УДК 004.9

# МЕТОДИКА СКОРИНГА ИСТОЧНИКОВ И ИНЦИДЕНТОВ В МНОГОУРОВНЕВЫХ СИСТЕМАХ КАЧЕСТВА ДАННЫХ

**Лопатин И.Н.,** аспирант, Российский институт стандартизации, г. Москва

*Цель работы: разработка методики скоринга источников в многоуровневой системе качества данных для оценки состояния критичных источников информационной безопасности, в частности SOC (Security operation center), в антифрод-системах, системах искусственного интеллекта, использующих SLM (Small Language Models – малые языковые модели).*

*Методы: аккумуляции знаний научных публикаций в области алгоритмов оценки качества данных, технологий работы корреляционных ядер, механизмов скоринга весов в антифрод-системах; обобщены методики и алгоритмы систем качества данных.*

*Результаты: показано, что предложенная методика оценки качества данных в многоуровневой системе мониторинга обеспечивает снижение влияния низкого качества данных на критичные системы путем своевременного информирования об отклонениях в данных; проведен анализ алгоритмов расчета качества данных в международных системах оценки данных и показаны преимущества представленного подхода в работе.*

**Ключевые слова:** качество данных, кибербезопасность, антифрод-системы, SLM, мультиагентные системы, многоуровневый мониторинг, алгоритмы качества данных, методики оценки качества данных.

**Цитирование:** Лопатин И.Н. Методика скоринга источников инцидентов в многоуровневых системах качества данных // Информационно-экономические аспекты стандартизации и технического регулирования. 2025. № 4(85). С. 101–107.

## ВВЕДЕНИЕ

В современных системах информационной безопасности (Security Operations Center, SOC) [1] и антифрод-платформах эффективность выявления угроз и мошеннических действий напрямую зависит от качества данных. Ошибки или неполнота информации зачастую приводят к ложным сработкам систем, избыточной нагрузке на операторов и пропущенным инцидентам. Также недостаточный контроль данных в других сферах, где критичность информации играет решающую роль, например многоагентные системы искусственного интеллекта[1] (СИИ), где принятие интеллектуальным агентом [2] решений, основанных на системе знаний с низким уровнем качества данных, может привести к неправильным действиям агентов внутри организации.

Обычно подход к оценке качества ограничивается одним уровнем анализа, например, проверкой корректности форматов в базе данных. Однако практика показывает, что проблемы могут появляться как на инфраструктурном уровне: поток данных, состояние топиков, непрерывность логирования, так и на уровне содержательного анализа: достоверность полей, аномалии в поведении и другие [3].

В документах, посвященных качеству данных, например ISO 8000 [4], DAMA DMBOK [5], подчеркивается необходимость структурированного контроля, однако отсутствует точная информация об алгоритмах скоринга, пригодных для источников SOC и антифрод.

Задача представленной работы – объединить подходы из сферы управления качеством данных с системами: информационной безопасности, антифрод-систем, СИИ, ввести многоуровневые алгоритмы расчета и показать их эффективность.

Вышеизложенные задачи можно декомпозировать:

- провести анализ существующих на рынке решений и подходов и сравнить между собой;
- проанализировать особенности многоуровневой методики при мониторинге данных на уровнях: инфраструктурном, статистическом, базовом, поведенческом и синтетическом;
- разработать математический аппарат скоринга (формулы, алгоритмы) с учетом критичности источников и проверок;
- предложить новый коэффициент учета уровня мониторинга, позволяющий корректировать итоговую оценку в зависимости от уровня обнаружения проблем;
- продемонстрировать практические примеры расчетов, подтверждающие эффективность метода для снижения

[1] ГОСТ Р 59277–2020 Системы искусственного интеллекта. Классификация систем искусственного интеллекта. – М.: Стандартинформ, 2021. (п. 3.31).

