в виде метаданных

XML-элемент	Пример
DQ_DataQuality	
scope: MD_Scope	
level: MD_ScopeCode	Тип объекта
levelDescription: MD_ScopeDescription	
features: GF_FeatureType	Дорога
report: DQ_Commission	
measure: DQ_MeasureReference	
nameOfMeasure: CharacterString	Количество избыточных элементов
measureIdentification: MD_Identifier	
code: CharacterString	2
measureDescription: CharacterString	Количество элементов в наборе данных, которые не должны присутствовать в наборе данных
evaluation: DQ_SampleBasedInspection	
evaluationMethodType: DQ_EvaluationMethodTypeCode	directExternal
evaluationMethodDescription: Character String	Многоступенчатый отбор. Отбор необходимого количества единиц выборки для выполнения соотношения выборки. Отбор выборки осуществляется на основе взвешенных объектов
evaluationProcedure: CI_Citation	
title: CharacterString	Приложение F
date: CI_Date	
date: Date	2010-07-05
dateType: CI_DateTypeCode	Приложение
referenceDoc: CI_Citation	
title: CharacterString	[14]
date: CI_Date	
date: Date	1999-11-18
dateType: CI_DateTypeCode	Публикация
lotDescription: CharacterString	Партия — это группа БД (лист карты масштаба 1:10 000), которые отбираются для контроля. Размер партии — это количество объектов в партии. Все дороги в наборе данных (одна партия для всего набора данных)
samplingScheme: CharacterString	Из партии отбирается область из множества квадратов 1 × 1 км таким образом, что количество дорог в выборке по крайней мере такое же, как требует AQL = 4
samplingRatio: CharacterString	В среднем область, включающая форматные листы (16 баз данных) с 6—10 квадратами (1 × 1 км), рекомендуется в качестве практического размера партии

Приложение F
(справочное)

Методы выборочной оценки

F.1 Введение

Данное приложение содержит рекомендации по определению выборки и разработке методов выборочного контроля. Для формирования выборки при оценке соответствия спецификации информационного продукта могут применяться стандарты ГОСТ Р ИСО 2859, ГОСТ Р ИСО 3951-1. Эти стандарты изначально были разработаны для непространственного использования. В настоящем приложении описывается, как применять методы выборочного контроля стандартов серии ГОСТ Р ИСО 2859 и ГОСТ Р ИСО 3951-1, а также другие методы пространственной выборки для пространственных данных.

F.2 Партия и элемент

Партия и элемент являются важными понятиями в методе выборочного контроля согласно серии ГОСТ Р ИСО 2859 и ГОСТ Р ИСО 3951-1. Партия — это минимальная единица продукции, для которой может быть оценено качество. Элемент — это минимальная единица, которая должна быть определена разработчиком данных в соответствии со спецификацией информационного продукта.

F.3 Объем выборки

Объем совокупности, а следовательно, и объем выборки могут оцениваться в зависимости от различных базисов к элементам. Определение объема выборки требует точного выявления элементов. Примеры различных базисов даны в таблице Е.1.

На рисунке Е.1 показана разница между проекциями. Вся фигура представляет данные в области определения качества данных. Фигура изображает возможную площадь выборки из приблизительно 15 % общей площади области определения качества данных, только около 10 % от длины кривой в области определения выборки и 0 % вершин.

Чтобы избежать проблем с выборкой, представленных на рисунке F.1, необходимо определить объем и расположение выборки с использованием комбинации различных критериев, при этом улучшается репрезентативность выборки.

Пример — Выборка включает 10 % площади, охваченной набором данных, и содержит не менее 5 % общей длины дорог в наборе данных.

Таблица F.1 — Различные базисы для определения совокупности

Базис	Размер набора данных	Объем выборки
Объекты	Количество объектов данного типа	Количество объектов данного типа, выраженное в процентах от общего количества объектов
Охваченная область	Область, покрытая набором данных	Площадь, покрытая выборкой и выраженная в процентах от общей площади
Кривые	Общая длина кривых в наборе данных	Длина выборочных кривых, выраженная в процентах от общей длины
Вершины	Общее количество вершин, описывающих кривые или области в наборе данных	Количество вершин в выборке, выраженное в процентах от общего количества вершин

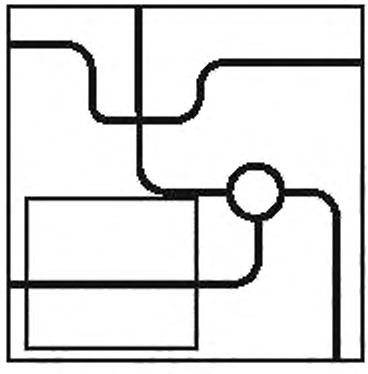


Рисунок F.1 — Влияние местоположения области выборки на репрезентативность элементов в выборке

Примечание — Область определения качества данных — это область внутри большого квадрата. Область выборки отмечена темным квадратом.

F.4 Стратегии создания выборки

F.4.1 Введение

В данном разделе приводятся рекомендации для определения выборок и выборочных методов с учетом конкретных аспектов пространственных данных. Стратегии выборочного построения, описанные в настоящем приложении, представлены в графической форме на рисунке F.2. Различают два аспекта стратегии создания выборки: определение элементов для выборки (область или объект) и способ, с помощью которого происходит отбор элементов (случайный или преднамеренный).



Рисунок F.2 — Отношения стратегий создания выборки

F.4.2 Вероятностная и детерминированная выборка

F.4.2.1 Различия

Вероятностная выборка применяет теорию выборочного метода и включает случайный выбор элементов выборки. Существенной характеристикой вероятностной выборки является то, что каждый компонент совокупности, из которой формируется выборка, имеет известную вероятность выбора. При случайной выборке можно делать статистические выводы о выборочной совокупности. Составление преднамеренной выборки сопряжено с отбором выборок на основе экспертных знаний или профессиональной оценки.

F.4.2.2 Простая случайная выборка

Простая случайная выборка основана на вероятности и включает в себя составление выборки в случайном порядке. Конкретная выборка (например, объекты, положение, время) осуществляется с помощью случайных чисел для отбора элементов, при этом любой выбор равновероятен. Простая случайная выборка применяется, когда генеральная совокупность является довольно однородной относительно характеристик отбираемых образцов, т. е. не содержит больших пропусков и скоплений. Данный метод не может привести к репрезентативному охвату области, т. е. существует вероятность, что полученная выборка будет содержать только часть области.

F.4.2.3 Стратифицированная случайная выборка

Стратифицированная выборка требует, чтобы генеральная совокупность была разделена на непересекающиеся слои или подсовокупности, которые являются более однородными по элементам выборки в одном слое, чем в разных слоях. Данная стратегия выборки перспективна для получения большей точности в определении средней и дисперсии, чем не стратифицированная для той же совокупности.

F.4.2.4 Полуслучайная выборка

Полуслучайная или систематическая выборка подразумевает случайный отбор исходных элементов выборки (например, положение, время, объект) и правила для отбора всех остальных элементов. Примером полуслучайной или систематической выборки служит решетчатая выборка, где начальное положение сетки определяется случайным образом, а образцы берутся через равные промежутки (ячейки сетки) пространственной области. Систематическая решетчатая выборка используется для поиска кластеров и вывода средних значений, процентов или других параметров и подходит для оценки пространственных тенденций или закономерностей. Данный метод предоставляет практичный и легкий способ обеспечения покрытия области.

F.4.3 Выборка на основе объектов и выборка на основе их пространственного расположения

F.4.3.1 Формирование выборки на основе объектов (непространственная выборка)

Особенностью данной выборки является отбор элементов выборки на основе непространственных атрибутов объектов, а не на их пространственном расположении. Выбор объектов в пределах области определения качества данных может производиться в случайном порядке при одинаковых условиях создания данных для всей области определения качества данных. В некоторых случаях в результате простой случайной выборки можно не

получить удовлетворительной выборки в силу того, что однородность может быть выявлена только для поднаборов, и тогда потребуется равномерное распределение выборок; т. е. пропуски и скопления встречаются в выбираемых характеристиках. В этом случае стратифицированная или полуслучайная выборки могут дать лучшие результаты.

Примечание — Если формирование выборки осуществляется путем случайного выбора объектов, то существует риск получения выборки, сосредоточенной на малой площади (которая может быть неприемлема).

Полуслучайная выборка может использоваться, чтобы гарантировать контроль различных критериев по объему выборки и/или расположению, ограниченных необходимостью сократить затраты на проведение контроля.

Пример — Энергетической компании необходимо оценить правильность атрибутов, определяемых для объектов различных типов. Были рассмотрены два метода: случайный и полуслучайный отбор (произвольный выбор объектов одного типа, а затем сбор объектов различных типов в соседстве с первым до заполнения выборки каждого типа), которые способствовали снижению затрат на контроль в процессе эксплуатации.

F.4.3.2 Формирование выборки на основе пространственного распределения (пространственная выборка)

В рамках данной выборки отбор единиц выборки основан на пространственном распределении. Элементами выборки могут выступать существующие географические единицы (например, политические или статистические) или другие объекты или компоненты объектов предметной области, для которых проводится контроль. Данный тип формирования выборки может использоваться в качестве первого этапа выборки с последующим использованием метода формирования выборки на основе объектов в пределах каждой подобласти.

Пример — Случайный выбор UTM области координатной сетки 1 × 1 км для того, чтобы оценить атрибуты объектов, содержащихся в этой области.

Рисунок F.3 иллюстрирует результат определения областей, которые будут представлены для проверки, полученных путем случайной генерации координат центральной точки квадратов равной площади (неперекрывающиеся).

