**Глава 1. Управление данными информационно-аналитической системы поддержки исследования микробиома озера Байкал**

**Краткий результат главы.** Сформирована формальная схема типа «сущность-связь» информационно-аналитической системы поддержки исследования микробиома озера Байкал (далее ИАС) на предметно-ориентированном языке JHipster Domain Language (далее JDL) (<https://www.jhipster.tech/jdl>). На основе подготовленной схемы спроектирована микро-сервисная архитектура ИАС, в которой взаимодействие серверной и клиентской частей организовано посредством сетевых коммуникационных протоколов в архитектурном стиле «Representational State Transfer» (REST) (<https://restfulapi.net>). Данная архитектура имплементирована с помощью комбинирования современных инструментальных платформ разработки «back-end» и «front-end» программного обеспечения (далее ПО). Серверная часть реализована на основе свободной платформы «Spring Framework» (<https://spring.io>), а клиентская часть — на основе платформы «Angular» ([https://angular.io](https://angular.io/)). Реализованная архитектура позволяет интегрировать сторонние микро-сервисы обработки и анализа метагеномных данных в ИАС посредством веб-ориентированного интерфейса прикладного программирования (REST API) доступа к базе данных ИАС.

**1.1 Материалы и методы**

Проектирование и реализация архитектуры ИАС выполнена на основе модельно-управляемой разработки программного обеспечения (MDSE — Model-Driven Software Engineering). Данный подход позволил сфокусироваться на проектировании информационной модели предметной области — исследовании микробиома озера Байкал, для генерации частей исходного кода ПО ИАС. Проектирование схемы типа «сущность-связь» ИАС осуществлялось исходя из постановки задачи и предоставленных исследовательских данных ЛИН СО РАН. Для формального представления платформо-независимой модели (PIM) ИАС использовался предметно-ориентированный декларативный язык JDL, а для ее визуализации и редактирования — интегрированная среда разработки JDL-studio ([https://www.jhipster.tech/jdl-studio](https://www.jhipster.tech/jdl-studio/)). Для трансформации исходной модели в платформо-зависимое представление (PSM) с последующей генерацией исходного кода каркаса ПО ИАС использовались инструментальные средства JHipster (<https://www.jhipster.tech>).

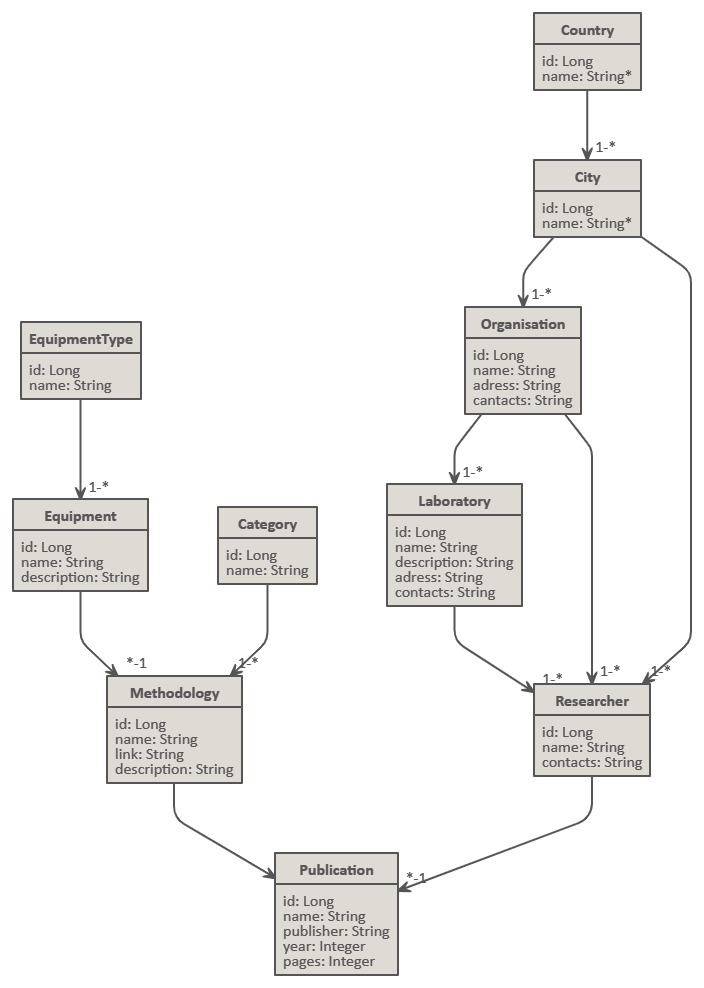
В соответствие со сформулированными общими критериями функционирования веб-ориентированной ИАС для ее разработки был выбран микро-сервисный тип архитектуры, обеспечивающий интеграцию с различными сторонними средствами обработки и анализа метагеномных данных. Для реализации архитектуры использовался следующий стек современных технологий разработки ПО: «Spring Boot» (<https://spring.io/projects/spring-boot>) — платформа разработки микро-сервисов серверной части ИАС; «Hibernate» (<https://hibernate.org/orm>) обеспечивает объектно-реляционное отображение; «JSON Web Token» (<https://jwt.io>) — аутентификация пользователя без сохранения состояния; PostgreSQL (<https://www.postgresql.org>) — реляционная система управления базами данных (далее СУБД) для постоянного хранения объектов модели ИАС; H2 ([https://www.h2database.com](https://www.h2database.com/)) — резидентная СУБД для кэширования первого уровня приложения; Maven (<https://maven.apache.org>) — сборка приложения; Angular — платформа разработки веб-ориентированных приложений.