ложных сработок и улучшения показателей работы SOC и антифрод-систем;

- продемонстрировать применимость данной системы для оценки данных для обучения SLM моделей, а также повышения эффективности интеллектуальных агентов за счет повышения качества данных.

Основой методики служит декомпозиция мониторинга по уровням: инфраструктура, статистика, базовое качество, поведенческий, синтетический анализ.

Цель настоящего исследования – сформировать научно обоснованный алгоритмический аппарат, позволяющий рассчитывать итоговый скоринг качества данных с учетом:

- критичности проверок;
- критичности источников (вес источников);
- критичности данных;
- многоуровневости контроля (от инфраструктурного до синтетического).

## 1. АНАЛИЗ СУЩЕСТВУЮЩИХ ПОДХОДОВ

Для получения полной картины эффективности работы алгоритмов и подходов качества данных необходимо провести сравнительный анализ по базовым критериям аналогичных продуктов.

Ниже представлена таблица 1 анализа источников и сравнения подходов многоуровневой системы качества данных, описанной в данной статье (в составе продукта «Мониторинг X360»), с лидерами ранка оценки качества данных, таких как: IBM, Oracle, SAP, Informatica.

Таблица 1

**Сравнительная таблица систем качества данных**

| КРИТЕРИИ | INFORMATICA DATA QUALITY (IDQ) [7] | IBM INFOSPHERE INFORMATION SERVER [6] | SAP INFORMATION STEWARD [8] | ORACLE ENTERPRISE DATA QUALITY (EDQ) [9] | МОНИТОРИНГ X360 (МНОГОУРОВНЕВАЯ СИСТЕМА КАЧЕСТВА ДАННЫХ) [3] |
|---|---|---|---|---|---|
| ОСНОВНАЯ НАПРАВЛЕННОСТЬ | Профилирование, очистка и стандартизация данных | Профилирование, очистка и интеграция данных | Управление качеством и метаданными | Очистка и стандартизация данных | Многоурвневый скоринг, управление качеством данных, мониторинг данных, очистка, стандартизация |
| НАЛИЧИЕ ФОРМАЛИЗОВАННЫХ АЛГОРИТМОВ И ФОРМУЛ | Частично формализованы | Частично формализованы | Частично формализованы | Средний уровень формализации | Полностью формализованы и детализированы |
| КОЛИЧЕСТВО УРОВНЕЙ МОНИТОРИНГА | Отсутствуют явно выраженные уровни | Отсутствуют явно выраженные уровни | Отсутствуют явно выраженные уровни | 2 уровня: базовый и расширенный | 5 уровней: инфраструктурный, статистический, базовый, поведенческий, синтетический |
| ВЕСОВЫЕ КОЭФФИЦИЕНТЫ ИСТОЧНИКОВ И ПРОВЕРОК | Частично используются | Частично используются | Ограниченно используются | Частично используются | Высоко используются. Полностью формализованы и детализированы |
| УДОБСТВО ПРАКТИЧЕСКОГО ВНЕДРЕНИЯ | Высокое при интеграции с Informatica | Высокое при интеграции с IBM | Высокое при интеграции с SAP | Высокое при интеграции с Oracle | Высокое, благодаря универсальным адаптерам стека Apache |
| ПОДХОД К ОЦЕНКЕ РИСКОВ И КРИТИЧНОСТИ | Частично реализован | Частично реализован | Ограниченно реализован | Ограниченно реализован | Полностью формализован с использованием скоринга и коэффициентов |
| ПРИМЕНИМОСТЬ В SOC И АНТИФРОД | Средняя | Средняя | Низкая | Средняя | Высокая, специально разработана |
| ПРИМЕНИМОСТЬ К АНАЛИЗУ ДАННЫХ SLM МОДЕЛЕЙ | Низкая | Средняя | Низкая | Средняя | Высокая, предназначена для анализа данных SLM Средний |
| УРОВЕНЬ ДЕТАЛИЗАЦИИ МЕТОДИКИ | Средний | Средний | Средний | Средний | Высокий, детализированные алгоритмы и примеры |

**Сравнение существующих методологических подходов DAMA DMBOK и ISO 8000.** В фундаментальных стандартах управления качеством данных, таких как DAMA DMBOK (Data Management Body of Knowledge) и ISO 8000, представлен преимущественно методологический взгляд на проблему. Эти документы задают общие рамки и направления работы с качеством данных, однако не предлагают конкретных алгоритмов оценки и расчета интегральных показателей.

