

Научная статья
УДК 004.9:005:352
DOI: 10.17586/2713-1874-2026-3-82-93

ЦИКЛ РАБОТЫ С ДАННЫМИ КАК КОНТУР УПРАВЛЕНИЯ В ОРГАНАХ ПУБЛИЧНОЙ ВЛАСТИ

Данил Иванович Шенцов^{1✉}, Мария Яковлевна Марусина²

¹Высшая школа экономики, Москва, Россия, dshentsov@hse.ru[✉], ORCID: 0009-0007-8554-1402

²Университет ИТМО, Санкт-Петербург, Россия, myamarusina@itmo.ru, ORCID: 0000-0002-5611-8812

Язык статьи – русский

Аннотация: Цифровая трансформация государственного управления сместила фокус управленческой деятельности на работу с данными, что выдвигает задачу построения устойчивого и управляемого цикла работы с ними. Целью исследования является повышение обоснованности управленческих решений в организационной системе органа публичной власти за счёт представления цикла работы с данными как замкнутого контура управления с обратной связью по состоянию данных. Материалом для демонстрации послужил открытый набор данных «Адресный реестр объектов недвижимости города Москвы» объёмом более 548 тысяч записей. Методологической основой выступает теория управления организационными системами; цикл работы с данными формализован в парадигме извлечения, преобразования и загрузки данных и рассмотрен относительно эталонной модели жизненного цикла данных. Разработан метод организации цикла работы с данными как контура управления, в котором узел контроля качества выполняет идентификацию состояния системы, а корректирующее воздействие направляется адресно, в зависимости от источника выявленного дефекта. Предложена реализация контура на открытом инструментальном стеке распределённой обработки, аналитического хранения и визуализации данных. Показано, что контроль качества, опираясь на нормативные требования к атрибутам данных, выявляет и различает по природе несколько типов дефектов, а обратная связь замыкается на соответствующие этапы обработки. Сделан вывод о применимости предложенного контура в организационных системах органов государственного управления. Ограничением работы является иллюстративный характер демонстрации; направлением дальнейших исследований выступает количественная оценка эффективности и распространение контура на задачи формирования синтетических данных.

Ключевые слова: государственное управление, жизненный цикл данных, контроль качества данных, органы публичной власти, организационные системы, открытые данные, цифровая трансформация, ETL-процесс

Ссылка для цитирования: Шенцов Д. И., Марусина М. Я. Цикл работы с данными как контур управления в органах публичной власти // Экономика. Право. Инновации. – 2026. – Т. 14. – № 3. – С. 82–93. – DOI: 10.17586/2713-1874-2026-3-82-93.

DATA HANDLING CYCLE AS A CONTROL LOOP IN PUBLIC ADMINISTRATION

Danil I. Shentsov^{1✉}, Maria Ya. Marusina²

¹Higher School of Economics, Moscow, Russia, dshentsov@hse.ru[✉], ORCID: 0009-0007-8554-1402

²ITMO University, Saint Petersburg, Russia, myamarusina@itmo.ru, ORCID: 0000-0002-5611-8812

Article in Russian

Abstract: The digital transformation of public administration has shifted the focus of management toward working with data, which raises the task of building a sustainable and controllable data-handling cycle. The aim of the study is to increase the validity of management decisions in the organizational system of a public authority by representing the data-handling cycle as a closed control loop with feedback on the state of the data. The demonstration is based on the open dataset «Address Register of Real Estate Objects of the City of Moscow» comprising more than 548 thousand records. The methodological basis is the theory of control of organizational systems; the data-handling cycle is formalized within the extract-transform-load paradigm and considered relative to a reference data-lifecycle model. A method is developed for organizing the data-handling cycle as a control loop, in which the quality-control node performs identification of the system state, and corrective action is routed in an addressed manner depending on the source of the detected defect. An implementation of the loop on an open-source stack for distributed processing, analytical storage, and data visualization is proposed. It is shown that quality control, relying on regulatory requirements for data attributes, detects and distinguishes several types of defects by their nature, and that feedback closes onto the corresponding processing stages. It is concluded that the proposed loop is applicable in the organizational systems of public-administration bodies. A limitation

of the work is the illustrative nature of the demonstration; a direction for further research is the quantitative assessment of effectiveness and the extension of the loop to the generation of synthetic data.

Keywords: public administration, data lifecycle, data quality control, public authorities, organizational systems, open data, digital transformation, ETL process

For citation: Shentsov D. I., Marusina M. Ya. Data-Handling Cycle as a Control Loop in Public Administration. *Ekonomika. Pravo. Innovacii*. 2026. Vol. 14. No. 3. pp. 82–93. (In Russ.). DOI: 10.17586/2713-1874-2026-3-82-93.

Введение. Цифровая трансформация государственного управления, которая затронула большинство развитых и развивающихся стран, сместила фокус управленческой деятельности с распорядительных процедур на работу с данными: сбор, хранение, обработку и интерпретацию больших массивов разнородной информации, на основе которых принимаются решения [1]. В организационных системах органов публичной власти данные перестают быть вспомогательным ресурсом и становятся объектом управления. От их полноты, консистентности, точности и актуальности напрямую зависит обоснованность управленческих решений и устойчивость работы организации в целом [2], что выдвигает на первый план задачу не разовой обработки данных, а построения устойчивого, воспроизводимого и управляемого цикла работы с ними.

