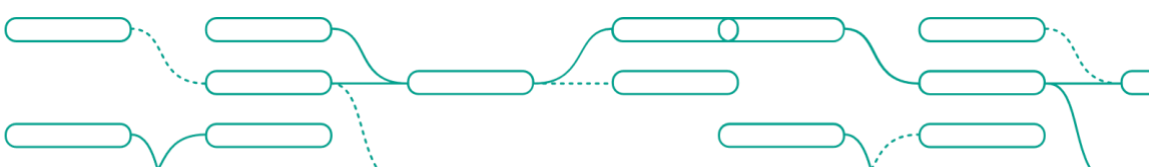
Arenadata Catalog

Arenadata Catalog

Руководство пользователя v 0.4.0

Москва 2023



Arenadata Catalog

Оглавление

Arenadata Catalog

Журнал изменений

Дата	Версия	Комментарий
25.10.2022	0.1	Начальная версия документа
05.03.2022	0.2	Обновление документа с релизом 0.2.2
15.06.2022	0.3	Обновление документа с релизом 0.3.0
18.07.2022	0.4	Обновление документа с релизом 0.4.0

Arenadata Catalog

Введение

Термины и определения

Термин	Значение
База данных	Совокупность данных, хранимых в соответствии со схемой, манипулирование которыми выполняют в соответствии с правилами средств моделирования данных
Бизнес-глоссарий	Словарь для бизнес-пользователей. Словарь состоит из бизнес-терминов, которые могут быть связаны друг с другом, и позволяет распределять их по Предметным областям, чтобы их можно было понимать в разных контекстах
SQL	Декларативный язык программирования, применяемый для создания, модификации и управления данными в реляционной базе данных, управляемой соответствующей СУБД

Сокращения и обозначения

Сокращение	Наименование
ПО	Программное обеспечение
СУБД	Система управления базами данных
ADC	Arenadata Catalog
Arenadata EDP	Arenadata Enterprise Data Platform

Общие положения

Настоящий документ является руководством пользователя программного обеспечения Arenadata Catalog (ADC). ADC– это масштабируемая отказоустойчивая система, предназначенная для удобного и простого формирования базы знаний о ландшафте данных и управления ей. Система реализована на базе технологии opensource решения Open-Metadata. Значимыми преимуществами Arenadata Catalog являются расширенный функционал Бизнес-глоссария, а также интеграция с другими компонентами платформы Arenadata EDP.

Основное назначение ПО ADC заключается в предоставлении удобного инструмента корпоративного уровня, в котором можно найти информацию по цифровым ресурсам, увидеть их структуру, взаимосвязи, проследить lineage (информацию о потоках, происхождении данных), посмотреть profiling (краткую статистику содержимого), sample (пример наполнения), историю проверок качества данных, увидеть владельцев ресурсов и связанные с ними бизнес понятия.

ADC отличается гибко настраиваемой структурой Бизнес-глоссария, широкими возможностями конфигурации рабочих процессов для его управления и интуитивно понятным для конечного пользователя интерфейсом.

С точки зрения архитектуры ADC является интеграционным многокомпонентным кластерным решением на базе ПО собственной разработки, объединяющим:

- приложение бизнес-пользователя с визуальным интерфейсом;
- SQL-хранилище для задач хранения служебных и метаданных;
- поисковую систему, предназначенную для анализа и полнотекстового поиска в режиме реального времени;
- платформу для моделирования и создания бизнес-процессов управления объектами каталога данных;
- приложение для создания, мониторинга и оркестрации задач по импорту метаданных, включающее в себя коннекторы к различным типам цифровых ресурсов.

Более подробная документация с видео-примерами функционала продукта доступна по ссылке. **Рекомендуем обращаться к web-документации** при возникновении вопросов.

Начало работы

Запуск ПО

Запуск программного обеспечения осуществляется в web-браузере по адресу:

```
http://[server_addres]:8585
```

Открывается домашняя страница Arenadata Catalog с новостной лентой деятельности. Пример ее интерфейса показан на рисунке 1.