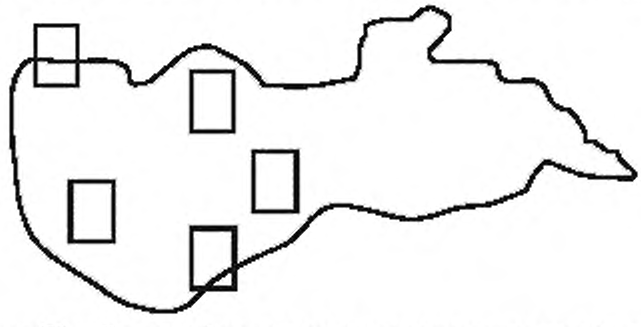
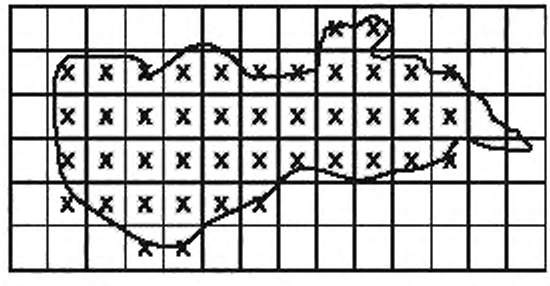


Рисунок F.3 — Пример выборки на основе пространственного распределения

Когда особое значение имеет покрытие всей площади, то местоположение выборки следует определять в соответствии с регулярной или полурегулярной сеткой. Рисунок F.4 иллюстрирует пример полуслучайной (систематической) выборки с выборочными объектами, распределенными вдоль регулярной сетки для оценки позиционной точности набора данных.



Примечание — «X» обозначает ячейки сетки, выбранные в соответствии с правилом для включения в выборку.

Рисунок F.4 — Пример регулярной или случайной выборки на основе пространственного распределения

Пространственное позиционирование с различным объемом в различных областях набора данных может быть необходимо в полуслучайной выборке, если распределение объектов неоднородно. При использовании сетки с постоянным размером ячейки необходимо правило для включения или исключения клеток, которые не представляют интереса.

F.5 Вероятностная выборка

F.5.1 Общее представление

При применении выборки необходимо учитывать следующее:

а) области, охватываемые географическим набором данных, могут образовывать непрерывное пространство. При разделении набора данных на партии особое внимание должно быть уделено присутствию или отсутствию элементов, проходящих через границы партии;

б) ряд факторов, включая качество исходных данных и мастерство операторов, могут повлиять на качество пространственных данных. Разработчикам данных необходимо проявлять осторожность при определении партии для получения однородности в контексте обеспечения качества.

F.5.2 Существующий стандарт выборочного контроля

F.5.2.1 Общие положения

Исходя из особенностей разработки и в соответствии со спецификацией информационного продукта для проведения выборочного контроля должны использоваться соответствующие существующие стандарты. Например, ГОСТ Р ИСО 2859-1 применим в первую очередь для контроля непрерывной серии партий. Для индивидуальных партий подходит ГОСТ Р 50779.72, а для процедур выборочного контроля независимо от партий применяется ГОСТ Р ИСО 2859-3. ГОСТ Р ИСО 3951-1 применяется при контроле по количественным переменным для определения процента несоответствующих элементов продукции.

Уровень соответствия качества набора данных указывается как AQL в соответствии с ГОСТ Р ИСО 3951-1.

Для определения соответствия каждого элемента должны быть определены допуски спецификации при применении ГОСТ Р ИСО серии 2859 и ГОСТ Р 50779.72. При применении ГОСТ Р ИСО 3951-1 статистические показатели качества следует указывать на основе спецификации информационного продукта.

F.5.2.2 Полезные таблицы на основе этих стандартов — объем выборки и пределы отбраковки

F.5.2.2.1 Общие положения

При выборочном контроле оцениваемый отсутствующий показатель нельзя непосредственно сравнить с AQL. Таблицы Е.2 и F.4 определяют основные указания по объему выборки в зависимости от объема набора данных, а также связанного предела отбраковки.

F.5.2.2.2 Оценка соответствующих/несоответствующих элементов с учетом выборки

В таблице Е.2 ниже представлен рекомендуемый объем выборки в соответствии с генеральной совокупностью и связанный предел отбраковки для оценки соответствующих/несоответствующих элементов, например для оценки полноты. Она основана на гипергеометрическом распределении (ссылка [20]). Предполагается, что отклонения соответствуют этому распределению.

Как пользоваться таблицей:

а) выбрать объем совокупности элемента для проверки;

б) выбрать объем выборки (n) из таблицы;

в) произвести оценку и рассчитать количество «ошибочных элементов»;

г) вся совокупность бракуется, если количество ошибок равно или превышает предел отбраковки для фактического n и p_0 (AQL).

Таблица F.2 — Статистические показатели соответствия/несоответствия элементов при 95-процентном уровне значимости

Объем совокупности		p_0 =	0.5 %	1,0 %	2,0 %	3,0 %	4,0 %	5,0 %
От	До	Объем выборки (n)	Уровень отбраковки					
1	8	Все	1	1	1	1	1	1
9	50	8	1	1	1	2	2	2
51	90	13	1	1	2	2	2	3
91	150	20	1	2	2	3	3	4
151	280	32	1	2	3	3	4	4
281	400	50	2	3	3	4	5	6
401	500	60	2	3	4	5	6	7

Окончание таблицы F.2

Объем совокупности		p_0 =	0,5 %	1,0 %	2,0 %	3,0 %	4,0 %	5,0 %
От	До	Объем выборки (n)	Уровень отбраковки					
501	1200	80	3	3	5	6	7	8
1201	3200	125	3	4	6	8	10	11
3201	10 000	200	4	6	8	11	14	16
10 001	35 000	315	5	7	12	16	20	23
35 001	150 000	500	6	10	16	23	28	34
150 001	500 000	800	9	14	24	33	42	51
> 500 000		1250	12	20	34	49	63	76

Примечания

1 Если объем выборки выше, чем минимальный объем, приведенный в таблице, то предел отбраковки следует рассчитывать индивидуально. Данная проверка справедлива для ситуаций, когда оценка качества основана на оценке пригодности/непригодности элементов.

2 Существуют другие диапазоны статических значений, помимо представленных в таблице Е.2.

Пример — Проверка отсутствующих домов (полнота/отсутствие) в определенной области.

Первая область выборки выбрана и в этой области проверен каждый дом на предмет его присутствия или отсутствия в наборе данных. Далее оценивается (подсчитывается) число отсутствующих домов и их общее количество. Вопрос в следующем: значительно ли полученный результат отличается от предела приемлемого качества (AQL)? Если да, то набор данных бракуется. Если нет, то набор данных принимается.

Набор данных для проверки состоит из 2440 зданий.

Объем выборки (из таблицы Е.2) – n = 125. Контроль по полю показывает, что пропущены 2 здания, выдавая примерный процент отсутствия: 2 / (125 + 2) · 100 % = 1,6 %.

AQL (из спецификации информационного продукта для набора данных): p_0 = 0,5 %.

1,6 % больше чем 0,5 %, но есть ли вероятность отбраковки набора данных? В случае применения выборки полученный показатель отсутствия нельзя непосредственно сравнивать с AQL. Для этого необходима односторонняя проверка гипотез, и в этом помогает таблица F.2.

Предел отбраковки (N = 125, p_0 = 0,5 %) равен 3. В результате полевого контроля обнаружены 2 пропущенных элемента.

Заключение: так как 2 меньше, чем 3 (предел отбраковки), набор данных не может быть отклонен и принимается.

F.5.2.2.3 Стандартное отклонение

В таблице F.4 представлен рекомендуемый объем выборки в соответствии с объемом совокупности, а также связанный предел отбраковки при измерении стандартного отклонения.

Данный статистический метод применяется для определения того, превышено ли стандартное отклонение для выборки AQL. Приведенная ниже таблица F.4 основана на нормальном распределении и предполагает нормальное распределение отклонений.

Символы и формулы к таблице F.4 представлены в таблице F.3.

Таблица F.3 — Символы и формулы

Стандартное отклонение, определенное на основе выборки	s
Объем выборки	n
AQL для стандартного отклонения	σ
F (из F-распределения)	$F_{0.05,\ n-1,\ \infty}$
Доверительный интервал	$\frac{s}{\sqrt{F}} \cdot s \cdot \sqrt{F}$
Стандартное отклонение превышено, если:	$\sigma < \frac{s}{\sqrt{F}}$

Набор данных не отвечает требованиям (т. е. может быть отклонен при значимости 95 %), если стандартное отклонение, разделенное на F-значение (взято из таблицы F.4), выше, чем AQL.

Таблица F.4 — Статистические показатели проверки стандартного отклонения. Уровень значимости 95 %

Объем совокупности		Объем выборки (n)	$\sqrt{F_{0,05,n-1,\infty}}$
От	До		
26	50	5	1,54
51	90	7	1,45
91	150	10	1,37
151	280	15	1,30
281	400	20	1,26
401	500	25	1,23
501	1200	35	1,20
1201	3200	50	1,16
3201	10 000	75	1,13
10 001	35 000	100	1,12
35 001	150 000	150	1,09
150 001	500 000	200	1,08
> 500 000		200	1,08

Пример — Оценка позиционной/абсолютной точности крышек люков.

В наборе данных из 450 крышек люков измеряются 25 (объем выборки n = 25). Приблизительное стандартное отклонение 21 см, приемлемый уровень качества (AQL) = 19 см.

