

## Метрологический дефицит в промышленных «больших данных»

Михаил Михайлович Елисейкин<sup>1✉</sup>, Валерий Фёдорович Очков<sup>2</sup>

<sup>1,2</sup>Московский энергетический институт (национальный исследовательский университет), Москва, Россия

<sup>1</sup>[muxa@muxa.ru](mailto:muxa@muxa.ru) ✉, <https://orcid.org/0000-0001-8754-0699>

<sup>2</sup>[OchkovVF@mpei.ru](mailto:OchkovVF@mpei.ru)

**Аннотация.** Широкое распространение киберфизических систем в промышленности, рост использования машинного обучения, искусственного интеллекта, цифровых двойников и другие технологии цифровизации предъявляют высокие требования к качеству собираемых данных, характеризующих объект исследования. Современные вычислительные мощности позволяют получать измерительную информацию со скоростью, недоступной уровню развития прошлых лет. Дефицит вычислительных мощностей приводил к тому, что анализ собираемых данных проводился на основании только лишь значений величин, без метрологических метаданных (погрешность, принцип измерений, характеристики средств измерений и др.). Прежние подходы строились на той аксиоме, что если средство измерений имеет статус поверенного и допущенного к использованию, то полученные данные имеют допустимую точность. В настоящее время нет дефицита в вычислительных мощностях и готовых программных решениях для анализа данных и даже не самые крупные предприятия могут проводить анализ данных одновременно и быстро, и точно. Появилась возможность учитывать метрологические метаданные, которым не уделяется должного внимания. По сути, можно говорить о том, что в промышленных «больших данных» имеет место метрологический дефицит, что создаёт риски, связанные с ошибочными выводами при анализе этих данных. Авторы показывают, как на основании существующей в России материальной и нормативной базы, можно решить проблемы метрологического дефицита в собираемых промышленных «больших данных».

**Ключевые слова:** большие данные, законодательная метрология, искусственный интеллект, машинное обучение, метрологический дефицит, цифровизация промышленности, цифровые двойники

**Финансирование:** отсутствует.

**Для цитирования:** Елисейкин М.М., Очков В. Ф. Метрологический дефицит в промышленных «больших данных» // Законодательная и прикладная метрология. 2024. № 4. С. 19–24. <https://doi.org/10.32446/2782-5418.2024-4-19-24>

## Metrological deficit in industrial “big data”

Mikhail M. Eliseikin<sup>1✉</sup>, Valerii F. Ochkov<sup>2</sup>

<sup>1,2</sup>Moscow Power Engineering Institute (National Research University), Moscow, Russia

<sup>1</sup>[muxa@muxa.ru](mailto:muxa@muxa.ru) ✉, <https://orcid.org/0000-0001-8754-0699>

<sup>2</sup>[OchkovVF@mpei.ru](mailto:OchkovVF@mpei.ru)

**Abstract.** The widespread use of cyber-physical systems in industry, the growing use of machine learning, artificial intelligence, digital twins and other digitalization technologies place high demands on the quality of collected data characterizing the object of research. Modern computing power allows obtaining measurement information at a speed inaccessible to the level of development of the past. The deficit of computing power led to the fact that the analysis of collected data was carried out on the basis of values of quantities only, without metrological metadata (error, principle of measurement, characteristics of measuring instruments, etc.). The former approaches were based on the axiom that if a measuring instrument has the status of verified and authorized for use, the data obtained have an acceptable accuracy. Nowadays there is no shortage of computing power and ready-made software solutions for data analysis, and even not the largest enterprises can analyze data both quickly and accurately. It has become possible to take into account metrological metadata, which

has not been given due attention. In essence, we can say that there is a metrological deficit in industrial Big Data, which creates risks associated with erroneous conclusions when analyzing this data. The authors show how, based on the existing material and regulatory framework in Russia, it is possible to solve the problems of metrological deficiency in the collected industrial “big data”.

