

- Отсутствие сред и инструментов разработчиков, позволяющих удобно разрабатывать именно распределённые системы

- Инструменты моделирования и тестирования также следует развивать для упрощения разработки и диагностики проблем в распределённых ИС.

Но для упрощения их разработки и поддержки стоит решить ряд проблем, которые замедляют или усложняют данные процессы.

Список литературы

1. Шамакина А.В. Обзор технологий распределённых вычислений // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2014. № 3. URL: <https://cyberleninka.ru/article/n/obzor-tehnologiy-raspredeennyh-vychisleniy> (дата обращения: 15.02.2025).
2. Радченко Г.И. Распределённые вычислительные системы. Челябинск: Фотохудожник, 2012. 184 с.
3. Филимонов Д.Н. Обзор решений в области распределённых вычислений // Молодой ученый. 2020. № 19 (309). С. 145-148. URL: <https://moluch.ru/archive/309/69713/> (дата обращения: 15.02.2025).
4. Сормов С.И. Обработка информации в распределённых системах // Современные материалы, техника и технологии. 2015. № 2 (2). URL: <https://cyberleninka.ru/article/n/obrabotka-informatsii-v-raspredeennyh-sistemah> (дата обращения: 15.02.2025).
5. Мельникова Т.В., Питолин М.В., Преображенский Ю.П. Моделирование обработки больших массивов данных в распределённых информационно-телекоммуникационных системах // Моделирование, оптимизация и информационные технологии. 2022. № 10(1). URL: <https://moitvvt.ru/ru/journal/pdf?id=1117> (дата обращения: 15.02.2025).

КОНЦЕПЦИЯ ПОСТРОЕНИЯ ПРОГРАММНО-АППАРАТНОЙ ПЛАТФОРМЫ НАУЧНЫХ ИССЛЕДОВАНИЙ В ОБРАЗОВАТЕЛЬНОЙ ОРГАНИЗАЦИИ

Данилин М.Д., Шестопалова Д.А.

*Российский экономический университет
им. Г. В. Плеханова, Москва,
e-mail: danilinbox2@yandex.ru*

Научный руководитель: Микрюков А.А.

Сегодня сбор и обработка данных является повсеместным процессом. В России в 2024 году стартовал национальный проект «Экономика данных и трансформация государственного управления», направленный на развитие экономики на основе данных. Он станет продолжением завершающегося национального проекта «Цифровая экономика». Одной из его целей является сбор данных из различных сфер, таких как промышленность, транспорт, телемедицина и наука, и их последующая обработка с использованием искусственного интеллекта для выявления тенденций, аномалий и проблем [1].

Новые цифровые стандарты и стратегии в экономике данных требуют подготовки специалистов, способных работать с большими данными, аналитикой и искусственным интеллектом и формируют потребность в адаптации

образовательной среды для подготовки таких специалистов [2]. Для соответствия современным реалиям образовательные учреждения комплексно развивают свою образовательную среду по трем направлениям: использование личностных факторов студента, создание материальных и социальных условий обучения в контексте преподавания отдельных дисциплин, а также внедрение концепции сетевого образования без границ [3]. В особенности, развитие в образовательной среде направлено на формирование личностей, готовых к реализации в различных областях: творческой, исследовательской, образовательной и других [4].

Для поддержки современных тенденций мы предлагаем создание платформы для работы над Data Science проектами университета. Это может быть информационный ресурс, на котором в основном студенты направлений Data Science и Machine Learning смогут практиковаться в решении исследовательских задач на данных, собранных в научной и образовательной среде института. Платформы такого формата уже существуют, например, Kaggle, GitHub, Pew Research и другие [5]. Преподаватели и студенты профильных направлений уже используют их в качестве источника данных для анализа. Научная новизна заключается в том, что мы предлагаем использовать концепцию такой платформы в контексте развития образовательной и научной среды университета.

Описание концепции

В образовательной среде университета платформа для обмена данными объединяет широкий круг заинтересованных сторон (рисунок 1). Ситуационные центры, научные лаборатории, отраслевые подразделения и сама образовательная организация предоставляют данные о результатах исследований, маркетинговых трендах, загрузке аудиторий и успеваемости. Студенты и преподаватели также могут делиться результатами своих исследований – от структурированных баз успеваемости и посещаемости до неструктурированных текстовых отчетов, полуструктурированных логов и временных рядов, что позволяет проводить полный цикл анализа данных.