Нижний предел доверительного интервала = 21 см / 1,23 (из таблицы F.4) = 17,1 см. AQL (19 см) находится в пределах доверительного интервала стандартного отклонения.

Вывод: стандартное отклонение по контролю не является значительно выше, чем AQL, и набор данных не может быть отбракован.

F.5.3 Процесс составления выборки

F.5.3.1 Определение элементов

Элементы необходимо определять в соответствии со спецификацией информационного продукта или требований. Если несоответствующие элементы статистически согласованы, то они обрабатываются как один элемент.

F.5.3.2 Определение областей определения качества контролируемого набора данных

Если область определения качества данных не является однородной, то их следует разделить на однородные поднаборы. Эти однородные поднаборы следует рассматривать как отдельные области определения качества данных.

Однородность может быть выведена там, где выполняются следующие условия:

- исходные данные продукции имеют почти одинаковое качество;
- системы производства данных (оборудование, программное обеспечение, квалификация оператора) практически не отличаются;
- другие факторы, которые могут повлиять на вероятность появления несоответствий, такие как сложность и плотность объектов, практически не отличаются.

F.5.3.3 Разделение области определения качества данных на партии

Генерирование партий осуществляется путем деления областей определения качества данных. При сильной положительной пространственной автокорреляции возникновения несоответствий предпочтительнее использовать меньший объем партии.

F.5.3.4 Разделение партий на единицы выборки

Единицей выборки может быть существующий географический район или другое разделение предметной области, для которой проводится проверка. Когда единицей выборки служит географическая область, то необходимо применять правила для частичного включения единиц продукции в единицу выборки.

F.5.3.5 Отбор единиц выборки с помощью простой случайной выборки для проверки

Общее число элементов, которые относятся к выбранным единицам выборки, следует устанавливать в соответствии с актуальными международными стандартами.

Примечание — Если партия является статистически гетерогенной, то применение простой случайной выборки с тем же уровнем выборки недопустимо. ИСО серии 2859 дополнительно предусматривают использование стратифицированной выборки.

F.5.3.6 Контроль выбранных единиц выборки

Все элементы, которые принадлежат отобранным единицам выборки, подвергаются проверке. Элементы в наборе данных сравниваются с предметной областью в соответствии с выбранной мерой качества.

Приложение G
(обязательное)

Базовые меры качества данных

G.1 Цель базовых мер качества данных

Понятие базовой меры качества данных вводится в настоящем стандарте с целью предотвращения повторного определения одного и того же понятия. Существуют меры качества данных, которые имеют определенную схожесть. Например, счетные меры качества данных работают по принципу подсчета ошибок. Число ошибок может использоваться для создания различного рода мер качества данных. Концепция построения этих мер качества данных направлена на обобщение базовых мер качества данных, которые используются для создания мер качества данных, разделяющих эту общность.

Можно выделить меры подсчета и неопределенности качества данных. Поэтому в данном приложении приведены две принципиальные категории базовых мер качества данных. Счетные меры качества данных основаны на концепции подсчета ошибок или правильных элементов. Меры неопределенности в обеспечении качества данных основаны на концепции моделирования неопределенности измерений с помощью статистических методов. Измеряемая величина может лежать в разных размерностях. В зависимости от размерности измеряемой величины применяются различные типы базовых мер качества данных для создания мер качества данных.

G.2 Базовые меры качества данных, связанные с подсчетом

Базовые меры качества данных, основанные на различных методах подсчета ошибок или количества правильных значений, приведены в таблице G.1.

Таблица G.1 — Базовые меры качества данных для измерения качества данных, связанного с подсчетом

Имя базовой меры качества данных	Определение базовой меры качества данных	Пример	Тип значения качества данных
Error indicator	Показатель того, что элемент ошибочный	False	Boolean (если значение «true» элемент неправильный)
Correctness indicator	Показатель того, что элемент правильный	True	Boolean (если значение «true» элемент правильный)
Error count	Общее количество элементов, которые являются предметом ошибки указанного типа	11	Integer
Correct items count	Общее количество элементов, которые свободны от ошибки указанного типа	571	Integer
Error rate	Количество ошибочных элементов по отношению к общему количеству элементов, которые должны присутствовать	0,0189	Real
Correct items rate	Количество правильных элементов по отношению к общему количеству элементов, которые должны присутствовать	0,9811	Real

Примечания

1 Показатель ошибок может быть представлен в процентах или как отношение. Единица значения в количественном результате (см. 7.5.4.2) может быть использована для указания того, в каком виде представлен результат в процентах или как отношение.

2 Показатель правильных элементов может быть представлен в процентах или как отношение. Единица значения в количественном результате (см. 7.5.4.2) может быть использована для указания того, в каком виде представлен результат в процентах или как отношение.

Примечание — Количество элементов определяется по числу элементов в предметной области набора данных, заданных областью определения качества данных.

Пример — Применение числа элементов из реального мира или эталонного набора данных.

G.3 Базовые меры качества данных, связанные с неопределенностью

G.3.1 Общие положения

Числовые значения, получаемые в результате измерения, можно производить только до определенной точности. Принимая измеряемую величину в виде случайной величины, данную неопределенность можно измерить количественно. Для определения базовых мер качества данных, связанных с неопределенностью, используют различные способы описания неопределенности посредством статистических методов.

Статистические методы, используемые для определения мер качества данных, связанных с неопределенностью, основаны на следующих предположениях:

- неопределенность является однородной для всех наблюдаемых значений;
- наблюдаемые значения не являются коррелированными;
- наблюдаемые значения имеют нормальное распределение.

G.3.2 Одномерная случайная величина Z

Для измеряемой величины невозможно определить истинное значение. Но можно найти вероятность истинного значения в пределах определенного интервала. Этот интервал называется доверительным интервалом. Он представлен вероятностью P истинного значения, находящегося между нижним и верхним пределами. Величину P также называют уровнем значимости

$$P \text{ (нижний предел} \le \text{истинное значение} \le \text{верхний предел)} = P.$$

Если стандартное отклонение σ известно, то пределы задаются квантилями u нормального (Гауссова) распределения

$$P(z_t - u \cdot \sigma \le \text{истинное значение} \le z_t + u \cdot \sigma) = P.$$

См. также таблицу G.2.

Таблица G.2 — Связь между квантилями нормального распределения и уровнем значимости

Вероятность P, %	Квантиль	Базовая мера качества данных	Имя	Тип значения качества данных
$P = 50$	$u_{50\ \%} = 0{,}6745$	$u_{50\ \%} \cdot \sigma_Z$	LE50	Мера
$P = 68{,}3$	$u_{68{,}3\ \%} = 1$	$u_{68{,}3\ \%} \cdot \sigma_Z$	LE68.3	Мера
$P = 90$	$u_{90\ \%} = 1{,}645$	$u_{90\ \%} \cdot \sigma_Z$	LE90	Мера
$P = 95$	$u_{95\ \%} = 1{,}960$	$u_{95\ \%} \cdot \sigma_Z$	LE95	Мера
$P = 99$	$u_{99\ \%} = 2{,}576$	$u_{99\ \%} \cdot \sigma_Z$	LE99	Мера
$P = 99{,}8$	$u_{99{,}8\ \%} = 3$	$u_{99{,}8\ \%} \cdot \sigma_Z$	LE99.8	Мера

Если стандартное отклонение σ неизвестно, но одномерная случайная величина Z измеряется избыточно посредством независимых наблюдений N, то стандартное отклонение можно определить из наблюдений (см. таблицу G.3).

z_{mi} представляет i-е измерение значения. Если истинное значение z_t для Z известно, то стандартное отклонение вычисляется по формуле

$$s_Z = \sqrt{\frac{1}{r}\sum_{i=1}^{N}\left(z_{mi} - z_t\right)^2},$$

где избыточность r представляет число наблюдений $r = N$. Если истинное значение неизвестно, то его можно вычислить как среднее арифметическое наблюдений

$$z_t = \sum_{i=1}^{N} z_{mi}.$$

Стандартное отклонение можно затем определить по той же формуле, при

$$r = N - 1.$$

Если стандартное отклонение определяется посредством избыточных измерений, то доверительный интервал можно вывести из t-распределения Стьюдента с параметром r

$$P\left(-t \cdot s_z \le Z - z_t \le t \cdot s_z\right) = P, \text{ при } \frac{(Z - z_t)}{s_z} \sim t(r).$$

Таблица G.3 — Связь между квантилями t-распределения Стьюдента и уровнем значимости для различных коэффициентов избыточности *r*

Вероятность *P*, %	Квантиль *r* = 10	Квантиль *r* = 5	Квантиль *r* = 4	Квантиль *r* = 3	Квантиль *r* = 2	Квантиль *r* = 1
P = 50	*t* = 1,221	*t* = 1,301	*t* = 1,344	*t* = 1,423	*t* = 1,604	*t* = 2,414
P = 68,3	*t* = 1,524	*t* = 1,657	*t* = 1,731	*t* = 1,868	*t* = 2,203	*t* = 3,933
P = 90	*t* = 2,228	*t* = 2,571	*t* = 2,776	*t* = 3,182	*t* = 4,303	*t* = 12,706
P = 95	*t* = 2,634	*t* = 3,163	*t* = 3,495	*t* = 4,177	*t* = 6,205	*t* = 25,452
P = 99	*t* = 3,581	*t* = 4,773	*t* = 5,598	*t* = 7,453	*t* = 14,089	*t* = 127,321
P = 99,8	*t* = 4,587	*t* = 6,869	*t* = 8,610	*t* = 12,924	*t* = 31,599	*t* = 636,619

Таблица G.4 — Базовые меры качества данных для различных вероятностей *P* одномерной величины, где стандартное отклонение определяется на основе избыточных измерений

Вероятность *P*, %	Базовая мера качества данных	Наименование	Тип значения качества данных
P = 50,0	$t_{50\,\%}(r) \cdot s_Z$	LE50(r)	Мера
P = 68,3	$t_{68.3\,\%}(r) \cdot s_Z$	LE68.3(r)	Мера
P = 90,0	$t_{90\,\%}(r) \cdot s_Z$	LE90(r)	Мера
P = 95,0	$t_{95\,\%}(r) \cdot s_Z$	LE95(r)	Мера
P = 99,0	$t_{99\,\%}(r) \cdot s_Z$	LE99(r)	Мера
P = 99,8	$t_{99,8\,\%}(r) \cdot s_Z$	LE99.8(r)	Мера
Примечание — Значения *t* для числа избыточности можно получить из таблицы G.3.			