Исходный код объектной модели ИАС представлен на языке программирования общего назначения Java (<https://www.java.com>) с использованием шаблона проектирования «Model-View-Controller» (далее MVC) для разделения данных приложения, пользовательского интерфейса и управляющей логики. База данных ИАС управляется реляционной СУБД PostgreSQL с обеспечением постоянного хранения и объектно-реляционного отображения. Дополнительное использование современной резидентной СУБД H2 позволило ускорить процесс обработки пользовательских запросов и обеспечить сохранение данных при перезагрузке сессий подключения. Журналирование изменений базы данных ИАС выполняется инструментальными средствами Liquibase (<https://www.liquibase.org>). В качестве Исполнение веб-сервера и контейнера сервелатов используется свободный сервер приложений Jetty (<https://www.eclipse.org/jetty>). Сборка серверной часть приложения ИАС выполняется на базе пакетного файлового формата «Java Archive» (далее JAR).

Веб-ориентированный прикладной программный интерфейс ИАС для доступа к базе данных исследования микробиома разработан на основе архитектуры взаимодействия компонентов распределённого приложения в сети — REST. Интеграция ИАС со сторонними программными средствами обеспечивается по протоколу прикладного уровня передачи данных — «Hypertext Transfer Protocol» (далее HTTP) (<https://www.w3.org/Protocols/rfc2616/rfc2616.html>) через HTTP-запросы типа POST и GET с возможностью сериализации и десериализации объектов базы данных в открытом формате «JavaScript Object Notation» (далее JSON) (<https://www.json.org>). Интерфейс также поддерживает аутентификацию пользователя без сохранения состояния. Клиентская часть разработана в виде настольного пользовательского веб-приложения с помощью платформы Angular на скриптовом языке программирования TypeScript ([https://www.typescriptlang.org](https://www.typescriptlang.org/)).