Такое сотрудничество стимулирует развитие образовательной и научной среды: студенты получают возможность работать над реальными проектами и создавать ценные портфолио, преподаватели – находить подтверждения своим идеям, а университет усиливает научный рейтинг и конкурентное преимущество. Платформа также способствует установлению новых профессиональных связей, выявлению талантов и сотрудничеству с компаниями-партнёрами, использующими данные для исследований и коммерческих целей. Концепция описана подробнее в предыдущей статье.

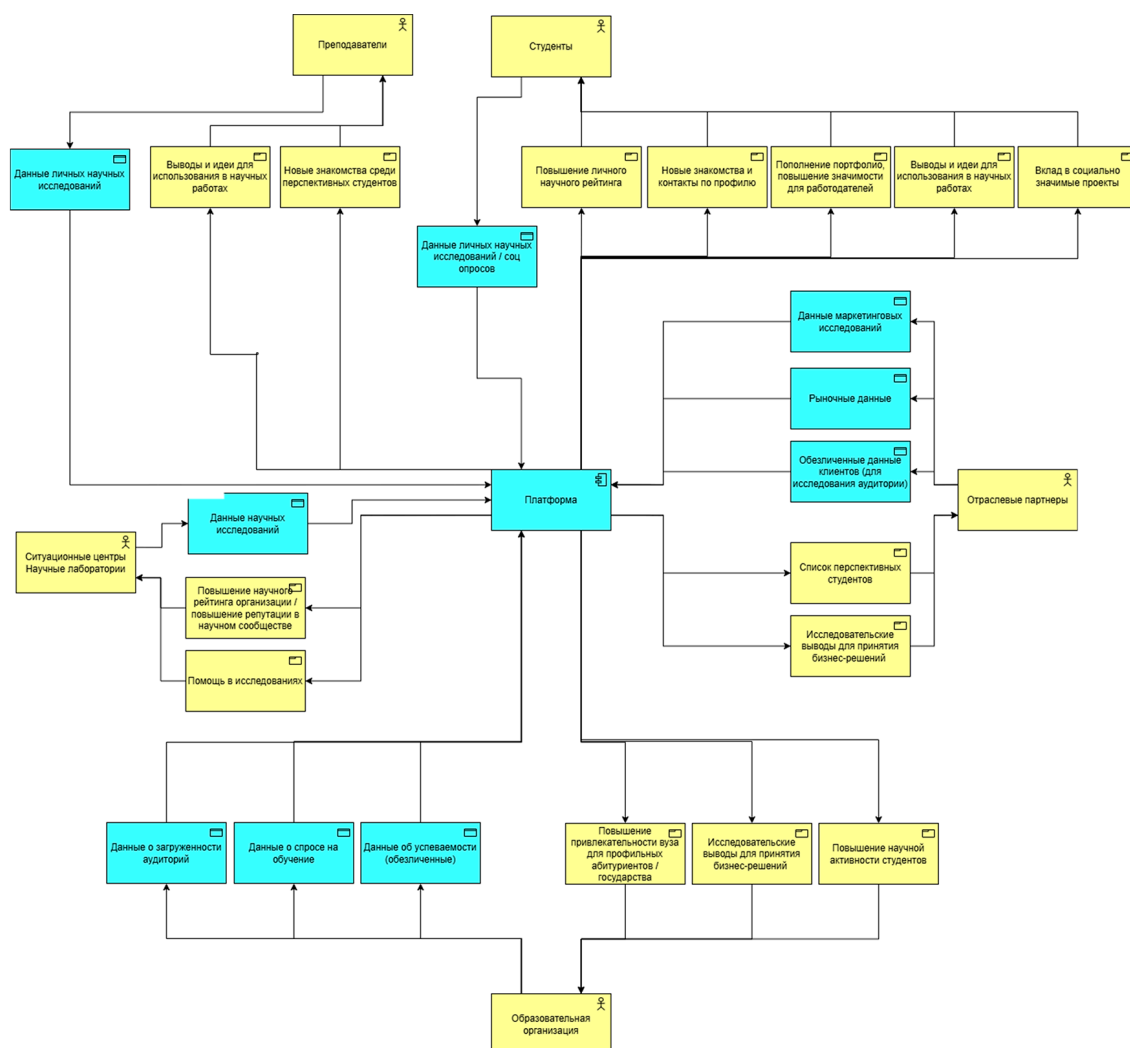


Рис. 1. Концепция использования платформы заинтересованными лицами

Аппаратно-программная архитектура

Создание платформы с технической точки зрения – это интеграция множества компонентов, и в современных условиях развития облачных технологий можно использовать гибридный подход к построению такого рода систем. Оптимальным решением может стать использование микросервисной архитектуры, развернутой на облачной платформе, такой как Yandex Cloud или VK Cloud. Применение готовых облачных сервисов позволит упростить развертывание и поддержку микросервисов, а также обеспечит гибкую масштабируемость всех компонентов системы. Интеграция локальных источников данных и систем по API позволит обеспечить требуемый уровень функциональности и интеграции между внешними и внутренними системами. Кроме того, Yandex Cloud предоставляет гранты и особые условия для образовательных учреждений, что делает

его привлекательным выбором для подобных проектов [6]. Пример такой архитектуры представлен на рисунке 2.

Модуль загрузки данных - один из основных в системе. При разработке модуля загрузки данных платформа должна поддерживать работу с различными форматами, включая CSV, JSON и SQL-данные. Для CSV важно учитывать кодировки, разделители и пропущенные значения, а для JSON – обработку вложенных структур. Использование ETL-инструментов (как на базе Yandex Cloud, так и пользовательских) позволит автоматически очищать и нормализовывать данные. SQL-данные могут загружаться либо через прямые подключения к базам, либо в виде дампов. Пользовательский интерфейс должен позволять выбирать источник данных, настраивать параметры загрузки и получать обратную связь о статусе. Такой подход обеспечит гибкость и удобство работы с данными для студентов и преподавателей.