Базовые меры качества данных для неопределенности одномерных величин приведены в таблицах G.2 и G.4. Они направлены на измерение неопределенности с указанием верхней и нижней границы доверительного интервала. Разница заключается в том, как получено стандартное отклонение. Если оно известно априори, то применяется таблица G.2. Если стандартное отклонение определяется избыточными измерениями, то применяется таблица G 4 в сочетании с таблицей G.3.

G.3.3 Двумерная случайная величина *X*, *Y*

Помимо одномерной случайной величины *Z* существует двумерная величина, которая всегда определяется двумя значениями. Результат представляется парой *X*, *Y*. Как и в случае одномерной случайной величины, она использует те же предположения.

Результаты наблюдений — x_m и y_m. Эквивалентностью доверительного интервала в одном измерении является доверительная область, которая обычно описывается как окружность вокруг лучшей оценки для истинного значения. Вероятность нахождения истинного значения в этой области вычисляется посредством интегрирования области по двумерной функции плотности нормального распределения. Окружность характеризуется его радиусом. Этот радиус, *R*, используется в качестве меры точности двумерных случайных величин (см. также таблицу G.5)

$$P(R, \sigma_X, \sigma_Y) = \frac{1}{2\pi\sigma_X\sigma_Y} \iint_{(x-x_t)^2+(y-y_t)^2=R^2} e^{-\frac{1}{2}\left[\frac{(x-x_t)^2}{\sigma_X^2}+\frac{(y-y_t)^2}{\sigma_Y^2}\right]} dxdy.$$

В некоторых особых случаях радиус может быть рассчитан в зависимости от стандартных отклонений σ_x и σ_y.

Таблица G.5 — Отношение между вероятностью *P* и соответствующим радиусом окружности

Вероятность *P*, %	Базовая мера качества данных	Наименование	Тип значения качества данных
P = 39,4	$\frac{1}{\sqrt{2}}\sqrt{\sigma_x^2+\sigma_y^2}$	CE39.4	Мера
P = 50	$\frac{1,1774}{\sqrt{2}}\sqrt{\sigma_x^2+\sigma_y^2}$	CE50	Мера
P = 90	$\frac{2,146}{\sqrt{2}}\sqrt{\sigma_x^2+\sigma_y^2}$	CE90	Мера
P = 95	$\frac{2,4477}{\sqrt{2}}\sqrt{\sigma_x^2+\sigma_y^2}$	CE95	Мера
P = 99,8	$\frac{3,5}{\sqrt{2}}\sqrt{\sigma_x^2+\sigma_y^2}$	CE99.8	Мера

G.3.4 Трехмерная случайная величина *X*, *Y*, *Z*

Одномерная случайная величина *Z* может быть расширена до трехмерной, где результат всегда наблюдается тремя значениями. Результат представляется *X*, *Y*, *Z*. В ее основе лежат те же предположения, что и в случае одномерной случайной величины.

Результаты наблюдений — x_{mi}, y_{mi} и z_{mi}. Эквивалентность доверительного интервала в одном измерении является доверительным объемом, который обычно описывается как сфера вокруг лучшей оценки для истинного значения. Вероятность нахождения истинного значения в этом объеме рассчитывается по интегрированию объема над трехмерной функцией плотности нормального распределения. Объем сферы характеризуется его радиусом. Этот радиус используется в качестве меры точности трехмерных случайных величин (см. таблицу G.6).

Таблица G.6 — Отношение между вероятностью *P* и соответствующим радиусом сферы

Вероятность *P*, %	Базовая мера качества данных	Наименование	Тип значения качества данных
P = 50	$0,51\cdot(\sigma_x+\sigma_y+\sigma_z)$	Вероятностная сферическая ошибка (SEP)	Мера
P = 61	$\sqrt{\sigma_x^2+\sigma_y^2+\sigma_z^2}$	Средняя радиальная сферическая ошибка (MRSE)	Мера
P = 90	$0,833\cdot(\sigma_x+\sigma_y+\sigma_z)$	90-процентный стандарт сферической точности	Мера
P = 99	$1,122\cdot(\sigma_x+\sigma_y+\sigma_z)$	99-процентный стандарт сферической точности	Мера

Приложение Н
(справочное)

Управление мерами качества данных

Н.1 Введение

В данном приложении дано описание хранения мер качества данных, базовых мер и параметров в реестре или каталоге.

Н.2 Хранение мер качества данных

Н.2.1 Общие положения

Полное описание мер качества данных, базовых мер качества данных и параметров могут храниться в реестре либо в каталоге. Эти две структуры совместимы и дополняют друг друга. Реестр носит глобальный характер (например, реестр для всех применяемых мер в структуре), каталог же представляет набор информации, относящейся к одному конкретному варианту использования (например, каталог комплекса мер, используемых для оценки качества данных одного конкретного набора данных).

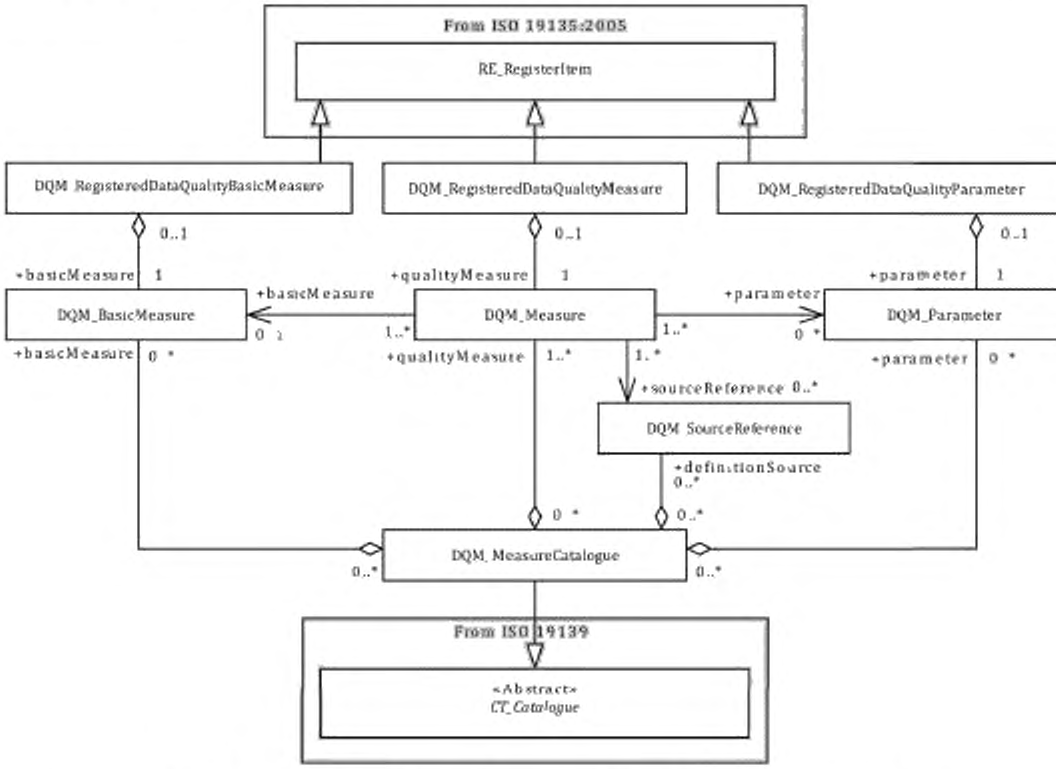


Рисунок Н.1 — Внесенные в реестр элементы, каталоги и меры качества данных

Н.2.2 Каталог мер качества данных

Меры, базовые меры, указатели источника и параметры могут быть представлены в каталоге мер: DQM_MeasureCatalogue, выведенного из класса CT_Catalogue согласно [4].

DQM_MeasureCatalogue должен объединять все нужные экземпляры DQM_Measure, DQM_BasicMeasure, DQM_SourceReference и DQM_Parameter, как показано на рисунке Н.1.

Н.2.3 Реестр мер качества данных

Для управления мерами качества данных можно создать реестр мер качества данных. В этом случае реестр мер качества данных должен соответствовать спецификации реестра, приведенного в [5], который описывает структуру и атрибуты внесенных в реестр элементов.

На рисунке Н.2 представлена структура класса RE_RegisteredItem в сравнении с классами DQM_Measure, DQM_BasicMeasure и DQM_Parameter.