Например, в стандарте ISO 8000 сформулированы принципы оценки качества данных, такие как точность (accuracy), полнота (completeness), актуальность (timeliness) и согласованность (consistency). Однако стандарт не раскрывает механизм количественной оценки этих критериев и не предлагает практических формул расчета интегрального показателя, а ограничивается лишь общей рамочной структурой и качественными индикаторами (ISO 8000-61:2016). Аналогично DAMA DMBOK представляет структурированную модель управления данными, акцентируя внимание на процессах Data Governance и Data Quality Management. В руководстве DAMA подчеркивается необходимость системного подхода, регулярного мониторинга и управления качеством данных, однако отсутствуют примеры алгоритмов скоринга или формул расчета интегральных показателей, которые могли бы быть непосредственно применены в системах информационной безопасности или антифрод-платформах [3].

Таким образом, методологические подходы, несмотря на их фундаментальность и популярность, зачастую требуют значительной доработки и конкретизации для использования в практических задачах кибербезопасности и антифрод-аналитики.

## 2. ОСОБЕННОСТИ КОНЦЕПЦИИ МНОГОУРОВНЕВОЙ МЕТОДИКИ

Предлагаемая в данной статье методика и алгоритмы как раз и призваны объединить достоинства подходов DAMA DMBOK и ISO 8000, добавив новый многоуровневый коэффициент и новую систему распределения весов критичностей. Основой методики является многоуровневый мониторинг, включающий уровни:

- ***Инфраструктурный*** *(проверка доступности и целостности передачи данных)*
- ***Статистический*** *(контроль статистических свойств данных)*
- ***Базовое качество данных*** *(точность, полнота, форматы)*
- ***Поведенческий анализ*** *(аномалии и отклонения)*
- ***Синтетический анализ*** *(проверки данных по бизнес-логике)*

Каждому уровню мониторинга и источнику присваивается специальный весовой коэффициент, который позволяет получать интегральный скоринг качества данных. Например, ошибки, выявленные только на последнем синтетическом уровне, получают существенно более высокий вес, чем ошибки, выявленные на ранних этапах, что отражает их большую потенциальную опасность и влияние на бизнес-процессы, правила корреляции и другие сущности.

### 2.1. Математический аппарат скоринга с учетом критичности источников

В данной статье рассматриваются четыре ключевые формулы (S_base, S_type, S_source, S_dataset), а также дополнительный множитель W_level. Ниже приведено их детальное описание.

**Базовая формула доли ошибок (S_base)**

Данная формула является классической метрикой контроля качества (Quality Control Ratio) и отражает, насколько часто данные содержат ошибки на проверяемом уровне. Базовая формула расчета доли ошибочных записей в проверяемом наборе данных выглядит следующим образом:

$$S_{base} = \frac{C_{fail}}{C_{total}}, \quad (1)$$

где: $C_{fail}$ – число проблемных проверок, $C_{total}$ – общее число проверок.

**Скоринг по типу проверки (S_type)**

Проверки в многоуровневом мониторинге могут обладать различной важностью: например, критичная проверка, нарушение формата ключевых полей, без которых невозможно обнаружить инцидент и базовая проверка – общая корректность форматов.

Вводится весовой коэффициент, отражающий уровень критичности:

$$S_{type} = S_{base} * W_{type}, \quad (2)$$

где:

- $W_{type} = 1$ – задается экспертно, в зависимости от потенциального влияния ошибок;
- критичные проверки (могут привести к серьезным инцидентам) имеют $W_{type} = 1 = 10$;
- среднекритичные проверки (ограниченное влияние на процессы) $W_{type} = 1 = 5$;
- базовые проверки (малое влияние на результат) $W_{type} = 1$.