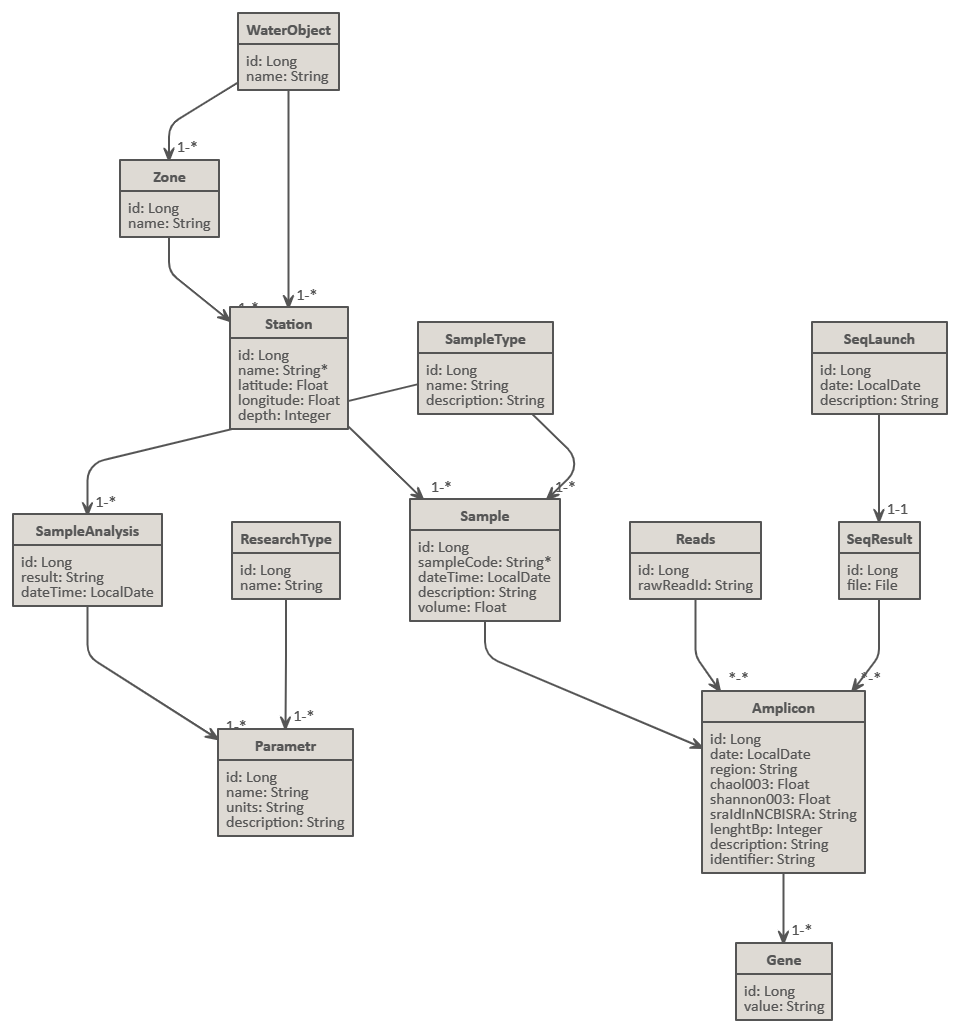
**1.2. Результаты с таблицами и картинками**

Реализованная объектная модель ИАС декомпозируется на следующие модули. Модуль «Research» обеспечивает общее описание проводимых исследовательских работ, включая данные по их исполнителям, публикациям, методикам, оборудованию и расходным материалам. Диаграмма сущностей данного модуля и их связи показаны на Рисунке 1.



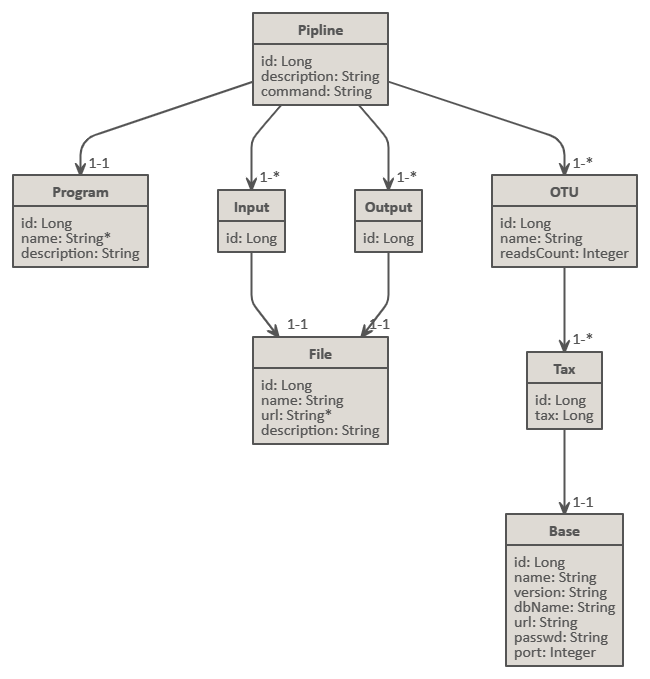
***Рисунок 1*. Диаграмма сущность-связь модуля описания исследования — «Research».**

Модуль «Sample» служит для представления специфических данных исследования микробиома озера Байкал, включая описания отбираемых проб, измеримые параметры, результаты анализа проб, изучаемые ампликоны, полученные прочтения, результаты секвенирования. Диаграмма сущностей данного модуля и их связи показаны на Рисунке 2.