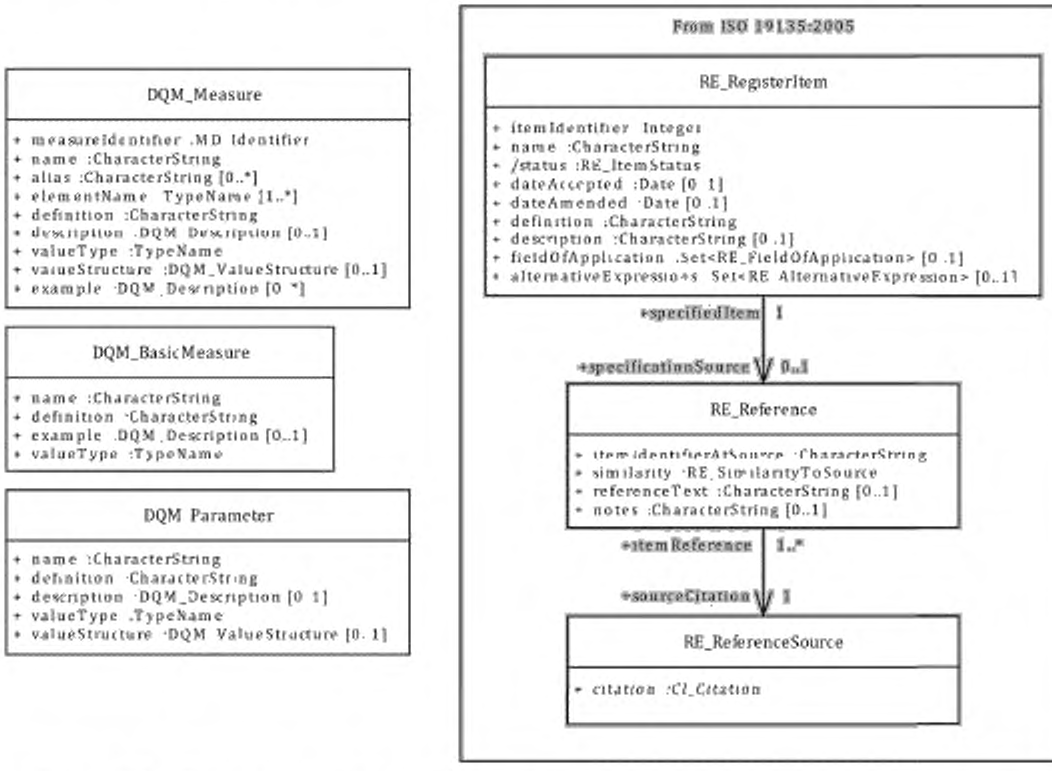


Рисунок Н.2 — Структурное сходство между внесенными в реестр элементами и мерами качества данных

Некоторые дескрипторы мер качества данных, базовых мер и параметров (как определено в разделе 8) могут быть повторно использованы в качестве атрибутов внесенных в реестр мер, базовых мер и параметров (см. рисунок Н.1 и таблицу Н.1) из RE_RegisteredItem, определенного в [5]. Иные дескрипторы внесенных в реестр элементов следует представлять в соответствии с [5].

Таблица Н.1 — Меры, базовые меры и атрибуты параметров, соответствующие атрибуту элемента

Элемент меры стандарта 19157	→	Элемент стандарта [5]
Зарегистрированная мера качества данных		
DQM_Measure.name	→	DQM_RegisteredDataQualityMeasure.name
DQM_Measure.definition		DQM_RegisteredDataQualityMeasure.definition
DQM_Measure.description.textDescription		DQM_RegisteredDataQualityMeasure.description
DQM_Measure.alias		DQM_RegisteredDataQualityMeasure.alternativeExpressions
DQM_Measure.measureIdentifier.code		DQM_RegisteredDataQualityMeasure.specifiedItem.itemIdAtSource
DQM_Measure.measureIdentifier.authority		DQM_RegisteredDataQualityMeasure.specifiedItem.sourceCitation

Окончание таблицы Н.1

Элемент меры стандарта 19157	»	Элемент стандарта [5]
Зарегистрированные базовые меры качества данных		
DQM_BasicMeasure.name	→	DQM_RegisteredDataQualityBasicMeasure.name
DQM_BasicMeasure.definition		DQM_RegisteredDataQualityBasicMeasure.definition
Зарегистрированные параметры качества данных		
DQM_Parameter.name	→	DQM_RegisteredDataQualityParameter.name

В таблице Н.2 дан пример зарегистрированной Меры 11 (см. таблицу D.11).

Таблица Н.2 — Пример зарегистрированного элемента для item — Мера 11

Зарегистрированный элемент для item	Пример значения
DQM_RegisteredDataQualityMeasure.itemIdentifier	Идентификатор item в пределах реестра Пример: «1»
DQM_RegisteredDataQualityMeasure.status	Статус item в пределах реестра
DQM_RegisteredDataQualityMeasure.name	«Количество некорректных перекрытий поверхности»
DQM_RegisteredDataQualityMeasure.definition	«Общее количество ошибочных перекрытий в данных»
DQM_RegisteredDataQualityMeasure.description	В зависимости от приложения одни поверхности могут перекрываться, а другие — нет. Не все перекрывающиеся поверхности обязательно ошибочны. При составлении отчета об этой мере качества данных нужно также представлять данные о классах объекта соответствующего некорректного перекрытия поверхности
DQM_RegisteredDataQualityMeasure.alternativeExpressions	«перекрывающиеся поверхности»
DQM_RegisteredDataQualityMeasure.specifiedItem.itemIdAtSource	«11»
DQM_RegisteredDataQualityMeasure.specifiedItem.sourceCitation	CI_Citation

Приложение I
(справочное)

Руководство по использованию элементов качества

I.1 Обзор

В некоторых случаях для одного конкретного требования к качеству может быть несколько возможных элементов качества, а в оценке качества обнаружена одна ошибка. Данное приложение содержит рекомендации по выбору элемента качества.

Примечание — Элементы качества описаны в 7.3.

I.2 Категории элементов качества данных

I.2.1 Общее представление

В разделе 7.3 определено шесть различных категорий элементов качества:

- полнота (7.3.2);
- логическая согласованность (7.3.3);
- позиционная точность (7.3.4);
- тематическая точность (7.3.5);
- временная точность (7.3.6);
- элемент применимости (7.3.7).

Элемент применимости используется для оценки качества в соответствии с потребностями пользователя, которые не могут быть охвачены остальными пятью категориями качества данных. Его также можно применять для предоставления результата агрегации, где обобщаются результаты из нескольких категорий качества данных (например, общее соответствие одной спецификации). Иные варианты применения в данном приложении не рассматриваются.

Из оставшихся пяти только логическая согласованность может быть оценена в полной мере без знания наземных данных. Установленные требования и оценка логической согласованности регулируют «внутренние отношения» в данных и контролируют соответствие данных правилам, установленным в спецификациях.

Три категории полноты, позиционной и тематической точности используются для описания того, насколько точно набор данных отображает предметную область.

Последняя категория (временная точность) состоит из комбинаций элементов качества данных, которые частично зависят от логических правил (в сравнении с логической согласованностью), и частично нуждается в данных наземного наблюдения (так же как категории полноты и точности).

I.2.2 Другие варианты

Все принятые элементы качества данных могут быть оценены, и результаты оценки не устареют по истечении времени. Два возможных (но не принятых) вида «Up-to-dateness» (актуальность) и «Timeless» (качество вне времени) описывают, насколько данные точно отвечают текущей реальной ситуации в мире.

При измерении актуальности (насколько набор данных отвечает современному реальному миру) результат будет действителен только короткое время. По истечении, например года, результат возможной хранимой меры актуальности будет неверным, т. е. будет отражать ситуацию в мире годовой давности, нежели текущее положение.

I.2.3 Последовательность при оценке качества данных

При оценке пространственных данных одна-единственная ошибка может повлиять на несколько элементов качества данных. Для вывода измерения в процентных показателях (например, процентные показатели аспектов полноты) важное значение представляет использование надлежащих знаменателей, которые описывают генеральную совокупность (см. рисунок I.1).

Оценка качества данных проводится в следующем порядке:

а) логическая согласованность/согласованность по формату: самое первое, что оценивается, — это читаемость (или интерпретируемость) данных с целью определения возможности их декодирования/чтения/понимания. Следует выводить информацию о неинтерпретируемых данных и игнорировать их при дальнейшей оценке. Результат согласованности по формату должен содержать информацию о том, какая часть данных не читается;

б) логическая согласованность: определить, применяются ли правила, установленные для набора данных. Части набора данных, не соответствующие правилам, должны игнорироваться при дальнейшей оценке;

в) полнота: следующий этап оценки — аспект существования объекта, охватываемый полнотой. Для ее оценки сравниваются объекты в фактическом наборе данных и данных наземного наблюдения, выводится информация об отсутствии и присутствии;

г) точность (позиционные, тематические и временные аспекты): последний этап оценки включает аспекты точности, которые измеряют отклонения между фактическими свойствами и свойствами объекта наземного наблюдения.

Эти измерения могут быть основаны только на тех фрагментах набора данных, которые представлены как в фактическом наборе данных, так и в предметной области.