Обоснование выбора весовых коэффициентов – весовые коэффициенты были выбраны на основе экспертного анализа влияния ошибок на бизнес-процессы и операционные риски в SOC и антифрод-системах. В системе МониторингX360 гибко представлена возможность настройки под потребности коэффициентов, но для расчета взяты следующие:

1. Критичные (коэффициент = 10).
2. Среднекритичные (коэффициент = 5).
3. Базовые (коэффициент = 1).

Чем выше S_type, тем больше проблем обнаружено с учетом важности (веса) самой проверки.

**Средневзвешенный скоринг источника (S_source)**

Каждый источник проходит разные категории проверок. Для агрегирования результатов вводится следующий алгоритм:

$$S_{source} = \left(\frac{\sum S_{type}}{N_{types}}\right) * W_{source}\ , \qquad (3)$$

где:

- $\sum S_{type}$ – сумма скорингов по типам проверок (критичной, среднекритичной и т. д.);
- $N_{type}$ – общее число категорий проверок (типов проверок);
- $W_{source}$ – вес критичности (важности) самого источника.

Выражение (3) для общего балла конкретного источника данных берет среднее значение скоринга по всем типам проверок для этого источника и умножает его на вес W_source источника. Усреднение в (3) $\sum S_{type}/N_{types}$ означает, что все виды проверок учитываются равномерно. Последующее умножение на W_source вводит важность самого источника.

Применение многоуровневого подхода в системе скоринга источника на этапе агрегирования источника:

$$S_{source} = \left(\frac{\sum S_{type}}{N_{types}}\right) * W_{source} * W_l, \qquad (4)$$

где:

- $\sum S_{type}$ – сумма скорингов по типам проверок (критичной, среднекритичной и т. д.);
- $N_{type}$ – общее число категорий проверок (типов проверок);
- $W_{source}$ – вес критичности (важности) самого источника.
- $W_l$ – вес уровня мониторинга, который проходит источник при анализе.

Если источник участвует в нескольких уровнях, например, данные источника проверяются на M2 и на M3, применяется комбинированный коэффициент.

## 2.2. Оценка качества набора данных SLM моделей (S_dataset)

Агрегирование нескольких источников в общий индикатор качества, используется формула:

$$S_{dataset} = \frac{\sum (Q_i * W_i)}{\sum W_i}, \qquad (5)$$

где $Q_i$ – скоринговая оценка $i$-го источника; $W_i$ – вес $i$-го источника, зависящий, например, от объема данных или уровня.

Данная формула представляет собой классическое взвешенное среднее показателей качества $Q_i$ с весами $W_i$. Формула взвешенного среднего используется как: каждому критерию «$i$» присваивается вес $W_i$ – отражающий его значимость, после чего вычисляется итоговый индекс качества. В документации IBM по мониторингу данных прямо указано, что общий балл качества рассчитывается как взвешенное среднее баллов всех проверок или показателей. Фактически, S_dataset совпадает с вычислением интегрального Data Quality Score на основе нескольких измерений качества.

Практический пример при оценке набора данных для моделей ИИ могут учитываться доля пропусков, уровень шума, сбалансированность классов и т.п.; присвоив каждому из этих показателей вес, аналитик вычисляет сводный показатель качества датасета по формуле средневзвешенного. Таким образом, S_dataset соответствует принципам многомерной оценки качества данных, широко описанным в литературе по управлению качеством данных (DQM). Отметим, что взвешенное суммирование индексов применяется не только в качества данных, но и например при расчете «score» модели машинного обучения по нескольким метрикам, или при агрегировании показателей данных в различных фазах их жизненного цикла.