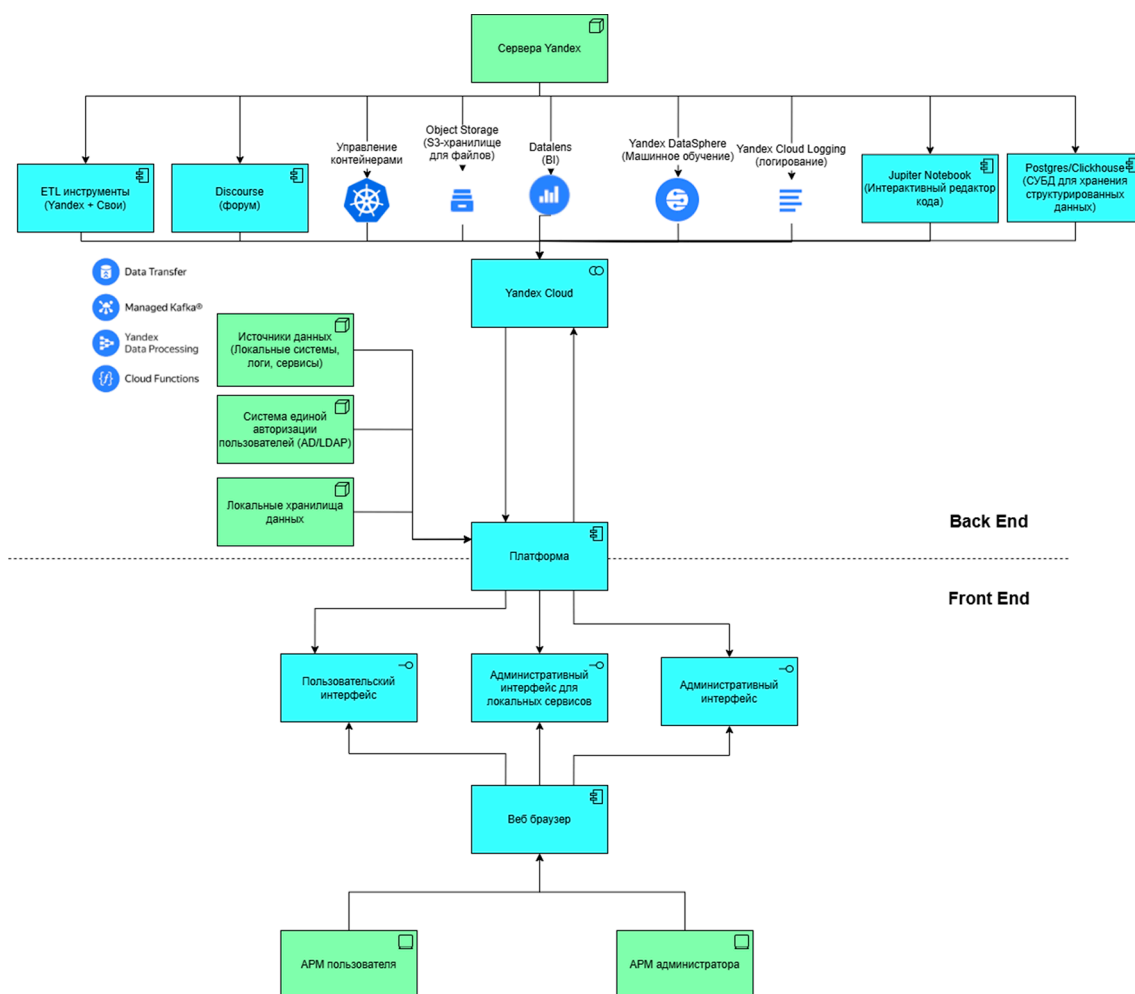


Рис. 2. Архитектура инфраструктурных компонентов для реализации платформы

Для хранения исследовательских данных в SQL-форматах в системе предусмотрено использование реляционной СУБД PostgreSQL и аналитической СУБД ClickHouse. Эти решения обеспечивают высокую производительность, отказоустойчивость и легко интегрируются с другими компонентами платформы. Сервисы Yandex Cloud, такие как Managed Service for PostgreSQL и Managed Service for ClickHouse, позволяют автоматически масштабировать базы данных, обеспечивая их стабильную работу при росте нагрузки.

Для хранения данных, включая исследовательские наборы и пользовательские вложения, на платформе предусмотрена файловая система, совместимая с протоколом S3. В качестве основного решения может использоваться Yandex Object Storage, обеспечивающий надежное и масштабируемое хранение данных. Однако, при необходимости, возможно применение альтернативных решений, таких как MinIO или другие S3-совместимые хранилища, что позволяет гибко настраивать инфраструктуру в зависимости от требований проекта.

Модуль форума будет построен на Open-source платформе Discourse, которую используют OpenAI, GitLab, Docker и другие. Это масштабируемое решение с системой категорий, тем и тегов, инструментами модерации и управления доступом. Discourse поддерживает плагины, позволяющие добавить работу с SQL-таблицами, Jupyter Notebook и другие функции. Встроенный REST API обеспечивает интеграцию с сервисами университета, упрощая автоматизацию и масштабирование [7]. Пакет discourse-machine-learning предоставляет инструменты для анализа данных и машинного обучения: загрузку и структурирование данных, аналитику обсуждений, персонализированные рекомендации и автоматизацию модерации [8].