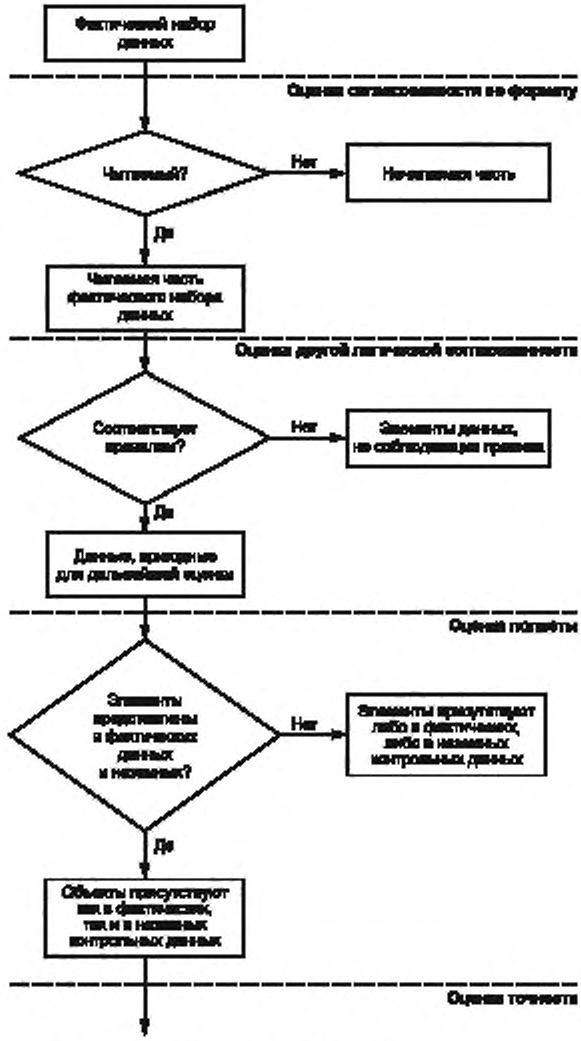


Рисунок I.1 — Порядок при оценке качества данных

I.3 Отношения между элементами качества данных

I.3.1 Общие положения

Многие элементы качества данных связаны между собой. В некоторых случаях это может привести к неопределенности относительно вывода информации о выявленных отклонениях/ошибок в данных. В данном разделе рассматриваются отношения между элементами качества данных.

I.3.2 Элементы качества данных, связанные с отсутствующими значениями атрибутов

По крайней мере три различных значения следует рассматривать как идентификатор «недоступности значения». Их способ применения может влиять на элементы качества данных, выбранные для представления информации об отсутствующем значении. Эти три значения имеют различную семантику:

- Пустое значение. В этом случае атрибут не имеет никакого значения.
- Неприемлемое значение означает, что для этого конкретного объекта атрибут неверен, т. е. не имеет никакого значения.

Пример 1 — Дата смерти живых людей.

- Неизвестное значение. В этом случае атрибут присутствует, т. е. значение должно быть, но оно неизвестно.

Информацию об обязательных атрибутах с пустыми значениями следует выдавать в виде последовательности логических ошибок. При оценке атрибутивной полноты не следует учитывать неприменимые обязательные атрибуты. Информацию о количестве неизвестных событий следует представлять в виде атрибутивной полноты.

Способом увеличения атрибутивной полноты является добавление искусственных значений к набору данных. Благодаря чему набор данных улучшится с точки зрения согласованности атрибутов, но при этом уменьшится атрибутивная точность.

Пример 2 — Набор данных содержит 50 экземпляров объекта типа дерево. 45 из них имеют хранящееся значение атрибута HeightOfTree. Точность этого атрибута (45 экземпляров) равняется ± 1 м (стандартное отклонение) и полнота атрибута 45/50, т. е. 90 %. Если однако эти отсутствующие HeightOfTree-значения дали неправильные (фиктивные) значения, например 10 м, то атрибутивная полнота увеличится (100 %), а атрибутивная точность, вероятно, уменьшится.

I.3.3 Отношения между различными аспектами точности

Отклонения фактических данных от предметной области могут быть измерены с помощью позиционной (временной) и атрибутивной (тематической) точности. Примерами альтернативных способов выражения отклонения являются:

- Атрибут в противоположность пространству: для атрибутов, где географическое распределение известно, отклонение может быть выражено тематическим или позиционным компонентами. Значение высоты контурной линии может рассматриваться как атрибут контурной линии. Отклонение текущего положения от истинного местоположения может быть измерено атрибутивным компонентом («полметра слишком высоко») или пространственным компонентом («контур линии имеет смещение 10 м в северном направлении»).
- Пространство в противоположность времени: если движение объекта известно, то разница между измеренным и реальным положением может быть выражена временным или позиционного компонентом: например, позиционная ошибка для движущегося по дороге автомобиля может быть выражена как «местоположение было верно 20 с назад» или «местоположение в настоящее время отличается на 400 м».
- Атрибут в противоположность времени: «Цена ($/м^2) конкретной посылки ошибочна на $20» или «эта цена была верной 10 лет назад».

I.3.4 Зависимость между полнотой и точностью

Оценка полноты обычно основана на сравнении набора данных и предметной области.

Критическая операция реализует связь между объектами в наборе данных и предметной областью. Уникальный идентификатор, как правило, формируется на основе их взаимосвязи.

При обработке объектов без такой идентификации элементов необходимо применять методы, основанные на близости атрибутов и их значений. При компоновке пространственных объектов следует учитывать два аспекта:

а) тематическую близость (обычно выражается в виде типа объекта);

б) географическую близость объектов.

Когда два объекта (один в наборе данных, а другой в наземных наблюдениях) приняты в качестве представления одного и того же реального явления, то отклонения между ними обрабатываются в виде точности. Если пара объектов взята для представления различных явлений, то информация об отклонении между ними выдается с использованием полноты (отсутствие и/или присутствие). Например, при оценке полноты и точности для 1 типа объекта (см. рисунок I.2) проблем с позицией A, B, C и D не возникает. При этом классификация идентична (тематическое отклонение равно нулю) и географические отклонения фактического и реального положения находятся в пределах допустимого уровня. Объекты связаны между собой, а отклонения описываются посредством позиционной точности. В положении E два экземпляра имеют различную тематическую классификацию, но расположены очень близко друг к другу. Необходимо принять решение о том, находится ли разница в классификации в пределах допустимого уровня для соединения. Если да, то два экземпляра будут способствовать точности оценки (позиционной и/или тематической), если нет, то дело в полноте (одна точка отсутствует и одна в избытке). В позиции F

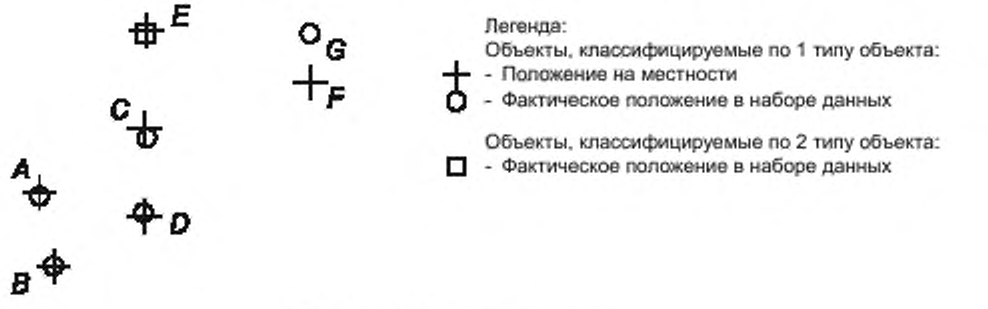


Рисунок I.2 — Точность и полнота

и G, два экземпляра имеют одинаковую классификацию, но отличаются положением. Если это географическое отклонение находится в пределах допустимого уровня для соединения, то отклонение будет способствовать позиционной точности (вероятно, отклоняющееся значение), если нет, то дело в полноте (отсутствие и присутствие).

I.4 Элементы качества данных — пример использования

I.4.1 Полнота

I.4.1.1 Общие положения

Для описания наличия и отсутствия объектов можно использовать элементы качества данных «отсутствие» и «присутствие». Полноту следует в основном применять на уровне типа объекта, который указывает на то, что объекты из предметной области обнаружены или не обнаружены в наборе данных.

Полнота может также иметь значение для свойств объекта («полнота атрибута» и «полнота отношений»). Перед использованием полноты в данных целях следует помнить о логической согласованности/концептуальной согласованности.

I.4.1.2 Присутствие — избыточность данных в наборе данных

Может применяться на уровне экземпляра объекта. Означает, что данные находятся в «избытке», если это целый экземпляр объекта. Если в экземпляре объекта или атрибуте экземпляра объекта есть необязательные данные, то о присутствии речь не идет.

Данное определение включает в себя экземпляры объектов, которые присутствуют в наборе данных, но которые не входят в область применения (согласно спецификации).

Правило для приведенных ниже примеров определяется так: «В набор данных должны включаться только те объекты, которые присутствуют в предметной области».

Пример 1 — Наличие данных из «Шотландия» подобно тем, что исключены из области определения набора данных («Англия»).

Пример 2 — Только здания с площадью больше 5 м^2 должны быть включены в набор данных. Информация о наличии зданий до 5 м^2 представляется в виде избыточности.

I.4.1.3 Отсутствие — данные отсутствуют в наборе данных

Так же как и присутствие, может применяться на уровне экземпляра объекта. На практике это означает отсутствие экземпляров объектов, включение которых указано в спецификации.

Отсутствие следует в основном использовать, когда «целый элемент», например экземпляр объекта отсутствует. Если обязательная часть элемента, например обязательный атрибут экземпляра объекта отсутствует, то информацию следует выдавать в виде ошибки концептуальной согласованности.

Правило для приведенного ниже примера определяется следующим образом: «Вся жилая недвижимость Англии и Уэльса должна быть включена в набор данных».

Пример — Отсутствие жилой недвижимости Англии или Уэльса в наборе данных.

I.4.2 Логическая согласованность

I.4.2.1 Общие положения

Степень соответствия логическим правилам структуры данных, распределения и отношений (структура данных может быть концептуальной, логической или физической) может быть описана посредством следующих элементов качества данных.