Так для многоагентных (мультиагентных) систем использующих ai-агенты, контроль используемых данных в системах знаний имеет критичную значимость. В случае отсутствия многоуровневого подхода оценки качества данных, в системах знаний агентов происходит неправильное формирование итоговых действий агентов за счет неточности в векторизации данных в RAG (Retrieval-Augmented Generation). Многоуровневый алгоритмический подход к оценке данных значительно повышает точность и уровень доверия к ai-агентам в организации.

При этом для моделей ИИ может быть важно не только усреднить показатели, но и исключить из датасета те сег-

менты, у которых скоринг существенно выходит за рамки допустимого.

**Коэффициент многоуровневости (W_level)**

Согласно идее многоуровневого мониторинга (инфраструктурный, статистический, базовое качество, поведенческий и синтетический уровни), проблемы, не обнаруженные на ранних фазах, потенциально более опасны. В связи с этим вводится дополнительный множитель:

$$S^*_{source} = S_{source} * W_{level}, \quad (6)$$

где $W_{level}$ – выбирается экспертно, исходя из конкретной архитектуры мониторинга. Чем выше уровень (например, синтетический), тем заметнее может быть вклад, если ошибка прошла через предыдущие ступени.

Практически введение коэффициента многоуровневости выглядит следующим образом: для каждого уровня M1 – M5 определяется вес $W_{M1} - W_{M5}$ (Таблица 2). При расчете скоринга правила или источника учитывать этот вес, например,

$$S_{type}\ (M3) = S_{base} \times W_{type} \times W_{M3} \quad (7)$$

Принципиально важно иметь возможность усиления или ослабления значения итогового скоринга исходя из того, на каком этапе выявлены проблемы.

Таблица 2

**Сравнительная таблица систем качества данных**

| УРОВЕНЬ МОНИТОРИНГА | IКОЭФФИЦИЕНТ $W_{level}$ | ОБОСНОВАНИЕ |
|---|---|---|
| Инфраструктурный ( $W_{M1}$ ) | 1.0 | Ошибки легко устранимы |
| Статистический ( $W_{M2}$ ) | 1.1 | Ошибки требуют внимания |
| Базового качества ( $W_{M3}$ ) | 1.2 | Ошибки более существенные |
| Поведенческий ( $W_{M4}$ ) | 1.3 | Ошибки критические |
| Синтетический ( $W_{M5}$ ) | 1.4 | Ошибки максимально серьезны |

## 3. ПРИМЕР РАСЧЕТА И РЕКОМЕНДАЦИИ ПО ПРИМЕНЕНИЮ

Чтобы закрепить описанные формулы на практическом примере, рассмотрим пример в SOC:

- имеется источник «S», от которого в сутки поступает в среднем 20 тысяч событий;
- контроль выполняется на трех типах проверок: критичных (5 штук), среднекритичных (10 штук) и базовых (10 штук). Всего 25 проверок;
- фиксируем, что из 5 критичных проверок 2 не пройдены (проблемы в обязательных полях), среди среднекритичных 3 дали сбой, а среди базовых – 2 сбоя;
- исходя из договоренности, W_type, crit = 10, W_type, mid = 5, W_type, base = 1;
- сам источник «S» обозначен как «высокой важности» (W_source = 2).

**Шаг 1: Расчет S_base по каждой группе**

1. Критичные: C_fail,crit = 2.
2. C_total, crit = 5, подставив S_base, crit = 0,4.
3. Среднекритичные: 3/10 = 0,3.
4. Базовые: 2/10 = 0,2.

**Шаг 2: Применение весовых коэффициентов типов**

1. S_type,crit = 0,4 × 10 = 4,0.
2. S_type,mid = 0,3 × 5 = 1,5.
3. S_type,base = 0,2 × 1 = 0,2.

тогда:

- S_type, crit = 4.0,
- S_type, mid = 1.5,
- S_type, base = 0.33,
- число категорий N_types = 3,

то:
если источник «S» имеет вес W_source = 2 (предположим, это критичный источник), тогда:
при интерпретации итогового балла необходимо определить, в каких пределах он может лежать. Если, к примеру, верхней границей считается 10, то значение 3,88 свидетельствует о заметном, но не критическом уровне проблем.