**Keywords:** artificial intelligence, big data, digital twins, digitalization of industry, legal metrology, machine learning, metrological deficit

**Financial Support:** none.

**For citation:** Eliseykin M. M., Ochkov V. F. Metrological deficit in industrial “big data”. *Legal and applied metrology*. 2024, no. 4, pp. 19–24. <https://doi.org/10.32446/2782-5418.2024-4-19-24>

**Актуальность проблемы.** В настоящее время наблюдается рост использования в промышленности технологий машинного обучения, искусственного интеллекта, цифровых двойников и других решений для анализа состояния и моделирования работы как отдельных агрегатов, так и предприятий в целом. Суть подобных решений состоит в том, чтобы собрать как можно больше численных данных о моделируемом объекте и, используя методы статистического анализа, попытаться выявить корреляцию между различными параметрами объекта. Уже сейчас потребителю доступны не только вычислительные мощности и программное обеспечение (ПО), позволяющее решать поставленные задачи, но и квалифицированные исполнители, специализирующиеся на таком анализе.

Однако кроме новых возможностей, такие современные технологии могут принести и новые риски. Нередко аналитики, использующие технологии машинного обучения и цифровых двойников, работают только с численной частью анализируемых данных и упускают из вида как этиологию этих данных, так и их физический смысл.

Известно, что данные, получаемые в результате измерений, не являются абсолютно достоверными и неизбежно содержат погрешность измерений. Раньше, до широкого развития цифровых технологий обработки данных, было достаточно выполнения условия, чтобы измерения проводились с использованием поверенных средств измерений (СИ) и их результаты были метрологически совместимыми, что позволяло контролировать производственные процессы. Актуальные данные можно было сравнивать с диапазоном допустимых значений и делать выводы о работе отдельного агрегата или всего предприятия в целом. При этом анализ накопленных данных вызывал трудности в связи с отсутствием у предприятий вычислительных мощностей и необходимого ПО. Расчёты производились

либо быстро, но недостаточно точно, либо более точно, но долго.

В настоящее время высокий уровень развития вычислительной техники позволяет работать как с актуальными данными в режиме реального времени, так и с накопленными ретроспективными данными. И в случае с промышленным оборудованием глубина наблюдений может измеряться в годах и десятилетиях. Например, в работе Андрияшина А. В. [1] приводятся результаты 6-летних наблюдений с 9586 данными измерений температуры окружающего воздуха и давления в конденсаторе паровой турбины.

Однако данные, о которых идёт речь, собирались для того, чтобы на их основании принимать решения на момент сбора этих данных. При этом специалисты, которые принимают решения (технологи, операторы, руководители), должны владеть дополнительной информацией о том, в каких условиях произведены измерения. Эта дополнительная информация не всегда отражается в записях результатов измерений.

Именно в этом и кроется проблема, которая станет актуальной в ближайшем будущем — современные инструменты анализа данных способны обнаружить ранее считавшиеся незначительными изменения в числовых значениях величин, но они не способны отличить изменения значения физической величины в реальности от изменения погрешности, возникающей в процессе измерения этой величины.

Например, при замене СИ на однотипное, систематическая погрешность измерений может измениться с  $-1\%$  на  $+1\%$ , при этом оставаясь в допустимых пределах  $\pm 1\%$ . При замене СИ на более совершенное и с более широким диапазоном измерений значение, бывшее на краю диапазона и имевшее погрешность, например,  $3\%$  будет измеряться ближе к центру диапазона и иметь погрешность  $1\%$ . При использовании СИ, осуществляющих

перерасчёт или корректировку результатов измерений (например, газовые расходомеры, вносящие поправку на температуру [2]) может измениться методика корректировки. При различной разрешающей способности СИ и точности записи, одно и то же действительное значение измеряемой величины может быть представлено в виде разных числовых значений.

Таким образом, существует много разных ситуаций, в которых изменение числового значения является метрологическим казусом, а не реальным изменением параметров отслеживаемого рабочего процесса. И если бы в исходных данных были не только числовые значения, но и метаданные, содержащие информацию о погрешности измерений и СИ, то изменение в данных могло бы быть сопоставлено с фактом замены СИ и учтено при дальнейшем анализе. Без этого аналитик, ищущий в данных закономерности, отражающие рабочие процессы, может найти несуществующие закономерности, что приведёт к созданию неправильного цифрового двойника и ошибочным решениям, принимаемым на основании этого цифрового двойника.