I.4.2.2 Концептуальная согласованность — соответствие правилам концептуальной схемы

Как правило, приложения имеют концептуальную схему, описывающую требования к структуре данных. Эта концептуальная схема может включать в себя:

- имена всех классов (типов объектов, типов данных и т. д.);
- имена атрибутов для всех классов, а также ограничения множественности;
- домены для всех атрибутов;
- отношения между классами;
- топологические отношения между типами объектов, например отношение между некоей областью и границей;
- отношения между атрибутами типа объекта для различных типов объектов, например отношение между значением высоты над уровнем моря от контурной линии и от дороги в географической точке пересечения для двух экземпляров объекта.

Концептуальная последовательность может охватывать все эти аспекты качества данных.

Другие подэлементы логической согласованности (доменная согласованность, топологическая согласованность) также могут рассматриваться для некоторых аспектов, перечисленных выше, если концептуальная согласованность используется только для обеспечения корректных свойств объекта для каждого экземпляра объекта.

I.4.2.3 Доменная согласованность — соответствие значений атрибутов области допустимых значений

Для описания области допустимых значений, как правило, используется концептуальная схема приложения, при этом информация о ней выдается как часть концептуальной или доменной согласованности. Если в концеп-

туальной схеме не существуют или не действуют доменные характеристики, то тогда можно использовать только подэлемент качества доменная согласованность.

Пример 1 — Организация определяет правильную область значений для каждого поля в терминах длины, типа данных и содержания. Доменная согласованность используется для обеспечения соблюдения этих условий со следующими исключениями:

- если поле содержит данные о местоположении (т. е. восточное и северное), то рассматривается как позиционная точность;
- если поле содержит данные о дате/времени, то рассматривается как временное качество;
- если поле содержит первичный ключ, в этом случае рассматривается в рамках логической согласованности.

Правило для приведенного ниже примера определяется следующим образом: поле LANGUAGE должно содержать «ENG» или «CYM».

Пример 2 — Пример ошибки доменной согласованности: «COR».

I.4.2.4 Согласованность по формату — степень соответствия накопленных данных физической структуре набора данных

Согласованность по формату следует в основном применять в качестве первой проверки оценки качества для удостоверения того, что набор данных сформирован в правильном формате в соответствии со спецификацией (продукта).

Если установлены определенные правила для определения формата определенных атрибутов, например для сгенерированных идентификаторов, то согласованность по формату также может иметь отношение к отдельным значениям атрибута. Если атрибуты значения проверяются в сравнении со списком допустимых значений (домен), то следует использовать доменную согласованность.

Пример 1 — Спецификация информационного продукта определяет GML как формат распространения. Если набор данных не является GML-файлом, то эту информацию об ошибке следует выдавать в виде ошибки согласованности по формату. Если один элемент в GML-файле отображается «в неправильном формате», например текст вместо числа, то информацию следует выдавать в виде ошибки концептуальной согласованности или ошибки доменной согласованности.

Пример 2 — В пределах организации данная классификация используется для описания проверок, которые обеспечивают соответствие правил спецификации информационного продукта и включает в себя:

- наличие, достоверность и уникальность значений первичного ключа. Пример правила: каждый экземпляр объекта должен иметь уникальный идентификатор. Пример ошибки согласованности по формату — «NULL»;

- внешние ключи, которые соотносят идентификатор для другого экземпляра объекта, не присутствуют в наборе данных. Пример правила — поле PARENT_UPRN должно содержать идентификатор, связанный с существующим экземпляром объекта UPRN.

I.4.2.5 Топологическая согласованность — корректность явно закодированных топологических характеристик набора данных

Топологические характеристики набора данных описывают геометрические отношения между элементами набора данных, которые не изменяются в результате трансформаций по методу «резинового листа». Предполагается, что основные части топологических ограничений описываются в концептуальной схеме, при этом информация может выводиться в виде концептуальной или топологической согласованности. Только топологическая согласованность может применяться в случае, если соответствующие топологические требования не являются частью концептуальной схемы.

Пример 1 — Для набора данных с типами объектов, определенных по береговой линии водных объектов (такие типы объектов, как побережье, гавань, эллинг), а также с типами объектов водоемов (озера, моря и т. д.). Топологические отношения между типами объектов четко определены в концептуальной схеме, и для выдачи информации о том, что геометрия прибрежных линий (одномерная) совпадает с геометрией водоемов (двумерная) используется подэлемент качества концептуальная согласованность.

Пример 2 — В сетевом наборе данных с неопределенными требованиями в концептуальной схеме для «чистой сети» информацию о «загрязненных частях» («недоход», «переход», перекрытие, самопересечение и т. д.) следует представлять как ошибки топологической согласованности.

I.4.3 Точность позиционирования

Точность положения объектов по отношению к Земле может быть описана с использованием элементов качества данных этого раздела.

Измерение позиционной точности посредством наземного наблюдения подразумевает создание «согласованных пар» с одним экземпляром объекта из набора данных и соответствующим в контрольном (наземное наблю-

дение) наборе данных. Если объекты имеют уникальные идентификаторы (например, для кадастровых участков), то это соответствие может быть установлено с помощью идентификаторов, при этом грубые ошибки, смещения, стандартное отклонение могут быть оценены и отражены в виде позиционной точности.

При отсутствии доступных идентификаторов соответствие следует устанавливать на основе позиции. Должен быть определен «предел расстояния для соответствия», который облегчает вычисление грубых ошибок. Данный «предел расстояния для соответствия» должен быть задокументирован в отчете в следующих случаях:

- экземпляры объектов в наборе данных без соответствующих экземпляров объектов контрольного набора данных должны быть указаны как избыточные в отношении элемента полноты;
- экземпляры объектов в контрольном наборе данных, не имеющие соответствующих экземпляров объектов в наборе данных, должны быть указаны в отчете как отсутствующие в отношении элемента полноты.

I.4.4 Временная точность

I.4.4.1 Общие положения

Точность временных атрибутов и временных отношений объектов может быть описана с помощью следующих параметров качества данных.

I.4.4.2 Точность измерения времени — соответствие заявленных измерений времени значениям, принятым или считающимся правильными.

Пример — В пределах определенной организации точность измерения времени применяется для удостоверения в том, что:

- значение не противоречит конкретному условию в поле (дополнительным условиям, обусловленным характером данных дата/время).

Пример правила — поле START_DATE не может содержать значение в будущем.

I.4.4.3 Временная согласованность — правильность порядка событий

Правила, описывающие аспект «правильность порядка последовательности событий», могут являться частью концептуальной схемы. Информация выдается в виде временной или концептуальной согласованности, если правила являются частью концептуальной схемы.

Пример — В рамках определенной организации временная последовательность используется для:

- подтверждения соответствия между значениями даты/времени, связанных с жизненным циклом объекта реального мира;

- обеспечения согласованности значений даты/времени, используемых при управлении экземплярами объекта в наборе данных.

Пример правила — END_DATE должна быть такой же или быть более поздней, чем START_DATE.

Пример ошибки временной согласованности — START_DATE = "2010-02-02", END_DATE = "2000-01-01".

I.4.4.4 Временная достоверность — достоверность данных по отношению ко времени

Правила, описывающие аспект «достоверность данных по отношению ко времени», могут являться частью концептуальной схемы.

Информация выдается как временная достоверность либо как концептуальная согласованность, если правила являются частью концептуальной схемы.

Пример — В рамках определенной организации точность измерения времени используется для:

- удостоверения в том, что содержимое поля даты или времени находится в правильном формате и использует календарь, определенный в спецификации.

Пример правила — значение даты должно быть указано в формате [15] — "ВВГГ-ММ-ДД".

Пример ошибки временной достоверности — "01.01.2010" или "2010-51-15".

I.4.5 Тематическая точность

I.4.5.1 Общие положения

Точность количественных атрибутов и правильность неколичественных атрибутов и классификаций объектов и их отношений могут быть описаны с помощью следующих элементов качества данных.

I.4.5.2 Правильность классификации — соответствие классификации объектов или их атрибутов предметной области (т. е. реальной местности или эталонному набору данных)

Пример — Данная характеристика используется строго в рамках одной организации. Классификации, которые не определены в спецификации набора данных, не рассматриваются в качестве правильности классификации (а только как доменная согласованность).

I.5 Рассмотрение особых случаев

I.5.1 Отношение между неправильной классификацией и полнотой на уровне типа объекта

На уровне типа объекта полнота и тематическая точность/правильность классификации тесно связаны друг с другом. Действительно при неправильной классификации одного экземпляра объекта (отнесении его к другому

типу объекта) ошибка классификации возникнет в оценке полноты для обоих типов объектов (одна — присутствия и одна — отсутствия).

Поэтому рекомендуется при оценке полноты на уровне объекта знать, что некоторая ошибка в присутствии или отсутствии может быть результатом неправильной классификации. При представлении информации о правильности классификации информация об ошибке будет выдана дважды.

Чтобы избежать двойной выдачи сообщения об ошибках, можно вывести отчет о полноте в одном верхнем уровне (набор данных, группировка типов объектов и т. д.), а информацию о неправильной классификации — на уровне объектов.

В приложении Е приведен пример.

I.5.2 Элементы качества, связанные с уникальными идентификаторами

Ниже представлены некоторые варианты использования, имеющие отношение к соответствующим элементам качества данных, для описания проблем, связанных с уникальными идентификаторами (см. таблицу I.1).