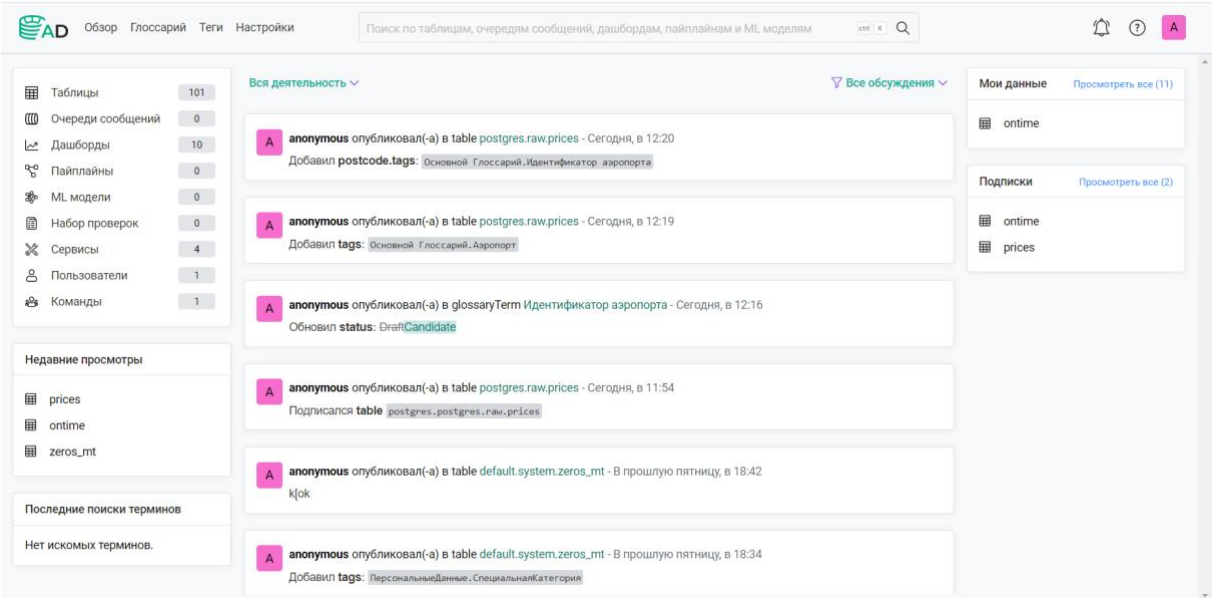


Рисунок 1— Стартовая страница Arenadata Catalog

В случае, если ПО не было установлено, воспользуйтесь инструкцией по установке.

Добавление сервисов

Данный шаг можно пропустить, если сервисы были добавлены ранее.

Необходимо перейти в раздел настройки в шапке программы, далее на левой панели выбрать необходимый для добавления сервис.

Например, необходимо добавление метаданных из новой СУБД. Для добавления мы выбираем «Базы данных», система открывает окно для просмотра подключенных БД (Рисунок 2).