Развитие электронного управления в России закреплено рядом программных документов («Электронная Россия», «Цифровая экономика Российской Федерации» и другие), что формирует институциональные условия для перехода к управлению на основе данных.

Особенность больших данных в государственном управлении состоит не столько в их объёме, сколько в разнородности источников и в их непрерывном поступлении из ведомственных информационных систем, административных регистров, сенсоров и обращений граждан.

Большие данные находят применение в различных областях государственного управления [3]. В сфере транспорта интеллектуальные транспортные системы, обрабатывая данные о передвижении, позволяют оптимизировать транспортные потоки и снижать заторы на дорогах [4]. В здравоохранении анализ медицинских данных применяется для раннего выявления заболеваний и повышения качества обслуживания. Так, в России развёрнута платформа искусственного интеллекта Минздрава России, а сервисы «СберМедИИ»

используются для поддержки врачебных решений. В сфере общественной безопасности данные систем видеонаблюдения используются для профилактики правонарушений. Например, в Московской области за 2023 год установлено более 20 тыс. камер, а связь распознавания лиц с контролем преступности подтверждена зарубежными исследованиями [5].

Специфика органов публичной власти как организационных систем состоит в том, что работа с данными здесь протекает в условиях, качественно отличных от коммерческого сектора: жёсткая регуляторная рамка, ограниченный и частично закрытый доступ к первичным данным, высокая цена ошибки в решениях и необходимость опираться одновременно на ретроспективную, текущую и экспертную информацию. При этом эмпирические исследования аналитики данных в государственном секторе развивающихся стран делают акцент на эффекте, обусловленном низким исходным уровнем цифровизации и внешним финансированием [6, 7], тогда как обобщающая библиометрическая картина показывает, что поле в целом формируется преимущественно развитыми экономиками, в том числе экономикой Китая [8].

При этом полный цикл работы с данными в государственном секторе, как правило, описывается либо как линейный конвейер извлечения, преобразования и загрузки, либо как итеративный процесс [9]. Существующие описания для государственного сектора, включая российский, носят преимущественно технологический или нормативный характер [10].

От организации работы с данными зависит их пригодность для последующего анализа и построения моделей, в том числе методами машинного обучения, что делает управление данными самостоятельной задачей, предшествующей их аналитическому использованию. Возникает объективная потребность в разработке метода управления

организационной системой на основе построения замкнутого цикла работы с данными, учитывающего условия российского государственного сектора.

Постановка задачи исследования. Несмотря на широкое практическое внедрение технологий обработки больших данных в государственном секторе, работа с данными редко описывается как замкнутый контур управления организационной системой. Инженерные решения (конвейеры извлечения, преобразования и загрузки данных) развиваются в отрыве от теории управления, а управленческие модели, как правило, не доходят до уровня программно-технической реализации. В результате цикл работы с данными в органах публичной власти воспроизводится как линейный конвейер без замкнутой обратной связи, что не позволяет управлять состоянием данных и, следовательно, качеством управленческих решений на регулярной основе.

В статье разрабатывается метод организации цикла работы с данными как замкнутого контура управления организационной системой органа публичной власти, обеспечивающего получение и обработку данных, идентификацию их состояния и корректирующее воздействие по обратной связи, и предлагается архитектура его программно-технической реализации на открытом инструментальном стеке, обеспечивающем подготовку данных, пригодных для последующих аналитических задач, включая машинное обучение.

Целью исследования является повышение обоснованности управленческих решений в организационной системе органа публичной власти за счёт построения замкнутого цикла работы с данными с обратной связью по состоянию их качества.

Литературный обзор. Теоретической основой рассмотрения работы с данными как контура управления выступает теория управления организационными системами, развитая школой Института проблем управления имени В. А. Трапезникова РАН. В этой традиции организационная система описывается через участников, механизмы управления (планирование, организацию, стимулирование, контроль) и обратные связи, а управление трактуется как целенаправленное воздействие на систему на основе информации о её состоянии; частным случаем является информационное управление, то есть воздействие

через информацию. Формальные модели рефлексивного и информационного управления, в которых управляющий центр действует, опираясь на наблюдаемое состояние управляемых элементов, представлены в работах той же школы [11]. Принципиально, что именно наличие замкнутой обратной связи отличает управление от разомкнутого исполнения регламентов. Это позволяет сопоставить узлам цикла обработки данных функции управления, где контроль качества выполняет идентификацию состояния, а обратная связь по его результатам, корректирующее воздействие. Такой подход задаёт язык, в котором обработка данных рассматривается не как технологическая процедура, а как контур управления.

Инженерная сторона работы с данными в международной практике оформилась в виде концепции управления данными на всех стадиях их жизненного цикла, в англоязычной литературе обозначаемой как DataOps [12]. Систематический обзор практик DataOps показывает, что эта концепция объединяет извлечение, преобразование и загрузку данных в единый управляемый процесс и нацелена на воспроизводимость и качество, однако рассматривается преимущественно в инженерном ключе, как организация конвейера, а не как контур управления организационной системой [12].

Близкие по существу решения описаны в литературе по архитектурам обработки больших данных: распределённые конвейеры извлечения, преобразования и загрузки данных обеспечивают интеграцию разнородных источников и обработку в режиме, близком к реальному времени [13]. Указанные работы детально проработаны технически, но управленческая интерпретация конвейера, то есть замыкание его результата в контур принятия решений, в них, как правило, не выполняется.