Для задач машинного обучения на платформе могут быть использованы инструменты Yandex DataSphere, которые можно дополнить стандартными средствами обработки данных на Python в среде Jupyter Notebook. Интеграция Jupyter Notebook позволит пользователям редактировать и запускать код непосредственно

на платформе, создавать интерактивные блокноты, а также делиться ими с другими участниками через форум, обеспечивая удобную совместную работу и воспроизводимость исследований.

Для визуализации данных на платформе, помимо стандартных инструментов Python, таких как Matplotlib, Seaborn и Plotly, может быть интегрирована полнофункциональная BI-система, например, Yandex DataLens. Этот инструмент позволяет создавать интерактивные дашборды, анализировать данные в режиме реального времени и работать с различными источниками данных. При необходимости можно использовать альтернативные Open-source решения, такие как Metabase, Superset или Redash, обеспечивающие гибкость в выборе инструментов и возможность локального развертывания [9].

Использование Kubernetes в рамках платформы позволяет организовать оркестрацию контейнеризированных микросервисов, обеспечивая высокую масштабируемость, отказоустойчивость и гибкость всей системы. Это особенно актуально для высоконагруженных сервисов, где требуется динамическое распределение рабочих нагрузок и автоматическое масштабирование компонентов в зависимости от текущих требований. Yandex Cloud предоставляет управляемый сервис Kubernetes, который значительно упрощает развертывание и администрирование кластеров, автоматизируя такие процессы, как обновление, балансировка нагрузки и восстановление после сбоев.