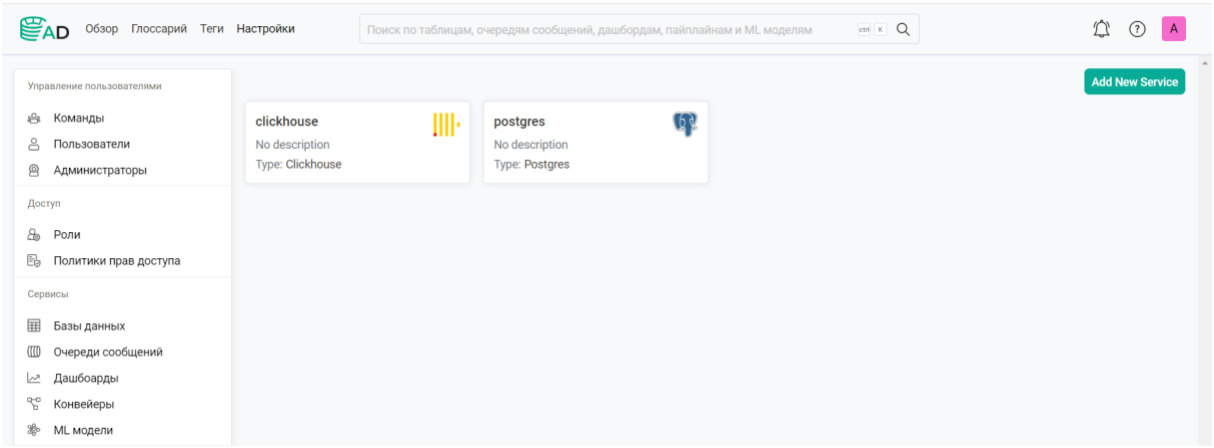


Рисунок 2 — Окно раздела «Базы данных»

Для нового подключения нажимаем на кнопку «Добавить новый сервис», открывается страница добавления нового сервиса (Рисунок 3).

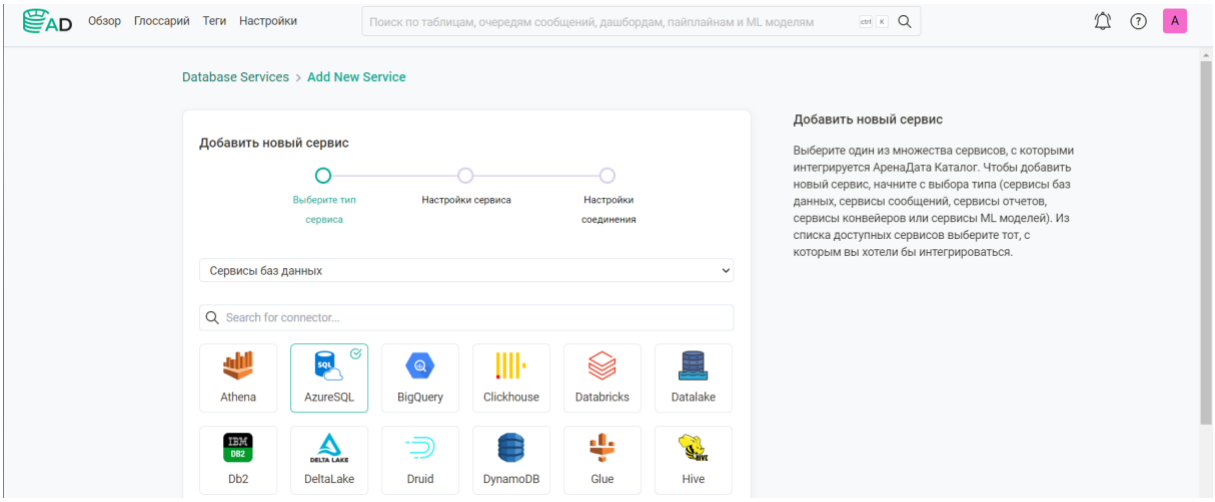


Рисунок 3 — Страница добавления нового сервиса

Далее из предложенного перечня интегрируемых с ADC сервисов выбираем необходимый.

Пошагово выполняется подключение необходимого сервиса:

1. Нажать кнопку «Далее».
2. Указать наименование подключаемого сервиса и заполнить описание.
3. Заполнить данные для доступа к БД, выбрать способ подключения.
4. Нажать кнопку «Тест подключения».

5. Сохранить добавленную настройку.

Каталог данных

Каталог данных- единое хранилище метаданных всех информационных активов организации. Arenadata Catalog позволяет не только обозревать метаданные, а также проводить проверки данных, отслеживать происхождение, запускать профилирование, устанавливать связи между объектами каталога и глоссарием и многое другое.

Чтобы открыть обзор каталога данных, достаточно кликнуть Обзор в хедере приложения. В обзоре отображаются 5 вкладок, согласно 6 типам, поддерживаемых сервисов: Таблицы, Очереди сообщений, Дашборды, Конвейеры, ML модели, Контейнеры. При переходе на одну из вкладок будет открыт список соответствующих источников.
В левом боковом меню предложены быстрые фильтры. Здесь вы можете отфильтровать перечень объектов по структуре сервисов или тегам. В правой части экрана отображается краткий обзор выбранного источника, чтобы перейти в подробный обзор, кликните по наименованию объекта данных.