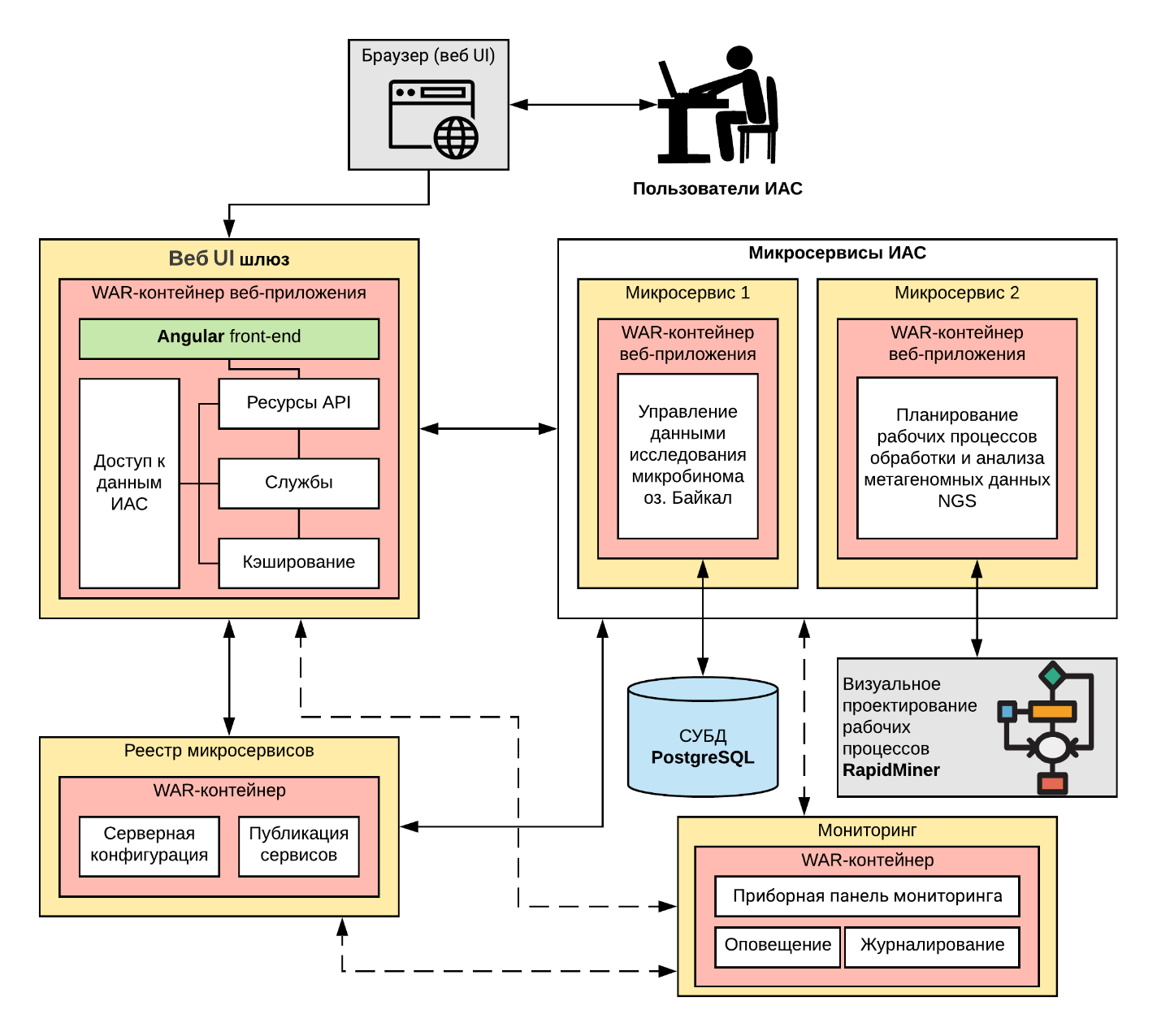
***Рисунок 2*. Диаграмма сущность-связь модуля описания пробы — «Sample».**

Модуль «NGS» организует данные, связанные с проведением вычислений по высокопроизводительному секвенированию ДНК/РНК последовательностей исследуемых проб воды и донных осадков, включая информацию по операционным таксономическим единицам, идентификаторам таксонов, используемым таксономическим базам данных, рабочим процессам (вычислительным цепочкам) обработки последовательностей и анализа результатов, списку программ, задействованных в этих рабочих процессах. Диаграмма сущностей данного модуля и их связи показаны на Рисунке 3.