Применительно к государственному сектору предложены обобщённые модели жизненного цикла данных. Наиболее близка рамочная модель жизненного цикла данных для управляемых данными правительств, построенная по результатам систематического обзора и охватывающая полный цикл от сбора до использования [9]. Модель является итеративной и допускает многократное прохождение стадий, оставаясь при этом описанием последовательности фаз.

Вопросы управления данными как ресурсом рассматриваются в литературе по data governance, где данные трактуются как объект управления, а качество данных – как условие обоснованности решений и доверия к ним [2]. Концептуальные обзоры data governance систематизируют механизмы распределения прав и ответственности за данные в организации [14], однако сосредоточены на организационно-правовом контуре и обычно не доходят до программно-технической реализации цикла обработки.

В государственном секторе управление данными изучается и в контексте цифровой трансформации: для России описан переход к дата-центричной модели государственного управления и связанным с ней задачам семантической интероперабельности [10]. Однако и эти работы рассматривают данные преимущественно на уровне архитектуры и нормативного регулирования, не формализуя цикл их обработки как контур управления с обратной связью. Географическая и мотивационная неоднородность этих исследований, отмеченная выше, лишь усиливает разрыв: сложившиеся описания либо привязаны к иным институциональным условиям, либо не доведены до управленческой формализации.

Следует отметить, что отдельные составляющие рассматриваемой задачи проработаны: инженерные конвейеры обработки данных, обобщённые модели жизненного цикла данных для государственного сектора, механизмы data governance и теоретический аппарат управления организационными системами. Вместе с тем, их сочетание, при котором полный цикл работы с данными в организационной системе государственного сектора описывается как замкнутый контур управления с обратной связью по состоянию данных, а сам контур доводится до программно-технической реализации, в известной литературе систематически не представлено.

Это определяет направление настоящего исследования.

Методы и материалы исследования. Метод основан на представлении цикла

работы с данными как замкнутого контура управления организационной системой. Контур реализован в парадигме извлечения, преобразования и загрузки данных (ETL) и содержательно проходит пять этапов: сбор данных, очистку, трансформацию, контроль качества и загрузку. В отличие от линейного конвейера, в котором данные проходят эти этапы однократно и в одном направлении, предлагаемый контур замыкается обратной связью: результат контроля качества служит основанием для корректировки предшествующих этапов на последующих периодах обработки.

Общая структура контура приведена на рисунке 1.

На этапе сбора данные поступают из внешних по отношению к контуру источников: государственных информационных систем и порталов, административных данных, а также, в общем случае, сенсоров и обращений граждан. Здесь реализуется функция наблюдения состояния организационной системы. Данные, поступающие в контур, отражают состояние управляемого объекта на момент их получения. В условиях государственного сектора этап характеризуется ограниченным и регламентированным доступом к первичным данным, что определяет необходимость опоры одновременно на ретроспективную, текущую и экспертную информацию.

Очистка обеспечивает исправление ошибок, нормализацию данных, удаление дубликатов и обработку пропусков. Трансформация формирует признаковое пространство: приведение категориальных признаков к пригодному для обработки виду, построение производных признаков и снижение размерности. Оба этапа выполняются средствами распределённой обработки, что обеспечивает работу с большими объёмами разнородных данных. С точки зрения контура управления эти этапы реализуют преобразование наблюдаемых данных в форму, пригодную для оценки состояния системы.