ADC предоставляет широкие возможности поиска и фильтрации.

В центральной части хедера приложения расположен основной поиск по каталогу данных. В нем вы можете искать по наименованию и описанию сервиса, наименованию элементов схемы и их описанию. Для уточнения поиска можно использовать специальные символы.

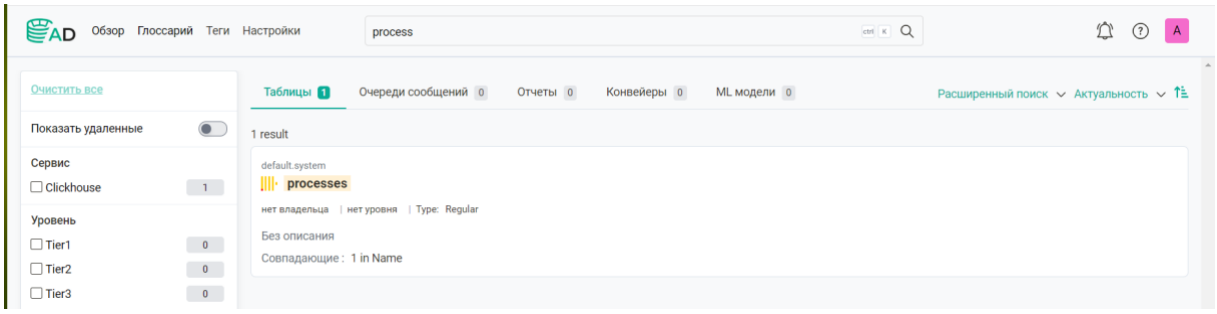


Рисунок 4 — Результаты поиска по каталогу данных

Управление объектами данных

В карточке объекта данных отображается подробная информация об объекте. Для перехода в карточку необходимо кликнуть по **Наименованию объекта**, находясь в разделе **Обзор**.

Для каждого источника данных в Arenadata Catalog можно указать **Владельца.** Владелец-пользователь ответственный за управление объектом. На основе владельцев можно настраивать ролевую модель: запрещать или разрешать редактирование и просмотр сущностей.

Для изменения владельца объекта необходимо иметь определенные полномочия в системе. Чтобы назначить владельца необходимо **перейти в карточку объекта** и кликнуть под ее наименованием на **иконку редактирования.** После клика разворачивается поиск по пользователям и командам, где необходимо указать нужного владельца. Удалить или изменить владельца также можно в этом окне.

Профилирование данных

Профилирование данных считает количество строк в таблице, оценивает уникальность значений в каждом столбце и выгружать пример данных в систему. Для профилирования данных необходимо настроить рабочий процесс загрузки данных типа Profiler.

Чтобы просмотреть **Пример данных** из таблицы, необходимо в карточке объекта открыть одноименную вкладку. Возможность просмотра примера данных управляется ролевой моделью, во избежание показа конфиденциальной информации.

Просмотреть профилирование таблицы и ее столбцов можно на вкладке **Профилирование & Качество данных.**
Для просмотра детальной информации по столбцу- кликните по его наименованию.