Система логирования необходима для мониторинга работы платформы, быстрого выявления ошибок и анализа производительности. Она фиксирует все ключевые события, включая загрузку данных, выполнение кода, аутентификацию пользователей и обращения к API, что позволяет оперативно реагировать на сбои и предотвращать потенциальные проблемы. В экосистеме Yandex Cloud существует встроенный сервис логирования, интегрированный с другими облачными решениями, такими как Yandex Monitoring и Yandex Data Streams. Он поддерживает сбор, хранение и анализ логов в режиме реального времени, а также позволяет подключать внешние приложения через API, что делает его удобным решением для централизованного управления логами на платформе.

В университете уже имеются ключевые инфраструктурные компоненты, которые обеспечивают работоспособность платформы. Например, система единой авторизации пользователей, которая уже используется для всех сервисов университета, может быть интегрирована с платформой для управления правами доступа. Это позволит унифицировать процессы аутентификации и авторизации, обеспечив удобство использования платформы и повышенную безопасность. Такой подход минимизирует потребность в дополнительных механизмах

управления доступом и упрощает техническую поддержку системы, снижая затраты на разработку и интеграцию. Также платформу можно интегрировать с уже имеющимися локальными СУБД и хранилищами данных для автоматизации пополнения платформы данными.

Для доступа к ресурсам платформы будет разработан централизованный пользовательский интерфейс, который обеспечит удобную навигацию и доступ ко всем основным функциональным возможностям: использование форума, загрузка данных, интерактивные элементы визуализации, написание и выполнение кода и другие функции. Этот интерфейс будет обеспечивать интуитивно понятное взаимодействие с платформой для пользователей всех уровней.

Административный интерфейс будет разделен по сервисам: отдельный для облачных сервисов и для локальных компонентов. Такой подход позволяет эффективно управлять различными частями инфраструктуры, при этом поддерживая гибкость в их настройке. В перспективе можно рассмотреть разработку централизованного интерфейса для управления всей платформой или возможный перенос локальных сервисов в облако, что обеспечит дополнительную масштабируемость и упростит техническую поддержку, а также повысит общую устойчивость системы.

Сложности и вызовы

При проработке программно-аппаратной части платформы появляются дополнительные вопросы, которые надо рассмотреть.

Одной из ключевых проблем при использовании облачных сервисов, таких как Yandex Cloud, является риск Vendor Lock – ситуации, при которой зависимость от одного поставщика облачных решений становится критической, ограничивая гибкость инфраструктуры и повышая затраты на миграцию. Это может выражаться в использовании проприетарных API, несовместимых с другими платформами, высоких комиссиях за вывод данных или отсутствии альтернативных инструментов с аналогичным функционалом. Чтобы минимизировать эти риски, необходимо проектировать систему с возможностью переноса данных и вычислительных процессов на другие облака или локальные мощности. Отчасти Yandex Cloud уже обеспечивает совместимость, например, для хранилищ S3, а Kubernetes разворачивается на аналогичных провайдерах без дополнительных усилий. Важно применять открытые стандарты и кросс-совместимые технологии для баз данных, таких как PostgreSQL или MySQL, вместо проприетарных решений. Такой гибридный подход снижает зависимость от одного провайдера, позволяя в случае необходимости масштабироваться, менять инфраструктуру и оптимизировать расходы.

Другой сложностью при реализации и поддержке платформы является разрозненность используемых сервисов. Наличие нескольких независимых компонентов, как облачных, так и локальных, требует значительных усилий для их интеграции, управления и поддержки. Это может привести к трудностям с мониторингом, обновлением, а также увеличению сложности при решении проблем, возникающих на различных уровнях инфраструктуры. В долгосрочной перспективе такая архитектура требует значительных затрат на поддержание совместимости сервисов и обеспечение их стабильной работы. Для решения этой проблемы можно рассмотреть возможность разработки централизованной системы управления, которая обеспечивала бы единую точку мониторинга и администрирования всех сервисов платформы. Это позволит упростить процессы обслуживания и сократить время на решение возникающих проблем. Также стоит проработать возможность постепенного переноса локальных сервисов в облако, что обеспечит более простое масштабирование, лучшую интеграцию компонентов и снизит операционные затраты.

Поскольку на платформе предполагается работа с большими и разнообразными данными, что может стать проблемой в контексте их качества и согласованности, потребуется разработать системы для автоматической очистки, проверки и стандартизации данных перед их загрузкой на платформу, а также для мониторинга и коррекции ошибок в процессе их использования.