Можно сделать вывод, что у решений, связанных с машинным обучением и цифровыми двойниками, наблюдается хорошо развитая математическая составляющая при очевидном отставании метрологической.

По сути, мы имеем дело с явлением, которое можно обозначить как *метрологический дефицит* — у нас есть только числовое значение величины, без информации о том, как и с какой погрешностью это значение было получено.

С сожалением приходится отметить, что данная проблема связана не только с тем, как работают программисты и аналитики, занятые сейчас в цифровизации промышленности, но и с нормативной базой.

**Отсутствие в ГОСТах метрологической составляющей.** Начиная с 2014 года в России было создано несколько технических комитетов [3], в задачи которых входит разработка новых цифровых технологий. При участии этих комитетов

был разработан ряд ГОСТов, регламентирующих использование компьютерного моделирования в промышленности.

Так, в настоящий момент действует семейство стандартов ГОСТ Р 57700, посвящённых численному моделированию физических процессов. В этих стандартах уделяется много внимания погрешности компьютерных вычислений, но практически не говорится о погрешности исходных экспериментальных данных.

Исключение составляют три стандарта, два из которых посвящены моделированию высокоскоростных процессов: ГОСТ Р 57700.11–2018<sup>1</sup> и ГОСТ Р 57700.16–2018<sup>2</sup>.

В данных документах содержится требование указывать погрешность получения первичных экспериментальных данных в матрице валидации модели.

Учитывая специфику данных стандартов, посвящённых ударным процессам, можно сделать вывод о том, что информация о погрешности экспериментальных данных является исключением и не рассматривается в качестве обязательной части процесса численного моделирования.

Некоторую долю оптимизма внушает более поздний ГОСТ Р 57700.25–2020<sup>3</sup>, в котором есть отдельный термин «3.1.13 экспериментальная погрешность: Снижение точности измерений в результате натурального эксперимента, обусловленная случайной (статистической) и систематической ошибками». Однако и в этом случае речь не идёт о систематическом сборе метрологической информации.

К сожалению, специализированный ГОСТ Р 57700.37–2021<sup>4</sup> вообще не затрагивает вопрос погрешности собираемых и используемых исходных данных.

Подобная ситуация является проблемой потому, что при создании цифровых двойников как разработчик, так и заказчик руководствуются существующей нормативной базой и устоявшейся практикой, но ни нормативы, ни практика на данном рынке не требуют от разработчика, чтобы они сохраняли и использовали метаданные метрологического характера.

<sup>1</sup> ГОСТ Р 57700.11–2018. Численное моделирование физических процессов. Процессы ударного взаимодействия. Верификация и валидация численных моделей низкоскоростных ударов и внедрений. Требования.

<sup>2</sup> ГОСТ Р 57700.16–2018. Численное моделирование физических процессов. Процессы ударного взаимодействия. Верификация и валидация численных моделей высокоскоростных ударов и внедрений. Общие требования.

<sup>3</sup> ГОСТ Р 57700.25–2020. Компьютерные модели и моделирование. Процедуры валидации.

<sup>4</sup> ГОСТ Р 57700.37–2021. Компьютерные модели и моделирование. Цифровые двойники изделий. Общие положения.

Боле того, в силу отсутствия такого требования, специалисты, занимающиеся анализом промышленных данных, не всегда понимают о чём идёт речь. Один из авторов статьи участвовал в обсуждении интернет-публикаций, посвящённых вопросам машинного обучения и цифровых двойников в промышленности, и данные обсуждения показывают, что аналитики больше интересуются математическими методами анализа, чем происхождением анализируемых данных.

Может показаться, что отсутствие в ГОСТах указаний на необходимость сохранения метрологических метаданных делает решение указанной проблемы трудновыполнимым. Раз нет требований, значит нет и решений, удовлетворяющих требованиям.

На самом же деле, имеющиеся технические стандарты позволяют приступить к решению данной задачи уже сейчас.