Таблица I.1 — Элементы качества, связанные с уникальными идентификаторами

Вариант использования	Применимый элемент качества данных
Все уникальные идентификаторы должны иметь формат, который отвечает правилам их определения	Согласованность по формату
Все используемые уникальные идентификаторы допустимы согласно списку зарезервированных уникальных идентификаторов	Доменная согласованность
Один и тот же экземпляр объекта дважды присутствует с таким же уникальным идентификатором	Полнота концептуальной согласованности (уникальные идентификаторы должны быть уникальными)
Один и тот же экземпляр объекта дважды присутствует с различными уникальными идентификаторами. Примечание — Задача заключается в необходимости удостоверения в том, что эти два экземпляра объекта действительно представляют один и тот же объект реального мира	Избыточность

Приложение J
(справочное)

Обобщение результатов оценки качества

J.1 Введение

Оценка на основе одного элемента качества данных, как правило, недостаточно удовлетворяет запросам пользователя.

Разработчик данных обычно разрабатывает (в сотрудничестве с потенциальными пользователями продукта) спецификацию информационного продукта с учетом всех требований, установленных для данного продукта.

Для потенциального пользователя представит интерес заключение, в котором говорится о том, что продукт оценивается на основе спецификации. Таким заключением является обобщенный показатель качества данных, который может быть полезен в других ситуациях, а не только при отчете соответствия спецификации.

Качество набора данных может быть представлено одним или более обобщенным показателем качества данных (ADQR). ADQR включает результаты оценки качества данных на основе различных элементов качества данных или различных областей определения качества данных.

Примеры методов, которые используются для создания ADQR, приведены в J.2—J.4. Следует учитывать, что при таком способе оценки качества набор данных может получить общую положительную оценку, даже если один или более показателей качества данных не прошли проверку. Поэтому обобщение необходимо применять только при наличии веских причин. В любом случае смысл обобщающего оценки качества данных должен быть всегда четко определен.

Так как ADQR может вызвать затруднения при определении, то смысл обобщающей оценки качества данных следует определить еще до извлечения выводов на основе обобщенного показателя качества данных для обеспечения качества набора данных.

Описание обобщенного показателя качества дано в 10.2.1.

J.2 Однозначная оценка пригодности/непригодности

Каждому показателю качества данных, вовлеченному в вычисления, придается логическое значение, равное единице (1), если значение показателя соответствует требованиям, и нулю (0), если нет. Обобщенный показатель качества определяется уравнением

$$ADQR = v_1 \cdot v_2 \cdot v_3 \cdot \ldots \cdot v_n,$$

где n — число групп определения качества данных.

Если ADQR = 1, то общее качество набора данных считается полностью соответствующим требованиям, а значит, пригодно. Если ADQR = 0, то качество считается не соответствующим требованиям, а значит, непригодно. Данный метод не обеспечивает результат, который показывает местоположение или величину несоответствия.

J.3 Взвешенная оценка пригодности/непригодности

Каждому показателю качества данных, вовлеченному в вычисление, придается логическое значение, равное единице (1), если значение показателя соответствует требованиям, и нулю (0), если нет. Кроме того, на основании значимости показателя для оценки качества в целом каждому из них присваивается весовое значение в интервале от 0 и 1 включительно. Сумма всех весов должна равняться 1. Выбор весов является субъективным решением, принимаемым разработчиком данных или пользователем данных. Причину своего решения разработчику данных следует указывать в виде части результата. Обобщенное качество определяется уравнением

$$ADQR = v_1 \cdot w_1 + v_2 \cdot w_2 + v_3 \cdot w_3 + \ldots + v_n \cdot w_n,$$

где n — число групп определения качества данных.

Данный метод обеспечивает получение значения величины, показывающей, насколько набор данных близок к полному соответствию. Метод не обеспечивает получение количественного значения величины, указывающее, где имеет место соответствие или несоответствие.

Пример — Таблица ошибок (см. таблицу J.1) показывает количество обнаруженных ошибок и их классифицирование в соответствии со стандартной процедурой, используемой для баз данных дорог. В этом конкретном примере каждому типу ошибки присвоены весовые значения. Сумма весов равна 100 %. Полученное взвешенное значение представляет качество набора данных.

Таблица J.1 — Пример вычисления обобщенных показателей оценки качества

Объект	Количество элементов в партии	Количество несоответствующих элементов	Степень несоответствия	Соотношение точности (принятое за единицу)	Вес	Весовое значение (соотношение точности · вес)
Сегмент дороги	19					
Некорректный		1				
Отсутствующий		0	4 / 19	0,79	50 %	0,3950
Избыточный		3				
Название улицы						
Базовое	19	5	5 / 19	0,74	15 %	0,1110
Направление движения	19	1	1 / 19	0,95	25 %	0,2375
Гидрография	1	0	0 / 1	1,00	10 %	0,1000
Общая точность	(определена как сумма весовых значений · 100)					84,35 %

Примечания

1 Элемент определяется в виде сегмента дороги, ограниченного точками пересечения с другими дорогами или границами элемента выборки.

2 Обобщение информации по оценке качества данных, особенно при использовании весов, не имеет большого значения для конечных пользователей и может привести к неправильным выводам в зависимости от того, какие весы применял разработчик данных.

J.4 Оценка данных на основе максимальных/минимальных значений показателя качества данных

Каждому показателю качества данных придается значение, основанное на значимости данного показателя для предназначения продукта. Основанием для подобного решения разработчика данных должны являться результаты оценки качества отдельных групп данных. Обобщенный показатель качества определяется одним из двух уравнений:

$$ADQR = \text{MAX}\ (v_i,\ i_n = 1\ \ldots\ n) \text{ или } ADQR = \text{MIN}\ (v_i,\ i_n = 1\ \ldots\ n),$$

где n — число определений качества данных.

Данный метод обеспечивает получение значения величины, указывающей, насколько набор данных близок к полному соответствию, но только в терминах определения качества набора данных на основе максимума или минимума. Также метод задает количественное значение, указывающее, где имеет место наибольшее соответствие или несоответствие, когда заданное определение качества данных выдается вместе с ADQR. Однако данный тип ADQR не сообщает о ситуации с другими показателями качества данных.

Приложение ДА
(справочное)

Сопоставление структуры настоящего стандарта со структурой примененного в нем международного стандарта

Таблица ДА.1

Структура настоящего стандарта	Структура международного стандарта ИСО 19157
Приложение ДА	—
Приложение ДБ	—
Примечание — Сопоставление структуры стандартов приведено, начиная с приложения ДА, так как предыдущие разделы стандартов и их структурные элементы (за исключением предисловия) идентичны.	

Приложение ДБ
(справочное)

Сведения о соответствии ссылочных национальных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте

Таблица ДБ.1

Обозначение ссылочного национального стандарта	Степень соответствия	Обозначение и наименование ссылочного международного стандарта
ГОСТ Р 57668—2017	MOD	ISO 19115-1:2014 «Географическая информация. Метаданные. Часть 1. Основные положения»
ГОСТ Р 57656—2017	MOD	ISO 19115-2:2009 «Географическая информация. Метаданные. Часть 2. Расширения для изображений и матричных данных»
ГОСТ Р 57657—2017	MOD	ISO 19131—2007 «Географическая информация. Спецификация информационного продукта»
Примечание — В настоящей таблице использовано следующее условное обозначение степени соответствия стандартов: - MOD — модифицированные стандарты.		

Библиография

[1] ISO 19109:2005, Geographic information — Rules for application schema

[2] ISO 19108, Geographic information — Temporal schema

[3] ISO 19103:2015, Geographic information — Conceptual schema language

[4] ISO/TS 19139:2007, Geographic information — Metadata — XML schema implementation

[5] 19135:2005, Geographic information — Procedures for item registration

[6] ISO 19110:2016, Geographic information — Methodology for feature cataloguing

[7] ISO 19107:2003, Geographic information — Spatial schema

[8] ISO/IEC 23824:1999, Information technology — Vocabulary — Part 4: Organization of data

[9] ISO/TS 19129:2009, Geographic information — Imagery, gridded and coverage data framework

[10] ISO 19123:2005, Geographic information — Schema for coverage geometry and functions

[11] Environmental Systems Research Institute, Inc (ERSI) GSI Data ReViewer 4.2 User Guide

[12] CRC Handbook of Tables for Probability and Statistics. Second Edition, 1982

[13] ISO 19156:2011, Geographic information — Observations and measurements

[14] ISO 2859, Sampling procedures for inspection by attributes

[15] ISO 8601:2004, Data elements and interchange formats — Information interchange — Representation of dates and times

УДК 622.1:528:002:006.354 ОКС 35.240.70

Ключевые слова: оценка качества данных, элементы качества данных, меры качества, пространственные данные, метаданные, качество данных, поставка данных, формат данных

БЗ 11—2017/37

Редактор *Н.А. Аргунова*
Технический редактор *В.Н. Прусакова*
Корректор *Е.Р. Ароян*
Компьютерная верстка *Л.В. Софейчук*

Сдано в набор 09.10.2017 Подписано в печать 22.11.2017 Формат 60×84 1/8 Гарнитура Ариал.
Усл. печ. л. 15,81. Уч.-изд. л. 14,31. Тираж 25 экз. Зак. 2325.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

ИД «Юриспруденция», 115419, Москва, ул. Орджоникидзе, 11
gost.e-ecolog.ru gost.e-ecolog.ru

Издано и отпечатано во ФГУП «СТАНДАРТИНФОРМ», 123001, Москва, Гранатный пер., 4.
gost.e-ecolog.ru gost.e-ecolog.ru

gost.e-ecolog.ru