***Рисунок 3*. Диаграмма сущность-связь модуля описания метагеномных данных — «NGS».**

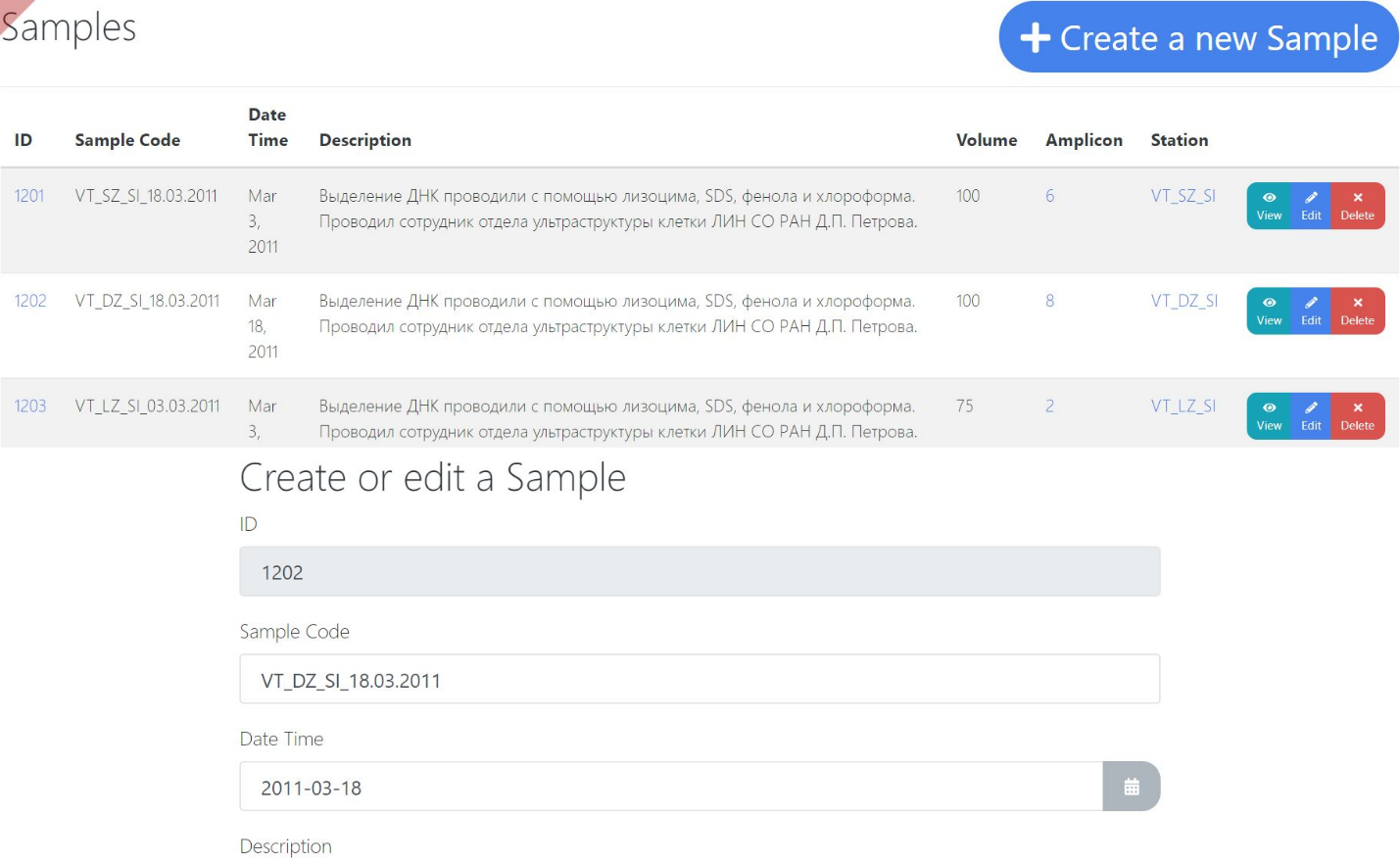
Реализуемая веб-ориентированная ИАС следует микросевисному архитектурному шаблону (см. Рисунок 4). Она включает следующие основные компоненты: «Веб UI шлюз» служит для построения веб-ориентированного пользовательского интерфейса, аутентификации пользователей, кэширования данных пользовательский сессий, а также загрузочной балансировки приложения; «Микросервис 1» обеспечивает управление данными исследования микробиома озера Байкал; «Микросервис 2» предоставляет интеграцию ИАС с системой визуального проектирования рабочих процессов обработки данных с целью планирования вычислительных конвейеров анализа метагеномных данных NGS. Кроме того, предлагаемая архитектура также включает служебные компоненты: «Реестр» для управления реестром зарегистрированных микросервисов и «Мониторинг» для отслеживания ошибок и замеров производительности микросервисов. Каждая компонента архитектуры ИАС реализуется в виде отдельного WAR-контейнера, исполняемого сервером веб-приложений.