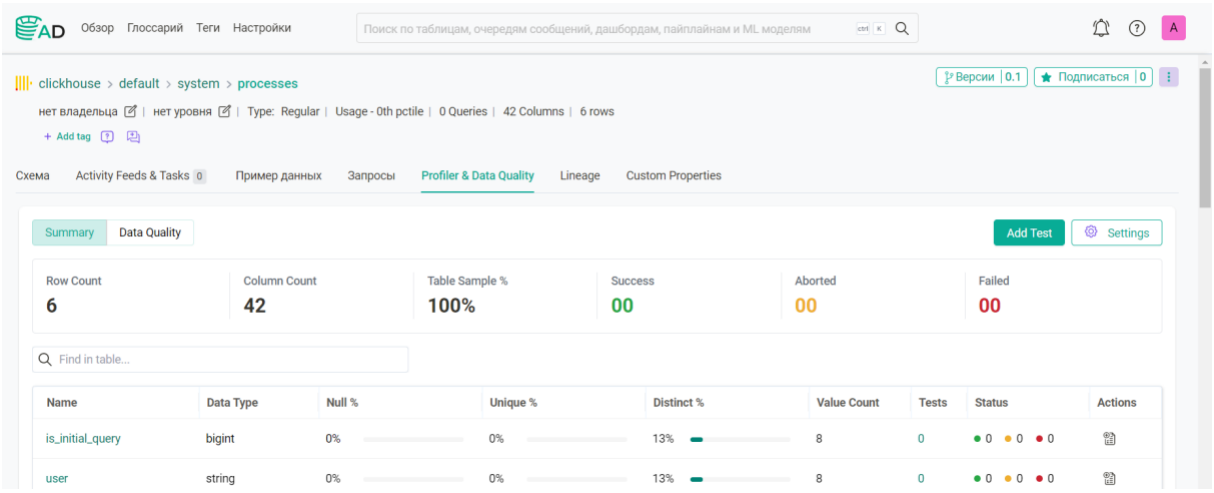


Рисунок 5 — Профилирование данных

Качество данных

Arenadata Catalog знает, что только достоверные и полные данные могут обеспечить точный анализ, который, в свою очередь, помогает принимать надежные бизнес-решения.
Для проверки качества данных используются пользовательские тесты, которые можно запускать по расписанию.

Если вы используете сторонние приложения для проверки качества данных, вы можете записывать результаты внешних проверок в Arenadata Catalog с помощью API

Результаты выполнения тестов можно проследить как в карточке таблицы, так и в наборе проверок. В карточке объекта на вкладке **Профилирование & Качество данных** вы можете выбрать раздел **Качество данных** в боковом меню, чтобы просмотреть статистику проверок данного ресурса

Создать новую проверку качества данных можно из карточки проверяемого объекта.
Для этого на вкладке **Профилирование & Качество** данных кликните на кнопку **Добавить тест.**

Первым этапом создания новой проверки является указание **набора проверок.** Группируйте проверки в наборы по оправданному признаку, это поможет вам настроить подходящие расписание и упростить анализ результатов.

Вторым этапом создания новой проверки является указание наименования проверки и ее типа. В ADC есть предопределенные типы проверок, которые помогают отследить состояние строк и столбцев, а также возможность создание проверок любой сложности с помощью SQL-запросов в едином интерфейса Arenadata Catalog.

Происхождение данных

Data lineage- информация, помогающая отследить путь формирования данных, точки использования, обработки и применения. В Arenadata Catalog есть 2 способа формирования происхождения данных: автоматическое и ручное.
Например, при добавлении загрузки метаданных из сервиса BI, вы можете указать какие сервисы баз данных использует этот инструмент (например, PostgreSQL) и Arenadata Catalog создаст связи происхождения между диаграммами/дашбордами сервиса с таблицами, на основе которых строятся инструменты.

Если признаки для автоматического связывания отсутствуют, пользователь может выстроить связи самостоятельно используя графический редактор происхождения данных.

Рисунок 6 — Визуализация происхождения данных

Лента активности и задач

Для коммуникаций внутри системы и коллективной работы существует раздел с лентой активности и задачами (Рисунок 7). У пользователей имеется возможность оставлять комментарии в соответствующем разделе.