**Шаг 3: Усреднение по типам**

Число типов N_types = 3, тогда среднее = 5,7 / 3 = 1,9.

**Шаг 4: Учет критичности источника**

Пусть W_source = 2 → S_source = 1,9 × 2 = 3,8.

**Шаг 5: Коэффициент уровня (W_level)**

Предположим, что данные проблемы были обнаружены лишь на уровне базового качества данных (M3), и организация расценивает такое позднее выявление с повышающим коэффициентом 1,2 ,

$$\text{тогда } S\_source^* = 3{,}8 \times 1{,}2 = 4{,}56.$$

**Интерпретация:** итоговая оценка 4,56 (при некоторой верхней границе 10) указывает на ощутимое количество ошибок. При переходе порогов (например, >4 – «желтая зона», >7 – «красная зона») источник S может быть помечен как требующий внимания.

Таким образом, используя цепочку формул, аналитики SOC или антифрод-систем могут ранжировать состояние источников, прошедших все уровни мониторинга, и получим итоговый скоринг.

**Реализация и тестирование методики**

Для проверки работоспособности предложенного подхода разработан программный модуль, позволяющий анализировать качество данных в реальном времени и корректировать оценку инцидентов на основе полученных данных. Эксперименты показали снижение числа ложных срабатываний на 15–20% и повышение точности детекции инцидентов.

## ЗАКЛЮЧЕНИЕ

Разработанная методика скоринга источников данных и инцидентов в многоуровневых системах качества данных позволяет снизить влияние некорректных данных на работу SOC и антифрод-систем. Использование адаптивных алгоритмов оценки качества данных повышает эффективность мониторинга и снижает издержки, связанные с обработкой ложных инцидентов.

Кроме того, включение коэффициента многоуровневости представляется обоснованным шагом для отражения архитектуры системы мониторинга. Многоуровневый анализ качества данных встраивается в формулы через дополнительный вес, что находит поддержку в концепциях иерархической оценки (как показали примеры из научных работ). Такой коэффициент позволит учесть важность уровня мониторинга и глубину проникновения ошибки в систему. В итоге методика скоринга, описанная в статье, имеет прочную научную основу, а ее развитие в направлении учета многоуровневости соответствует современным представлениям о комплексном мониторинге качества данных и рисков.

Рассмотренные выражения (1)–(5) скоринга (S_base, S_type, S_source, S_dataset), предложенные для многоуровневого мониторинга качества данных в SOC и AI, оказались согласованными с научными представлениями и практикой смежных областей. Базовая метрика S_base соответствует доле ошибок – фундаментальному показателю качества, формула S_type реализует взвешивание по критичности типа аналогично моделям риска (вероятность × влияние), агрегирование S_source через среднее по типам и вес источника повторяет известные схемы объединения показателей с учетом важности объекта, а итоговая оценка S_dataset прямо использует формулу взвешенного среднего, признанную и в промышленности, и в литературе по качеству данных. Научной же новизной выступают коэффициенты многоуровневости дающие большую гибкость и точность в выявлении сбоев.