**Ввод стандарта на IoT.** Одним из основных направлений в цифровизации промышленности и развитии киберфизических систем является концепция «интернета вещей», предполагающая использование на предприятиях СИ для автоматизированного сбора данных о производственном процессе и состоянии оборудования [4].

В рамках этой концепции используются протоколы, в которые уже заложены технические возможности, которые можно использовать для учёта метрологической составляющей при накоплении данных. Поэтому уже сейчас мы можем начать собирать данные, позволяющие идентифицировать СИ и фиксировать их замену.

15 января 2024 было объявлено о принятии ГОСТ Р 71168–2023<sup>5</sup>, который начинает действовать с 01 июля 2024 года взамен временного стандарта ПНСТ 516–2021<sup>6</sup>. Областью применения данного стандарта является беспроводной обмен данными с бытовым и промышленным оборудованием, таким как СИ, приборы, датчики, исполнительные устройства и т.п. [5]

В данном стандарте заложена возможность идентификации устройства, за которую отвечает идентификатор оконечного устройства DevEU I. DevEUI — это 64-битный идентификатор, являющийся аналогом MAC-адреса сетевых карт или IMEI-номера телефона. По сути, это глобально уникальный номер [6], содержащий информацию о производителе, модели и серийном номере

устройства. Гипотетически этот идентификатор может быть изменён, но на практике он является надёжным способом идентификации устройства с помощью набора символов, которые можно выразить не только в машиночитаемом, но и в человекочитаемом виде. Поэтому в рассматриваемом стандарте, в примечании к пункту 6.4.1.1.6 сказано, что DevEUI рекомендуется использовать в качестве номенклатурного номера устройства, а также для его учёта и сопровождения. То есть идентификатор DevEUI будет использован как в процессе общения сервера с конечным устройством, так и в бухгалтерской и технической документации всего производства.

Следовательно, при разработке систем мониторинга и сбора данных не придётся разрабатывать принципиально новых технических и организационных решений — надо будет просто начать сохранять идентификатор СИ вместе с полученными от него числовыми значениями, временной отметкой и другими данными.

Уже это позволит в будущем отличить изменения в данных, связанных с реальными физическими процессами от изменений, связанных с заменой измерительного устройства.

Также сохранение идентификатора DevEUI даст возможность определить модель СИ, а следовательно, и его метрологические характеристики: используемый физический принцип, погрешность, рабочий диапазон и т.п. (в том числе, в случае возникновения путаницы в технической документации предприятия или её утрате).

Кроме уже указанного риска возможного изменения DevEUI в течение жизненного цикла СИ, есть вероятность того, что СИ, передающий информацию, является не единым целым, а комплексом, в котором есть отдельный измерительный преобразователь и отдельное передающее устройство (например, импульсный расходомер и счётчик импульсов). В этом случае может возникнуть ситуация, когда конфигурация СИ была изменена, а идентификатор DevEUI остался прежним.

Тем не менее, несмотря на перечисленные риски, уже сейчас можно рассматривать ГОСТ Р 71168–2023 и идентификатор DevEUI как основу, позволяющую ввести метрологическую составляющую в процесс сбора данных, которые потом будут использоваться для машинного обучения и создания цифровых двойников.

<sup>5</sup> ГОСТ Р 71168–2023. Информационные технологии. Интернет вещей. Спецификация LoRaWAN RU.

<sup>6</sup> ПНСТ 516–2021. Информационные технологии. Интернет вещей. Спецификация LoRaWAN RU.



**Оценка затрат.** Устранение метрологического дефицита в собираемых данных может дать экономическую выгоду за счёт построения более точных моделей и увеличения эффективности предприятия, но сам процесс устранения метрологического дефицита не бесплатен. Причём, дополнительных затрат потребует не только внедрение процесса сбора дополнительных данных, но и их хранение.

И если финансовые затраты будут зависеть от множества условий и не поддаются предварительной оценке, то технические затраты на хранение оценить можно.

Предположим, что есть СИ, передающее температуру и необходимо сохранить массив данных в определённом временном диапазоне.

До начала сохранения DevEUI мы собирали данные «время» и «температура».