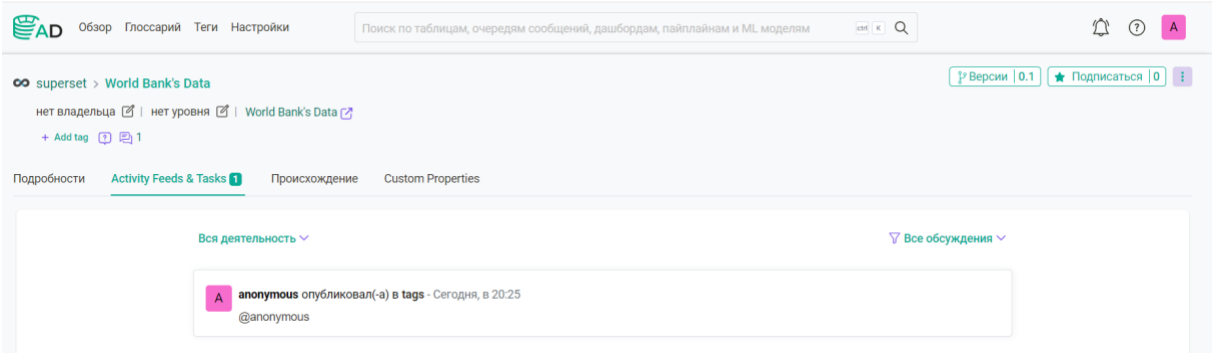


Рисунок 7— Лента активности и задачи на странице объекта данных

Глоссарий

В бизнес-глоссарии ADC существует 2 типа объектов, существующие для поддержания иерархии: Глоссарии и предметные области.
Вы можете создавать несколько отдельных глоссариев для разных подразделений организации или для любых других нужд. По умолчанию в ADC создан один глоссарий с названием «Основной Глоссарий».

Глоссарий содержит второй обязательный уровень иерархии- предметную область. В глоссарии может быть неограниченное количество предметных областей, а сами предметные области поддерживают многоуровневые связи, вы можете создать родительские и дочерние предметные области.

Термины в Arenadata Catalog типизируются

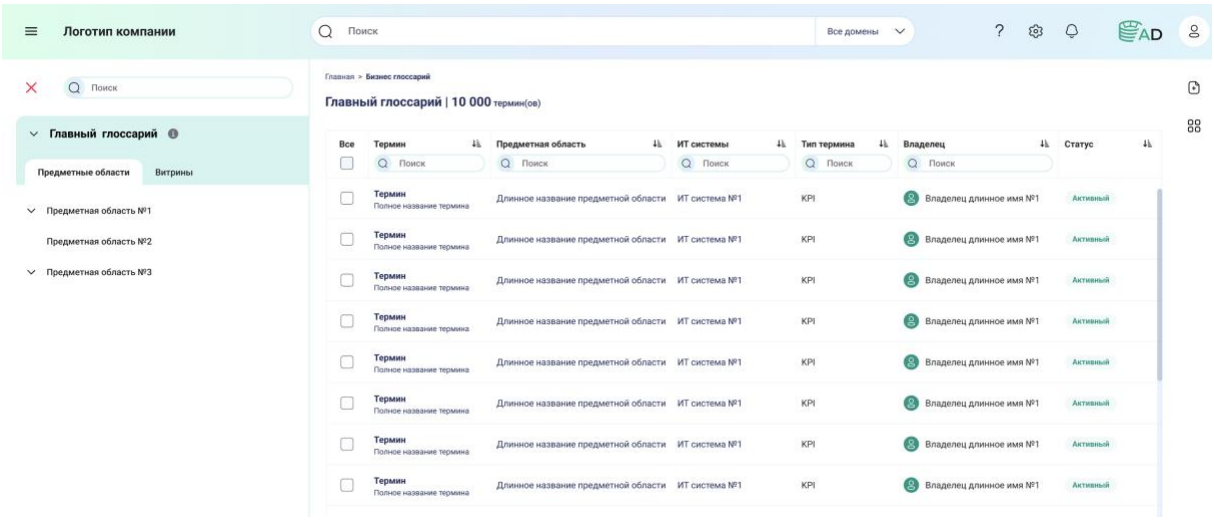


Рисунок 8 — Пользовательский интерфейс Глоссария

Получение информации о термине

В целях получения информации о необходимом термине достаточно перейти на страницу глоссария и в строке поиска ввести нужное название. Кликнув на название, осуществляется переход к карточке термина для изучения подробных данных (Рисунок 9).

В карточке термина вам доступны обсуждения, просмотр истории версий термина вместе с историей согласования и возможность подписки.