Список использованных источников и литературы / References

1. IBM Security operations center (SOC) – официальный сайт. URL: https://www.ibm.com/think/topics/security-operations-center?ysclid=maplv8abcf874764913 (дата обращения: 30.04.2025).
2. Бурый А.С., Фролов В.А., Куляница А.Л. Эволюция агентного моделирования. Часть 1. Архитектура интеллектуального агента // Информационно-экономические аспекты стандартизации и технического регулирования. 2023. № 5(74). С. 38–47.
3. Лопатин И.Н. Многоуровневые системы качественных данных на основе моделей искусственного интеллекта: проблемы и решения // Информационно-экономические аспекты стандартизации и технического регулирования. 2025. № 1 (82). С. 70–75.
4. DAMA International. DAMA-DMBOK: Data Management Body of Knowledge. Technics Publications, 2017.
5. ISO 8000-61:2016. Data quality – Part 61: Data quality management: Process reference model.
6. Informatica Data Quality (IDQ) – официальный сайт. URL: https://www.informatica.com/products/data-quality.html (дата обращения: 30.04.2025).
7. IBM InfoSphere Information Server (QualityStage) – официальный сайт. URL: https://www.ibm.com/products/infosphere-information-server (дата обращения: 30.04.2025).
8. SAP Information Steward – официальный сайт. URL: https://www.sap.com/products/technology-platform/information-steward.html (дата обращения: 30.04.2025).
9. Oracle Enterprise Data Quality (EDQ) – официальный сайт. URL: https://www.oracle.com/middleware/technologies/enterprise-data-quality.html (дата обращения: 30.04.2025).

10. Ataccama ONE Platform – официальный сайт. URL: https://www.ataccama.com/platform/one (дата обращения: 30.04.2025).
11. Collibra Data Quality – официальный сайт. URL: https://www.collibra.com/us/en/products/data-quality (дата обращения: 30.04.2025).
12. Experian Data Quality (EDQ) – официальный сайт. URL: https://www.edq.com (дата обращения: 30.04.2025).
13. FIRST.org. Common Vulnerability Scoring System v3.1: Specification Document. URL: https://www.first.org/cvss/v3.1/specification-document (дата обращения: 30.04.2025).
14. Talend Data Quality – официальный сайт. URL: https://www.talend.com/products/data-quality/ (дата обращения: 30.04.2025).
15. SAS Data Quality – официальный сайт. URL: https://www.sas.com/en_us/software/data-quality.html (дата обращения: 30.04.2025).
16. Precisely Trillium – официальный сайт. URL: https://www.precisely.com/product/precisely-trillium (дата обращения: 30.04.2025).
17. FIRST.org. Common Vulnerability Scoring System v3.1: Specification Document, 2019. CVSS v3.1 User Guide.
18. OWASP Risk Rating Methodology. OWASP Foundation, 2021. OWASP Risk Rating.
19. Redman T.C. Data Driven: Profiting from Your Most Important Business Asset. Harvard Business Press, 2008.

# SCORING METHODOLOGY FOR SOURCES AND INCIDENTS IN MULTILEVEL DATA QUALITY SYSTEMS

**Lopatin I.N.,** PhD student, Russian Standardization Institute, Moscow

*In the contemporary digital environment, data quality has become a crucial element affecting the reliability and effectiveness of cybersecurity systems, anti-fraud mechanisms, and artificial intelligence solutions utilizing Small Language Models (SLM). The implementation of a multilevel data quality scoring methodology, covering comprehensive monitoring from basic infrastructure checks, statistical analysis, primary data quality assessments, behavioral anomaly detection, to synthetic data testing, ensures effective and timely identification of data anomalies and inconsistencies. Poor data quality can significantly compromise critical systems, leading to operational disruptions, overlooked cybersecurity threats, financial losses, and increased risks. A systematic approach integrating accumulated scientific knowledge on data quality assessment algorithms, correlation engine technologies, and scoring mechanisms specifically designed for anti-fraud systems substantially enhances the effectiveness of data management. Comparative analysis with existing international data quality evaluation methodologies clearly demonstrates the advantages of the proposed multilevel scoring approach in terms of accuracy, practicality, and reliability, making investment in robust data quality management a strategic priority in today's data-driven landscape.*

**Keywords:** data quality, cybersecurity, anti-fraud systems, SLM, multi-agent systems, multilevel monitoring, data quality algorithms, data quality assessment methods.

**For citation:** Lopatin I.N. Scoring methodology for sources and incidents in multilevel data quality systems. Information and Economic Aspects of Standardization and Technical Regulation. 2025; 4(85): 101–107. (In Russ.).