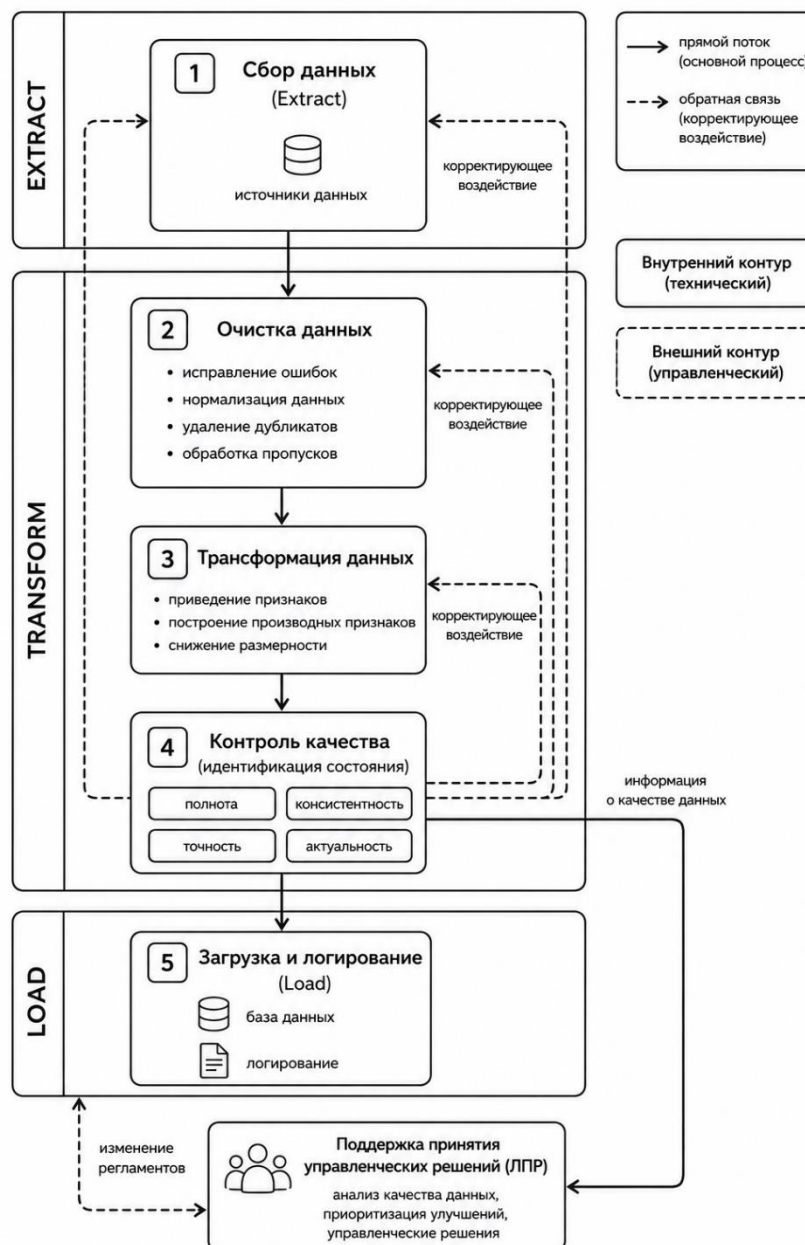


Рисунок 1 – Цикл работы с данными как контур управления
 Источник: составлен автором с использованием генеративной модели ИИ
 (см. «Декларацию о применении ИИ»).

Функцию идентификации состояния системы выполняет контроль качества. На этом этапе данные оцениваются по набору показателей качества (полнота, консистентность, точность, актуальность), конкретный состав которых определяется нормативными требованиями к предметной области. Результат оценки сопоставляется с требуемым уровнем и определяет, требуется ли корректирующее воздействие. Именно этот узел замыкает контур. Выявленное отклонение показателей качества не только фиксируется, но и локализуется по источнику. Если отклонение обусловлено данными на стороне поставщика, кор-

ректирующее воздействие адресуется этапу сбора; если оно вызвано обработкой внутри контура, воздействие адресуется этапам очистки или трансформации. Тем самым обратная связь является адресной: она направляется не в контур в целом, а на тот этап, на котором возник дефект, и учитывается на последующих периодах обработки. Такая организация реализует управление по обратной связи с идентификацией источника отклонения, опирающееся на ретроспективную и текущую информацию о состоянии данных.

Цикл обработки завершает загрузка. Подготовленные данные помещаются в аналити-

ческое хранилище, а операции регистрируются средствами логирования. Данные на выходе контура пригодны для последующего анализа и построения моделей, в том числе методами машинного обучения. Наряду с внутренним контуром, замыкающим обратную связь на этапы обработки, результат контроля качества образует основание для внешнего управленческого контура: показатели состояния данных доступны лицу, принимающему решения, и служат основанием для изменения регламентов сбора и обработки данных в организационной системе.

Предложенный контур соотносится с существующими моделями жизненного цикла данных. В качестве опорной принята рамочная модель DaLiF [9], представляющая собой наиболее полный на сегодня синтез моделей жизненного цикла данных для государственного сектора, построенный по результатам систематического обзора 76 моделей. DaLiF выделяет четырнадцать фаз, включая сбор, подготовку (интеграцию, фильтрацию и обогащение), сквозную фазу качества и хранение, и допускает прохождение фаз в разном порядке и многократно. Вместе с тем, DaLiF является описательной эталонной моделью: качество в ней рассматривается как проверка, применяемая на каждой фазе, а обратная связь ограничена реакцией на пользовательские замечания к опубликованным данным. DaLiF не формализует замкнутый контур, в котором результат контроля качества идентифицирует источник дефекта и направляет корректирующее воздействие на конкретный ответственный этап, не описывает внешний управленческий контур и не использует аппарат теории управления.

Идея контрольных точек качества с обратной связью в моделях жизненного цикла данных предлагалась и ранее [15]. Вклад настоящей работы состоит не в наличии обратной связи как таковой, а в представлении цикла работы с данными как замкнутого контура управления организационной системой в терминах теории управления организационными системами [16]. Узел контроля качества выполняет идентификацию состояния и направляет адресное корректирующее

воздействие на этап-источник отклонения, а внешний управленческий контур доводит информацию о состоянии данных до лица, принимающего решения. В отличие от DaLiF, носящей концептуальный характер и, по признанию авторов, не прошедшей эмпирической проверки, предложенный контур апробирован на реальных данных действующего органа публичной власти.

Программно-техническая реализация контура выполнена на открытом инструментальном стеке. Распределённая обработка данных на этапах сбора, очистки и трансформации реализована средствами Apache Hadoop и Apache Spark. Apache Hadoop представляет собой платформу распределённого хранения и пакетной обработки больших данных на кластере серверов, включающую распределённую файловую систему HDFS. Apache Spark представляет собой систему распределённых вычислений, обеспечивающую высокоскоростную обработку данных в оперативной памяти и выполнение операций соединения разнородных источников.