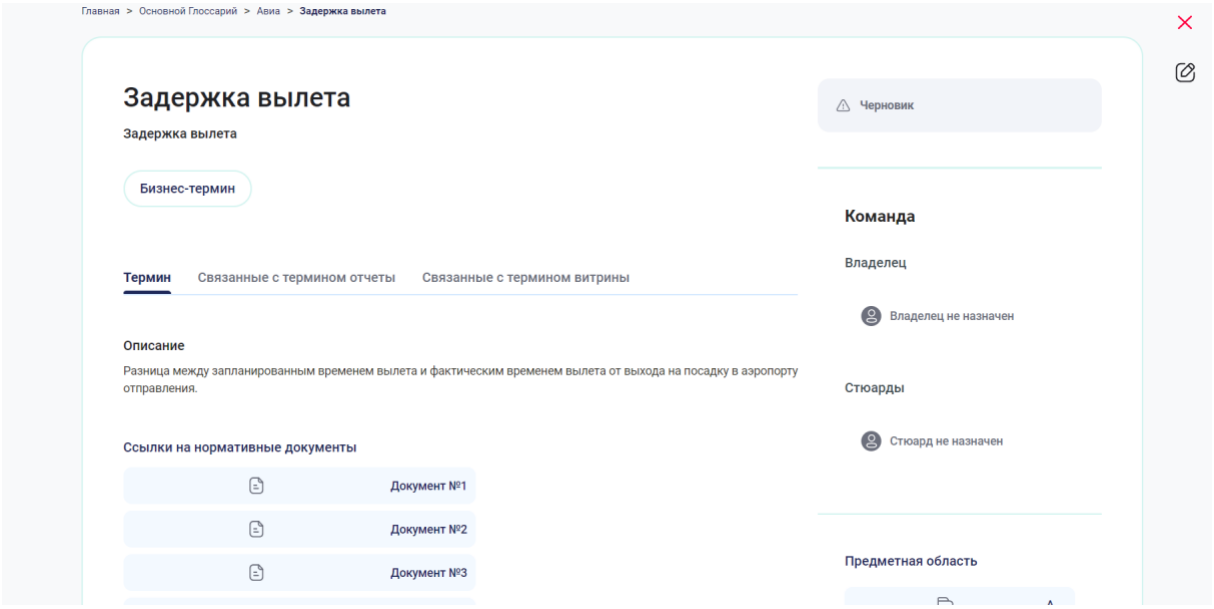


Рисунок 9 — Карточка термина в Глоссарии

Создание нового термина

Для создания нового термина на странице глоссария ADC используется кнопка  .

После ее нажатия открывается страница для создания нового термина (Рисунок 10).