При хранении информации в базе данных:

- «время» потребует 4 либо 8 байт, в зависимости от используемого типа данных;
- «температура» потребует 4 либо 8 байт, в зависимости от используемого типа данных.

Хранение 64-битного идентификатора DevEUI потребует 8 байт.

Таким образом, приблизительная оценка показывает, что при сборе данных о СИ, общий объём хранимой информации вырастет в 1,5–2 раза. Если же в базе данных будет храниться ещё какая-то информация (например, привязка показаний к конкретному оборудованию), а информация о СИ будет храниться не в виде самого DevEUI, а как внешний ключ, то коэффициент увеличения объёма хранимых данных будет ещё меньше.

Если же оценить это в абсолютных значениях, то окажется, что объём базы данных в миллион записей увеличится примерно на 8 мегабайт, что в современных условиях является пренебрежимо малым значением.

Таким образом, сбор дополнительных данных может потребовать дополнительных расходов на начальном этапе, но вряд ли в дальнейшем ляжет тяжким бременем на бюджет промышленного предприятия.

**Физический смысл данных.** Можно было бы сказать, что затронутая в данной статье проблема не игнорируется, а просто не имеет широкого освещения. Возможно, метрологические данные уже собираются, но, в силу очевидности, этот вопрос не считают нужным поднимать.

К сожалению, есть основание считать, что вопрос происхождения, физического смысла и реальной метрологической совместимости данных

действительно не интересует специалистов, связанных с анализом данных. Презентации, рекламные материалы, обсуждения в интернете и научные публикации часто содержат сведения о том, как с помощью статистических методов исключить «промахи» — данные, которые настолько сильно отличаются от остальных, что считаются заведомо ошибочными. Однако «промах» в числовых данных не всегда означает ошибочное значение.

Например, выброс данных может быть связан с использованием разных единиц измерений. Часто современные электронные устройства имеют возможность работать в разных системах. Навигационные модули имеют возможность работать с разными спутниковыми навигационными системами, а коммуникационные — с разными частотами и протоколами. Такой подход позволяет адаптировать одно и то же устройство к разным рынкам.

Аналогичная ситуация может быть и с единицами измерений — СИ может иметь возможность передачи данных с использованием разных единиц измерений. Это может быть как использование разных систем (метрическая или имперская), так и использование кратных или дольных единиц измерений.

Таким образом СИ, передающее информацию, может передавать температуру в градусах Фаренгейта, а не Цельсия, или расход в кубометрах в час, а не в литрах в секунду. С точки зрения анализа числовых данных, перед нами будет бесспорный «промах», который следует исключить. Однако это будут достоверные данные, которые можно и нужно использовать при анализе.

И хотя единицы измерений, в отличие от идентификаторов типа DevEUI, не является обязательной частью протокола, она может присутствовать в данных, передаваемых СИ. В этом случае будет правильным сохранять её вместе с накапливаемыми данными и учитывать при анализе текущей ситуации.

В связи с ростом числа киберфизических систем, последнее может быть очень актуально, потому что позволит обнаружить проблему в данных на этапе контроля размерности.

Также для контроля может быть использован сам факт наличия единиц измерений. В качестве примера можно привести аварию с АМС «Луна-25». Судя по официальным данным, причиной аварии стало то, что бортовой компьютер не видел разницы между отсутствием показаний акселерометра и показаниями, равными 0 (ноль).

И если бы вместе с числовыми данными передавались единицы измерений, то бортовой компьютер обнаружил бы сбой по их отсутствию.

**Выводы.** Развитие киберфизических систем, широкое использование машинного обучения, цифровых двойников и других современных цифровых технологий предъявляют высокие требования к качеству данных о состоянии как отдельных устройств, так и всего предприятия в целом.

В случае некорректности или неполноты этих данных имеется риск принятия ошибочных решений, которые позволят создать адекватную модель производства в виде цифрового двойника и могут привести к аварии. Поэтому уже сейчас надо готовиться к тому, что данные будут собираться годами, и что будет происходить замена СИ, изменение методик измерений погрешностей СИ.