***Рисунок 4*. Микросервисная архитектура ИАС поддержки исследования микробиома озера Байкал.**

Спроектированная архитектура обеспечивает улучшенную изоляцию, упрощенное тестирование и ускоренный запуск своих компонент, по сравнению с монолитным вариантом. Предполагается, что компонентное разделение реализуемой ИАС на отдельные микросервисы позволит в дальнейшем расширять функциональные возможности ИАС за счет интеграции со сторонним программным обеспечением сбора, публикации, анализа и визуализации исследовательских данных.

Веб-ориентированный пользовательский интерфейс ИАС следует архитектурному MVC-шаблону, реализованному платформой разработки «front-end» приложений — Angular. Пользовательский интерфейс обеспечивает навигацию (обращение, поиск, сортировку) и просмотр сохраненным данным объектной модели ИАС, а также предоставляет формы с стандартной CRUD (Create, Read, Update, Delete) функциональностью создания, редактирования, обновления и удаления записей для каждой сущности базы данных ИАС. Пример форм просмотра и добавления записей представлен на Рисунке 5.



***Рисунок 5*. Фрагменты снимков экрана веб-ориентированного пользовательского интерфейса ИАС с демонстрацией форм добавления результатов обработки новых проб в базу данных.**

Взаимодействие клиентской (навигационных и CRUD форм) и серверной части ИАС осуществляется посредством программного интерфейса (REST API) доступа к данным через отправку HTTP-запросов и получения в формате JSON данных в качестве ответов. Пример запроса и получения ответа представлен на ниже.

|  |
| --- |
| **HTTP Request:**  curl - H "Accept: application/json" - H "Authorization: Bearer eyJhbGciOiJIUzUxMiJ9.eyJzdWIiOiJhZG1pbiIsImF1dGgiOiJST0xFX0FETUlOLFJPTEVfVVNFUiIsImV4cCI6MTU3MjUxNDczNH0.MYLrZKjZ95chJYb\_0SECo9YGuFHAm5BTL5FcAlHYGDo6TcXnSNCfH0qD3\_EFZKWg5zJ9O2XP5\_dCBeS\_w4xjBA" http: //localhost:8080/api/stations  **HTTP Response:**    [{ "id": 1152, "name": "VT\_SZ\_SI", "latitude": 5154.009, "longitude": 10506.06, "depth": 250, "samples": null, "location": { "id": 1101, "name": "South basin, underside of the ice" } }, { "id": 1151, "name": "VT\_LZ\_SI", "latitude": 5154.08, "longitude": 10506.108, "depth": 40, "samples": null, "location": { "id": 1101, "name": "South basin, underside of the ice" } }, { "id": 1153, "name": "VT\_DZ\_SI", "latitude": 5154.371, "longitude": 10505.455, "depth": 600, "samples": null, "location": { "id": 1101, "name": "South basin, underside of the ice" } }, { "id": 1154, "name": "VT\_PZ\_SI", "latitude": 5154.618, "longitude": 10504.108, "depth": 1000, "samples": null, "location": { "id": 1101, "name": "South basin, underside of the ice" } } ] |

***Рисунок 5.*** Пример отправки HTTP-запроса и получение ответа в формате json

Следует отметить, что для доступа к системе через REST API, также как и через пользовательский интерфейс, необходима предварительная авторизация в системе. В представленном примере HTTP-запроса в заголовке Authorization содержится tokenID авторизованного пользователя, обеспечивающий доступ к ИАС.