Аналитический доступ к подготовленным данным реализован средствами двух систем. Apache Impala обеспечивает выполнение интерактивных SQL-запросов непосредственно поверх распределённого хранилища HDFS в единой экосистеме с Hadoop и Spark. ClickHouse относится к столбцовым аналитическим системам управления базами данных и обеспечивает высокую скорость запросов к витринам данных. Такое сочетание обеспечивает как интерактивный анализ данных на месте их хранения, так и быструю выдачу по подготовленным витринам.

Слой визуализации и поддержки принятия управленческих решений реализован средствами Apache Superset – платформой с открытым исходным кодом для построения интерактивных аналитических панелей и визуализации данных по показателям их состояния. Выбор инструментов с открытым исходным кодом обусловлен требованиями воспроизводимости и независимости реализации от проприетарных решений.

Соответствие узлов контура и компонентов стека приведено в таблице 1.

Этапы контура и компоненты инструментального стека*Источник: составлена автором*

Этап контура	Реализуемая функция управления	Компонент стека
Сбор данных	Наблюдение состояния организационной системы	Apache Hadoop
Очистка и трансформация данных	Подготовка данных к оценке состояния	Apache Spark
Контроль качества	Идентификация состояния	Apache Spark
Загрузка и аналитическое хранение	Фиксация результата	HDFS / Apache Impala, ClickHouse
Визуализация и поддержка решений	Представление состояния ЛПР	Apache Superset

Apache Spark в предлагаемой архитектуре рассматривается как основное средство обработки данных в объёмах, характерных для государственных информационных систем. Демонстрация выполнена на полном наборе данных средствами Python (pandas, NumPy) в среде Google Colab, логика проверок не зависит от средства исполнения и переносится на Apache Spark. В качестве материалов для демонстрации работы контура использованы открытые данные органов публичной власти, размещённые на портале открытых данных. Использование открытых данных обеспечивает воспроизводимость и не затрагивает сведения ограниченного доступа. Демонстрация предполагает прохождение выбранного набора данных через все этапы контура с фиксацией показателей качества на этапе контроля и замыканием обратной связи по результатам этого контроля, что рассматривается в разделе «Результаты».

Результаты. Работоспособность предложенного контура и замыкание обратной связи по результатам контроля качества показаны на открытых данных органа публичной власти. В качестве материала использован набор «Адресный реестр объектов недвижимости города Москвы», размещённый на портале открытых данных Правительства Москвы (data.mos.ru, набор 60562, версия 3, выпуск 681, дата обращения: 21.07.2026; набор сформирован в соответствии с Положением об Адресном реестре объектов недвижимости города Москвы, утверждённым постановлением Правительства Москвы от 26.02.2016 № 58-ПП). Объём выгрузки составил 548 408

записей по 37 полям, что обеспечивает представительность демонстрации на реальных данных без обращения к сведениям ограниченного доступа.

Набор проведён через все этапы контура. На этапе сбора данные получены с портала в виде выгрузки, при этом уже на входе выявлена особенность формата источника: двухуровневый заголовок (техническое имя поля и его текстовое описание), требующий корректной обработки при получении данных. На этапах очистки и трансформации выполнены приведение схемы и подготовка полей к оценке состояния. Узел контроля качества выполнил идентификацию состояния набора по показателям полноты, согласованности, уникальности и целостности, опираясь на нормативные требования к атрибутам адреса.

Нормативной основой для оценки послужили требования к государственному адресному реестру. В соответствии с Федеральным законом от 28.12.2013 № 443-ФЗ обеспечение полноты и достоверности сведений об адресах является принципом ведения реестра, а каждому объекту адресации, согласно Правилам присвоения, изменения и аннулирования адресов (постановление Правительства РФ от 19.11.2014 № 1221), должен быть присвоен уникальный адрес; уникальный номер записи в реестре устанавливается порядком его ведения (приказ Минфина России от 31.03.2016 № 37н). Полнота при этом рассматривается как формальная характеристика качества данных в соответствии с ГОСТ Р ИСО 8000. Указанные требования задают эталонное состояние,

относительно которого узел контроля качества выявляет отклонения. Контроль качества выявил в наборе несколько типов дефектов,

различных по природе. Результаты поиска дефектов приведены в таблице 2.

Таблица 2

Дефекты, выявленные узлом контроля качества

Источник: составлена автором

Тип дефекта	Объём	Нормативное / формальное основание
Несогласованность пары «номер – дата регистрации» ФИАС (есть номер, отсутствует дата)	104 413 записей	443-ФЗ (ст. 4); порядок ведения реестра
Неполнота: объект имеет адрес, но не имеет номера ФИАС	5 219 записей	ПП РФ № 1221 (обязательность адреса)
Нарушение уникальности номера ФИАС (один номер у нескольких разных объектов)	701 запись	ПП РФ № 1221 (уникальность)
Логическая аномалия: дата регистрации при отсутствии номера	5 записей	443-ФЗ (ст. 4)
Повреждённый идентификатор (некорректная длина номера)	1 запись	ГОСТ Р ИСО 8000 (целостность)

Ключевым результатом контроля стало не только выявление дефектов, но и различие их природы, что и отличает контур управления от линейного конвейера. Так, при анализе неполноты номеров ФИАС узел контроля разделил записи по дате регистрации адреса относительно даты вступления в силу Правил (2 декабря 2014 года): из 5 219 записей без номера лишь 83 относятся к адресам, присвоенным до этой даты (для которых отсутствие номера объяснимо переходным периодом), тогда как 5 136 записей зарегистрированы позднее и представляют собой действительный дефект актуальных данных, а не историческое наследие. Аналогично, при контроле уникальности выявленные повторы номеров отнесены преимущественно к однотипным объектам (здание – здание), что не подпадает под предусмотренное Правилами исключение для пары «земельный участок–здание» и, следовательно, квалифицируется как дефект.