Актуальной остается и проблема защиты пользовательских данных. Платформа будет работать с большими объемами исследовательских данных, что требует высокого уровня безопасности. Помимо стандартных мер защиты, важно будет предусмотреть механизмы шифрования данных на всех этапах их хранения и передачи, а также обеспечить надежный контроль доступа на основе ролей.

Выводы

Результаты проведенного исследования показали, что современные подходы и технологии разработки информационных систем позволяют реализовать идею цифровой платформы научных исследований в образовательной организации. В ходе работы были рассмотрены ключевые компоненты платформы, такие как модуль форума, инструменты работы с данными, а также возможность интеграции с облачными сервисами и использование Kubernetes для оркестрации микросервисов. Также предложены решения для масштабируемости и обеспечения безопасности данных, а также интеграции с существующими системами университета. Использование таких сервисов, как Yandex Cloud и их инструменты для обработки данных, визуализации и машинного обучения, существенно

ускоряют внедрение платформы и снижают затраты на разработку.

Тем не менее, несмотря на найденные решения, существует ряд проблем, требующих дальнейшей проработки. Среди них – интеграция разрозненных сервисов, обеспечение надежности и безопасности системы, а также решение вопросов, связанных с качеством и масштабируемостью данных. Дополнительное внимание следует уделить развитию системы централизованного управления и совершенствованию поддержки пользователей. В дальнейшем необходима детальная проработка этих аспектов для создания устойчивой и эффективной платформы, которая будет способствовать развитию научных исследований и образовательного процесса в университете.

Список литературы

1. В России появится новый нацпроект – «Экономика данных». 2023. [Электронный ресурс]. URL: <https://digital.gov.ru/ru/events/45686/> (дата обращения: 26.02.2025).
2. Специалисты по Data Science: основные навыки и спрос работодателей. HeadHunter. [Электронный ресурс]. URL: <https://hh.ru/article/27128> (дата обращения: 26.02.2025).
3. Ziyatdinova Yu.N. Innovative Educational Environment of a Research University // Modern problems of science and education. 2015. № 3 P. 463–463. URL: <https://science-education.ru/en/article/view?id=20087&ysclid=m2kokdms7w884904052> (дата обращения: 26.02.2025).
4. Андреева Е. А. Создание научно-исследовательской и инновационно-образовательной среды в вузах как одно из условий формирования образовательного и исследовательского пространства // Общество: социология, психология, педагогика. 2014. № 4. URL: <https://cyberleninka.ru/article/n/sozдание-nauchnoissledovatel'skoy-i-innovatsionno-obrazovatel'noy-sredy-v-vuzah-kak-odno-iz-usloviy-formirovaniya-obrazovatel'nogo-i> (дата обращения: 26.02.2025).
5. Где брать данные для анализа и машинного обучения: бесплатно и удобно. [Электронный ресурс]. URL: <https://thecode.media/5-big-data/> (дата обращения: 26.02.2025).
6. Программа поддержки цифровизации образовательных учреждений. HeadHunter. [Электронный ресурс]. URL: <https://yandex.cloud/ru/digitalization-of-universities> (дата обращения: 25.02.2025).
7. Discourse. GitHub. [Электронный ресурс]. URL: <https://github.com/discourse/discourse> (дата обращения: 20.02.2025).
8. Discourse-machine-learning. GitHub. [Электронный ресурс]. URL: <https://github.com/swamidass/discourse-machine-learning> (дата обращения: 20.02.2025).
9. 12 лучших open source BI инструментов – выбор профессионалов. BI Consult. [Электронный ресурс]. URL: <https://datafinder.ru/products/12-luchshih-open-source-bi-instrumentov-vybor-professionalov> (дата обращения: 26.02.2025).

ПРИНЦИП «ЕДИНОЕ ОКНО» ПРИ ОКАЗАНИИ УСЛУГ В ОБЛАСТИ ОБЕСПЕЧЕНИЯ ЕДИНСТВА ИЗМЕРЕНИЙ (НА ПРИМЕРЕ ДЕЯТЕЛЬНОСТИ ФБУ «ОРЕНБУРГСКИЙ ЦСМ»)

Денисова В.А., Третьяк Л.Н.

ФГБОУ ВО «Оренбургский государственный университет», Оренбург,
e-mail: denisovav960@gmail.com

Качество предоставляемых услуг различными организациями – важная составляющая удовлетворенности потребителя (клиента). Обе-