Список основных модулей клиентской части ИАС представлен в Таблице 1. Модуль entities содержит «front-end» компоненты для каждой сущности, представленной в модели. Для того, чтобы клиентское приложение следовало принципам MVC- шаблона, каждая форма сущности разделена на 3 части, а именно модель, контроллер и представление. Список модулей, реализованных для каждой сущности модели данных ИАС представлен в Таблице 2.

**Таблица 1**. Основные модули клиентской части ИАС.

|  |  |
| --- | --- |
| **Модуль** | **Описание** |
| account | Управление аккаунтами пользователей |
| admin | Администрирование |
| blocks | Основные блоки, такие как конфигурация |
| entities | Модули описания сущностей |
| home | Домашняя страница |
| layouts | Общие элементы компоновки страниц (навигация, ошибки) |
| shared | Общие службы (авторизация, многоязычность) |
| app.main.ts | Main-класс приложения |
| app.module.ts | Конфигурация модулей |
| app.route.ts | Основная маршрутизация приложения |

**Таблица 2.** Список модулей, реализуемы для каждой сущности модели данных ИАС.

|  |  |
| --- | --- |
| **Модуль** | **Описание** |
| entity.component.html | HTML-представление для отображения списка объектов сущности |
| entity.component.ts | Контроллер для отображения списка объектов сущности |
| entity.model.ts | Модель, представляющая сущность |
| entity.module.ts | Angular-модуль для сущности |
| entity.route.ts | Конфигурация Angular-маршрутизации сущности |
| entity.service.ts | Служба доступа к REST-ресурсу сущности |
| entity-delete-dialog.component.html | HTML-представление формы для удаления объекта сущности |
| entity-delete-dialog.component.ts | Контроллер для удаления объекта сущности |
| entity-detail.component.html | HTML-представление формы для детального просмотра объекта сущности |
| entity-detail.component.ts | Контроллер для детального просмотра объекта сущности |
| entity-dialog.component.html | HTML-представление формы для изменения объекта сущности |
| entity-dialog.component.ts | Контроллер для изменения объекта сущности |
| entity-popup.service.ts | Служба обработки диалоговой формы для добавления/изменения объекта |
| index.ts | Контейнер для отображения всех компонентов |

**1.3. Краткое обсуждение**

Спроектированная архитектура веб-ориентированной версии ИАС поддержки исследования микробиома озера Байкал позволила структурно организовать модули управления данными и планирования рабочих процессов обработки и анализа метагеномных данных NGS в виде отдельных микросервисов. Независимость развёртывания и обновления каждого из микросервисов обеспечивает возможность изменения и расширения ИАС. Дальнейшая работа предполагает развитие ИАС за счет внедрения современной среды оркестрации микросервисов, обеспечивающей отказоустойчивость и балансировку нагрузки, а также включение непосредственно в контур среды исполнения системы непрерывной интеграции, обеспечивающее автоматизацию обновления и развёртывания микросервисов.

Ожидается, что микросервисы, реализующие усовершенствованную объектную модель метагеномных данных NGS, обеспечат возможность управления метагеномными данными больших объемов на протяжении всего их жизненного цикла. В частности, это позволяет микробиологам хранить и публиковать в открытом и защищенном доступе свои первичные и обработанные данные, результаты исследования, а также самостоятельно создавать и хранить собственные рабочие процессы обработки и анализа метагеномных данных при исследовании микробиома озера Байкал.

**Список опубликованных работ с указанием базы, в которой индексируется издание**

1. Cherkashin E., Shigarov A., Malkov F., and Morozov A. An Instrumental Environment for Metagenomic Analysis. Proc. Int. Conf. on Information Technologies in the Research of Biodiversity. 2018. **(Scopus)**

2. Cherkashin E., Shigarov A., Paramonov V. Representation of MDA Transformation with Logical Objects // Proc. 7th Int. Conf. on Knowledge, Ontology, and Theory. (**Принято в печать**).

3. Малков Ф.С., Шигаров А.О., Лихошвай Е.В., Черкашин Е.А., Христюк В.В., Морозов А.А., Михайлов И.С., Галачьянц Ю.П. Свидетельство о государственной регистрации программы для ЭВМ № 2019613947 от 26 марта 2019 г. «Программное средство поддержки исследования микробиома оз. Байкал».

**Список процитированной литературы**

нет