По результатам идентификации состояния замкнута обратная связь с адресацией воздействия по источнику отклонения. Дефекты схемы (двухуровневый заголовок, приведение имён полей) отнесены к обработке внутри контура и устранены на этапах очистки и трансформации; повторный контроль подтвердил их отсутствие. Дефекты, обусловленные данными на стороне источника (отсутствие и несогласованность номеров ФИАС, нарушение уникальности), не подлежат восстановлению средствами обработки и адресованы этапу сбора как требование к поставщику данных на последующих периодах; при этом значения не достраиваются искусственно, а помечаются и локализуются. Тем самым обратная связь реализована адресно: тип воздействия определяется источником выявленного дефекта, что соответствует структуре контура, приведённой на рисунке 1.

Результаты демонстрации подтверждают, что предложенный контур реализуем на открытом инструментальном стеке и обеспечивает замыкание обратной связи по состоянию данных. Контроль качества идентифицирует отклонения, различает их природу и инициирует адресное корректирующее воздействие. Вместе с тем демонстрация носит иллюстративный характер и показывает работоспособность контура на одном наборе данных; количественная оценка эффективности управления качеством данных в сравнении с линейной организацией обработки выходит за рамки настоящей работы и составляет направление дальнейших исследований.

Выводы и рекомендации. В настоящем исследовании цикл работы с данными в организационной системе органа публичной власти представлен как замкнутый контур управления с обратной связью по состоянию данных. Такое представление позволяет связать инженерную реализацию обработки данных с управлением организационной системой и переводит задачу из области обработки информации в область управления в организационных системах.

Получены следующие основные результаты.

1. Разработан метод организации цикла работы с данными как замкнутого контура управления, в котором узел контроля качества выполняет идентификацию состояния организационной системы, а обратная связь реализована адресно: тип корректирующего воздействия определяется источником выявленного дефекта.

2. Предложена структура контура, объединяющая внутренний технический контур авторегулирования обработки и внешний управленческий контур поддержки принятия решений.

3. Разработана программно-техническая реализация контура на открытом инструментальном стеке распределённой обработки, аналитического хранения и визуализации данных.

4. Работоспособность контура и замыкание обратной связи показаны на актуальных открытых данных действующего органа публичной власти, где узел контроля качества, опираясь на нормативные требования к атрибутам данных, выявил и различил по природе несколько типов дефектов.

Результаты демонстрации показывают, что отличие контура от линейного конвейера состоит не в выявлении дефектов как таковых, а в различении их природы. Разделение записей без номера ФИАС на 83 записи переходного периода и 5 136 записей актуального периода показывает, что без учёта момента вступления нормы в силу исторические записи были бы ошибочно отнесены к дефектам, а масштаб действительной проблемы искажён. Наиболее массовый дефект, несогласованность пары «номер–дата регистрации» (104 413 записей), по масштабу носит системный характер и не устраняется средствами обработки, поэтому воздействие адресуется этапу сбора. В отличие от модели DaLiF [9], где качество рассматривается как сквозная проверка, предложенный контур связывает результат контроля с ответственным этапом. Вместе с тем адресная обратная связь на поставщика данных формализует требование, но не гарантирует устранения дефекта, поскольку источник находится вне зоны управления органа-обработчика.

Практическая значимость работы состоит в том, что предложенный контур применим в организационных системах ключевых органов государственного управления, работающих с большими объёмами данных, и, в отличие от существующих концептуальных моделей жизненного цикла данных, апробирован на реальных данных. Контур позволяет управлять состоянием данных на регулярной основе и тем самым повышать обоснованность управленческих решений.

По результатам работы органам публичной власти и операторам открытых данных рекомендуется:

- включать в регламенты загрузки проверку согласованности связанных атрибутов, в частности пары «идентификатор–дата регистрации»;
- при контроле качества разделять исторические и актуальные дефекты по дате вступления соответствующей нормы в силу;
- не дораивать отсутствующие значения, а маркировать и локализовать их;
- направлять поставщику данных формализованные требования по дефектам источника;
- публиковать показатели качества вместе с набором данных и описывать нестандартные элементы его структуры.

Работа не лишена ограничений. Проведённая демонстрация носит иллюстративный характер и показывает работоспособность контура на одном наборе данных. Количественная оценка эффективности управления качеством данных в сравнении с линейной организацией обработки в настоящей работе не выполнялась. Дальнейшие исследования

предполагают развитие узла контроля качества за счёт метрик качества данных, распространение контура на задачи формирования условно-реальных (синтетических) данных и идентификации моделей в условиях ограниченного доступа к данным, а также количественную оценку эффективности предложенного подхода.