В настоящий момент нормативы не требуют собирать метрологические метаданные в дополнение к числовым, но имеющиеся технические решения уже позволяют начать это делать.

И надо чётко понимать, что если не собирать данные сейчас, то в будущем нечего будет анализировать.

#### Список источников

1. Андрюшин А. В., Щербатов И. А., Цуриков Г. Н. Обнаружение выбросов данных в системах предиктивной аналитики для оборудования энергетике // Математические методы в технике и технологиях — ММТТ. 2019. Т. 10. С. 112–115. <https://www.elibrary.ru/item.asp?id=41261473>
2. Беляев А. О., Ковтун Д. Г. Электронная температурная коррекция объема в приборах учета потребления природного газа // Инженерный вестник Дона: сетевой журнал. 2016. № 4. URL: <http://www.ivdon.ru/ru/magazine/archive/n4y2016/3837> (дата обращения: 14.05.2024)
3. Туровец Ю. В., Вишневецкий К. О. Стандартизация цифрового производства: возможности для России и ЕАЭС // Бизнес-информатика. 2019. Т. 13, № 3. С. 78–96. <https://doi.org/10.17323/1998-0663.2019.3.78.96>.
4. Цуриков Г. Н., Щербатов И. А. Применение промышленного интернета вещей на объектах энергетике // Мехатроника, автоматика и робототехника. 2018. № 2. С. 97–100. <https://elibrary.ru/item.asp?id=32490604>
5. Якушков К. В. LoRaWAN RU — стандарт родом из России // Информатизация и системы управления в промышленности. 2023. № 2. С. 132–134. <https://elibrary.ru/item.asp?id=53944045>
6. Массеров Д. А., Массеров Д. Д. Обеспечение кибербезопасности умного города в процессе цифровизации городской среды // Информационные технологии и математическое моделирование в управлении сложными системами: сетевой журнал. 2022. № 1. С. 32–41. URL: [https://ismm.irknps.ru/sites/default/files/articles\\_pdf\\_files/ensuring\\_the\\_cyber\\_security.pdf](https://ismm.irknps.ru/sites/default/files/articles_pdf_files/ensuring_the_cyber_security.pdf) (дата обращения: 14.05.2024).

#### References

1. Andryushin AV, Shcherbatov IA, Tsurikov GN. Outlier detection in predictive analytics for energy equipment. *Mathematical Methods in Technique and Technologies* — ММТТ. 2019;10:112–115. (In Russ.)
2. Belyaev AO, Kovtun DG. Electronic volume temperature correction in natural gas accounting units. *Engineering Journal of Don: electronic scientific journal*. 2016;(4). Available at: <http://www.ivdon.ru/ru/magazine/archive/n4y2016/3837> (Accessed: 14.05.2024). (In Russ.)
3. Turovets YuV, Vishnevskiy KO. Standardization in digital manufacturing: implications for Russia and the EAEU. *Business Informatics*. 2019;13(3):78–96. (In Russ.)
4. Tsurikov GN, Shcherbatov IA. Primenenie promyshlennogo interneta veshchei na ob"ektakh energetiki [Application of the Industrial Internet of Things at Energy Facilities]. *Mekhatronika, avtomatika i robototekhnika*. 2018;(2): 97–100. (In Russ.)
5. Yakushkov KV. LoRaWAN RU — standart rodom iz Rossii [LoRaWAN RU — standard comes from Russia]. *Informatizatsiya i sistemy upravleniya v promyshlennosti*. 2023;(2):132–134. (In Russ.)
6. Masserov DA, Masserov DD. Ensuring the cyber security of the smart city in the process of digitalisation of the urban environment. *Information technology and mathematical modeling in the management of complex systems: electronic scientific journal*. 2022;1:32–42. Available at: [https://ismm.irknps.ru/sites/default/files/articles\\_pdf\\_files/ensuring\\_the\\_cyber\\_security.pdf](https://ismm.irknps.ru/sites/default/files/articles_pdf_files/ensuring_the_cyber_security.pdf) (Accessed: 14.05.2024). (In Russ.)

Дата принятия 12.07.2024