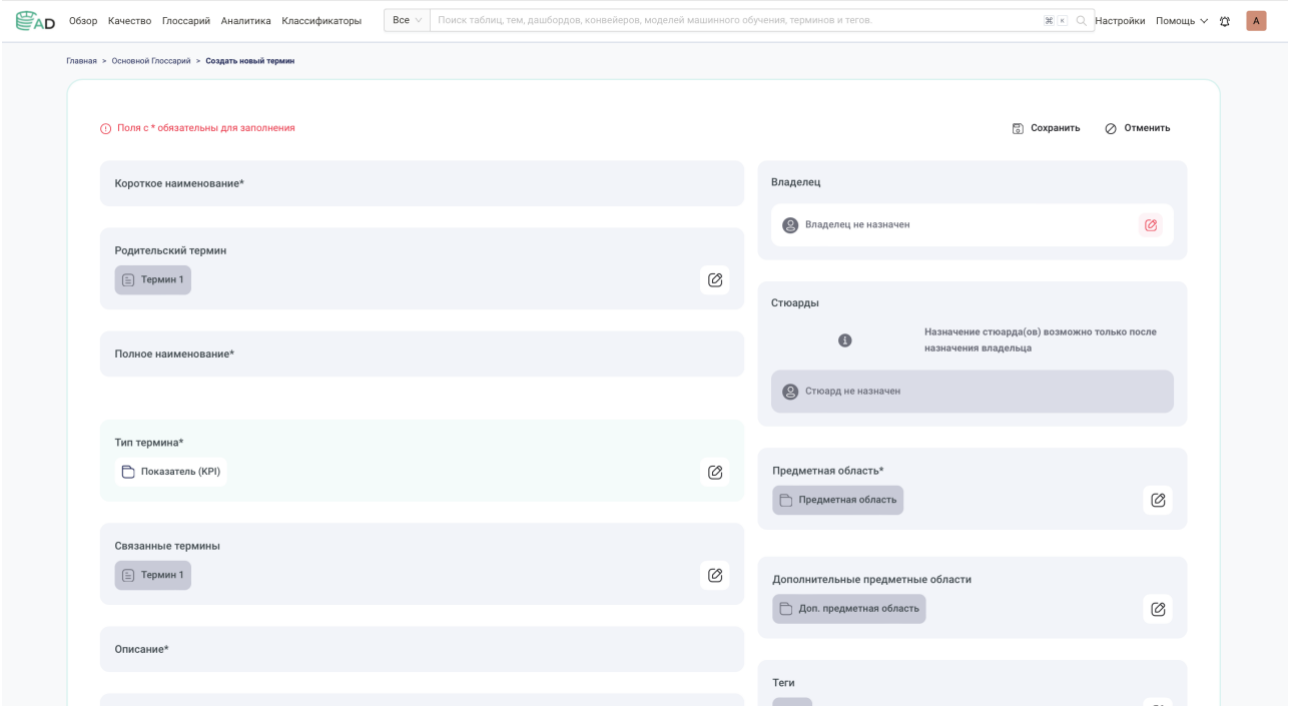


Рисунок 10 — Страница для создания нового термина

Заполняются обязательные поля, и необходимые пользователю дополнительные. Нажав кнопку сохранить, новый термин появляется в глоссарии и доступен к просмотру другим пользователям с соответствующими полномочиями. Созданный тестовый термин приведен на рисунке 11.

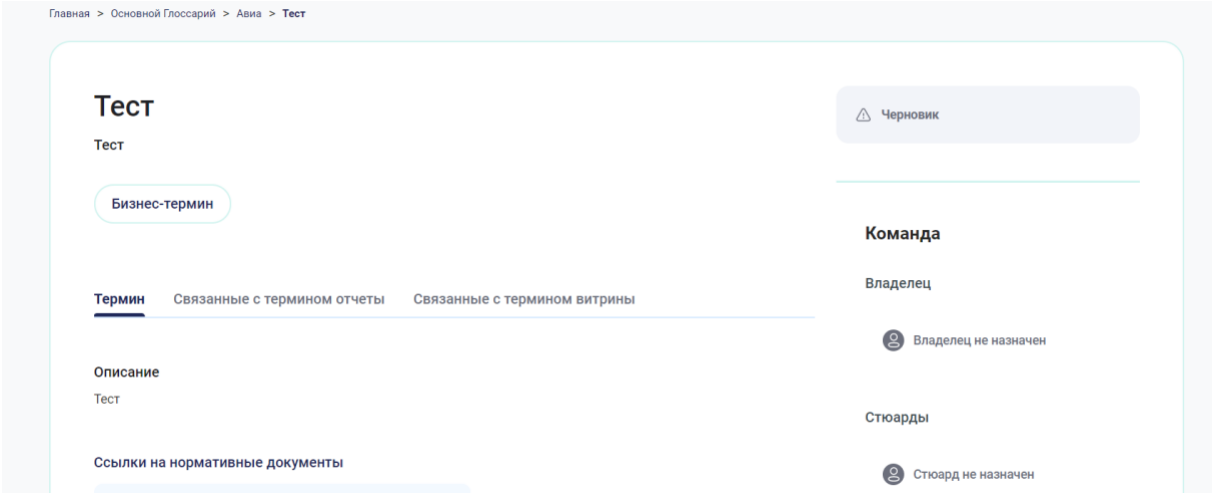


Рисунок 11 — Просмотр карточки термина

Контакты технической поддержки

Для связи с технической поддержкой по вопросам, связанным с руководством пользователя, воспользуйтесь электронным адресом: info@arenadc.io