Декларация о применении ИИ. При подготовке статьи автор использовал генеративную модель искусственного интеллекта (ИИ) Claude Opus 4.8 (Anthropic) в качестве вспомогательного инструмента для подготовки схемы рисунка 1, поиска литературных источников и написания программного кода демонстрации. Постановка задачи, правила проверок качества данных, интерпретация результатов и выводы сформулированы автором. Найденные источники и схема рисунка проанализированы автором и проверены на соответствие описанию метода. Код выполнен и проверен автором, результаты таблицы 2 воспроизводимы вручную. Код и промпты сохранены и предоставляются по запросу. Автор несёт полную ответственность за содержание статьи.

Список источников

1. Пенчева И. и др. Большие данные и ИИ – трансформационный сдвиг для государственного управления: что дальше ждет исследования? // *Public Policy and Administration*. – 2020. – Т. 35. – №. 1. – С. 24–44. – DOI: 10.1177/0952076718780537. – Текст: электронный. (In Eng.).
2. Янссен М., Броус П., Эстевес Э., Барбоса Л.С., Яновский Т. Управление данными: организация данных для обеспечения надежности искусственного интеллекта // *Government Information Quarterly*. – 2020. – Т. 37. – № 3. – Ст. 101493. – DOI: 10.1016/j.giq.2020.101493. – Текст: электронный. (In Eng.).
3. Бхупатираджу С. и др. Государственная аналитика с использованием данных о качестве административных процессов // *The Government Analytics Handbook*. – 2023. – С. 345–366. – DOI: 10.1596/978-1-4648-1957-5. – Текст: электронный. (In Eng.).
4. Аристова Д. А., Макеева Е. З., Федорова О. В. Эффекты внедрения интеллектуальных транспортных систем // *Транспортное дело России*. – 2022. – № 1. – С. 114–116. – DOI: 10.52375/20728689_2022_1_114. – Текст: электронный.
5. Джонсон Т. Л. и др. Применение систем распознавания лиц в полиции и борьба с насильственными преступлениями в городах США // *Cities*. – 2024. – Т. 155. – Ст. 105472. – DOI: 10.1016/j.cities.2024.105472. – Текст: электронный. (In Eng.).
6. Фаттах И. А. и др. Взаимодействие возможностей бизнес-аналитики и эффективности принятия решений в государственном секторе Индонезии // *Digital Business*. – 2025. – Т. 5. – № 2. – DOI: 10.1016/j.digbus.2025.100132. (In Eng.).

References

1. Pencheva I., Esteve M., Mikhaylov S.J. Big Data and AI – A Transformational Shift for Government: So, What Next for Research? *Public Policy and Administration*. 2020. Vol. 35. No. 1. pp. 24–44. DOI: 10.1177/0952076718780537.
2. Janssen M., Brous P., Estevez E., Barbosa L.S., Janowski T. Data Governance: Organizing Data for Trustworthy Artificial Intelligence. *Government Information Quarterly*. 2020. Vol. 37. No. 3. Art. 101493. DOI: 10.1016/j.giq.2020.101493.
3. Bhupatiraju S., Chen D., Jankin S., Kim G., Kupi M., Ramos Maqueda M. Government Analytics Using Data on the Quality of Administrative Processes. *The Government Analytics Handbook*. 2023. pp. 345–366. DOI: 10.1596/978-1-4648-1957-5.
4. Aristova D. A., Makeeva E. Z., Fedorova O. V. Effects of the Implementation of Intelligent Transport Systems. *Transportnoe delo Rossii*. 2022. No. 1. pp. 114–116. DOI: 10.52375/20728689_2022_1_114. (In Russ.).
5. Johnson T. L., Johnson N. N., Topalli V., McCurdy D., Wallace A. Police Facial Recognition Applications and Violent Crime Control in U.S. Cities. *Cities*. 2024. Vol. 155. Art. 105472. DOI: 10.1016/j.cities.2024.105472.
6. Fattah I.A., Prabowo H., Tjhin V.U., Rahim R.K. The Interplay Between Business Analytics Capabilities and Decision-Making Performance in Indonesia's Public Sector. *Digital Business*. 2025. Vol. 5. No. 2. DOI: 10.1016/j.digbus.2025.100132.
7. Afieroho P., Perkins R., Zhou X., Hoanca B., Protasel G. Adopting E-Government to Monitor Public Infrastructure Projects Execution in Nigeria:

7. Афиерохо П. и др. Внедрение электронного правительства для мониторинга выполнения проектов общественной инфраструктуры в Нигерии: взгляд общественности // *Heliyon*. – 2023. – Т. 9. – № 8. – Ст. e18552. – DOI: 10.1016/j.heliyon.2023.e18552. – Текст: электронный. (In Eng.).
8. Лян Ю. и др. Обзор исследований поставщиков данных с точки зрения государственного управления: библиометрический анализ с использованием CiteSpace // *Journal of Big Data*. – 2025. – Т. 12. – Ст. 248. – DOI: 10.1186/s40537-025-01294-4. – Текст: электронный. (In Eng.).
9. Шах С. И. Х. и др. DaLiF: Структура жизненного цикла данных для правительств, ориентированных на данные // *Journal of Big Data*. – 2021. – Т. 8. – Ст. 89. – DOI: 10.1186/s40537-021-00481-3. – Текст: электронный. (In Eng.).
10. Акаткин Ю. М., Ясиновская Е. Д. Ориентированность на данные как ключевой фактор цифрового правительства: готова ли Россия к цифровой трансформации государственного сектора? // *Electronic Governance and Open Society: Challenges in Eurasia (EGOSE 2018)*. Cham: Springer, 2019. – Т. 947. – С. 439–454. – DOI: 10.1007/978-3-030-13283-5_33. – Текст: электронный. (In Eng.).
11. Новиков Д. А., Чхартишвили А. Г. Рефлексия и управление: математические модели. – Leiden: CRC Press. Balkema, 2014. – 298 с. – DOI: 10.1201/b16625. – Текст: электронный. (In Eng.).
12. Фаннуш А. и др. Совершенствование практик DataOps с помощью инновационных моделей сотрудничества: систематический обзор // *International Journal of Information Management Data Insights*. – 2025. – Т. 5. – № 1. – Ст. 100321. – DOI: 10.1016/j.jjime. 2025.100321. – Текст: электронный. (In Eng.).
13. Мехмуд Э., Анис Т. Распределенная архитектура ETL в реальном времени для неструктурированных больших данных // *Knowledge and Information Systems*. – 2022. – Т. 64. – № 12. – С. 3419–3445. – DOI: 10.1007/s10115-022-01757-7. – Текст: электронный. (In Eng.).
14. Абрахам Р. и др. Управление данными: концептуальная основа, структурированный обзор и программа исследований // *International Journal of Information Management*. – 2019. – Т. 49. – С. 424–438. – DOI: 10.1016/j.ijinfomgt.2019.07.008. – Текст: электронный. (In Eng.).
15. Левитин А. В., Редман Т. К. Модель жизненных циклов данных применительно к качеству // *Information and Software Technology*. – 1993. – Т. 35. – № 4. – С. 217–223. – DOI: 10.1016/0950-5849(93)90069-F. – Текст: электронный. (In Eng.).
16. Бурков В. Н., Губко М., Коргин Н., Новиков Д. Введение в теорию управления в организациях. – Boca Raton: CRC Press, 2015. – 346 с. – DOI: 10.1201/b18152. – Текст: электронный. (In Eng.).
- The Public Perspective. *Heliyon*. 2023. Vol. 9. No. 8. Art. e18552. DOI: 10.1016/j.heliyon.2023.e18552.
8. Liang Y., You J., Wang R., Wang X., Dong L., Lyu M. Data Provider Research Overview from a Public Management Perspective: A Bibliometric Analysis Utilizing CiteSpace. *Journal of Big Data*. 2025. Vol. 12. Art. 248. DOI: 10.1186/s40537-025-01294-4.
9. Shah S. I. H., Peristeras V., Magnisalis I. DaLiF: A Data Lifecycle Framework for Data-Driven Governments. *Journal of Big Data*. 2021. Vol. 8. Art. 89. DOI: 10.1186/s40537-021-00481-3.
10. Akatkin Yu. M., Yasinovskaya E. D. Data-Centricity as the Key Enabler of Digital Government: Is Russia Ready for Digital Transformation of Public Sector. *Electronic Governance and Open Society: Challenges in Eurasia (EGOSE 2018)*. Cham: Springer, 2019. CCIS. Vol. 947. pp. 439–454. DOI: 10.1007/978-3-030-13283-5_33.
11. Novikov D. A., Chkhartishvili A. G. Reflexion and Control: Mathematical Models. *Leiden: CRC Press. Balkema*. 2014. 298 p. DOI: 10.1201/b16625. (In Eng.).
12. Fannouch A., Gharib J., Gahi Y. Enhancing DataOps Practices Through Innovative Collaborative Models: A Systematic Review. *International Journal of Information Management Data Insights*. 2025. Vol. 5. No. 1. Art. 100321. DOI: 10.1016/j.jjime.2025.100321.
13. Mehmood E., Anees T. Distributed Real-Time ETL Architecture for Unstructured Big Data. *Knowledge and Information Systems*. 2022. Vol. 64. No. 12. pp. 3419–3445. DOI: 10.1007/s10115-022-01757-7.
14. Abraham R., Schneider J., vom Brocke J. Data Governance: A Conceptual Framework, Structured Review, and Research Agenda. *International Journal of Information Management*. 2019. Vol. 49. pp. 424–438. DOI: 10.1016/j.ijinfomgt.2019.07.008.
15. Levitin A. V., Redman T. C. A Model of the Data (Life) Cycles with Application to Quality. *Information and Software Technology*. 1993. Vol. 35. No. 4. pp. 217–223. DOI: 10.1016/0950-5849(93)90069-F.
16. Burkov V. N., Goubko M., Korgin N., Novikov D. Introduction to Theory of Control in Organizations. *Boca Raton: CRC Press*, 2015. 346 p. DOI: 10.1201/b18152.

Шенцов Данил Иванович / Shentsov Danil I.

старший преподаватель / Senior Lecturer

Национальный исследовательский университет «Высшая школа экономики» / National Research University Higher School of Economics

г. Москва, Покровский бульвар, д. 11

E-mail: dshentsov@hse.ru

Марусина Мария Яковлевна / Marusina Maria Ya.

доктор технических наук, профессор / D.Sc., Professor

профессор факультета систем управления и робототехники / Professor at the Faculty of Control Systems and Robotics

федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО» / ITMO University

Санкт-Петербург, Кронверкский пр., д. 49, лит. А

E-mail: myamarusina@itmo.ru