HEDDA.IO

HEDDA.IO End User Documentation

Version 2.2.0

oh22information services GmbH

2025-07-30

smart
casual
datadesign

Contents

1 Introduction

Welcome to the in-text documentation and presentation of HEDDA.IO.

This document will explain what is HEDDA.IO, how does it work, and most importantly, how can you navigate HEDDA.IO in order to effectively use all the functionalities it provides.

First, we start with the most important question “What is HEDDA.IO?”.

HEDDA.IO is a central data quality management solution that connects departments, data stewards and data engineers. It helps to easily integrate standardization, cleansing, matching and enrichment tasks into existing processes.

With the help of core functions such as Domain Members, Actions, Rule Books, HEDDA.IO can validate, clean, and standardize you data. And with the implemented Pandas Data Profiling open-source framework, you can run extensive and complex analysis on your data.

2 Concepts

Let’s start by introducing you to the core concepts of HEDDA.IO. We’ll first list them and then take a closer look at each of them further into this section.

- Projects
- Git
- External Connections
- Project Service Principal
- Events
- User Management
- Knowledge Base
- Domains
- Members and Member Search
- Algorithms
- Rulebooks and Business Rules
- Business Rule Preparation
- Business Rule Actions
- Dataset Rules
- Data Links
- Runs
- Preview
- Mappings
- Knowledge Base Versioning
- Tags

2.1 Projects

Projects are essentially workspaces within HEDDA.IO, designed to cater to your distinct needs for utilizing the HEDDA.IO platform.

Unless you have a Project established, it's not possible to generate a Knowledge Base. As a result, initiating a Project stands as the primary step towards making use of HEDDA.IO. Here's a guide on how to Create a Project in HEDDA.IO.

The Project page encompasses several vital Project-wide features, including External Connections, Events, and User Management. Moreover, it presents the list of Knowledge Bases and provides the capability to add new ones.

A comprehensive guide on how to navigate and create a Project in HEDDA.IO is available here.

2.2 Git

Projects can be configured to store data in a Git Repository rather than a Microsoft SQL Server (MsSQL) database.

Git is a distributed version control system that tracks versions of files. With this integration, data is synchronized between the HEDDA.IO Server and a remote Git Server, preventing data loss.

This integration enables multiple independent Edit versions of your Knowledge Bases, giving you greater flexibility during development. Additionally, using Git significantly improves the read performance of Knowledge Base data.

It is best practice to store each Project in its own repository.

2.2.1 Merging

When working with several editing versions, it is possible that contradicting changes are made to the same parts of a Knowledge Base. Situations of conflicting changes are detected by HEDDA.IO and can be resolved within the application by deciding which change should take precedent above the other.

2.2.2 Synchronizing

When working with Git Repositories, it is possible that the remotely stored data were updated and that these information aren't in HEDDA.IO yet. HEDDA.IO will detect such data and offer to synchronize the information so the changes are available in HEDDA.IO.

2.2.3 Git Example: Dev-Test-Prod Project Flow with Knowledge Bases

A commonly adopted practice to ensure quality and reliability is to separate Environments for Development, Testing and Production within HEDDA.IO which is supported by using Git Projects.

This approach enables developers to safely implement and modify Knowledge Bases in a dedicated development system, without affecting the live data. Once development is complete, the Knowledge base is moved to the Quality Assurance (QA) Environment for testing. Any bugs and issues identified during QA can be resolved within the development system and revalidated in QA as needed.

With QA completed, the Knowledge Base is imported into the Production Environment, where it is available for immediate use.

The following example shows how to set this up in HEDDA.IO and use it.

2.2.4 Creating and linking Environments

Linked Environment Branches are meant to link Projects across Environments. Still, they can be used to import Knowledge Bases from other Projects within the same Environment. This will be used in this example to demonstrate linking within a single Environment.

To start, three Projects need to be created. Each of the following Projects represents an Environment. Please create the Projects `Development`, `QA` and `Production`.

- The Project `Development` will be used to develop and bugfix Knowledge Bases.
- The Project `QA` will be used to test changes.
- The Project `Production` will be used as productive Environment

After the Projects are created, their Git connection needs to be configured using the same Repository URL as seen in the following table:

Environment	Environment Branch Name	Linked Environment Branch
`Development`	`Dev`	
`QA`	`QA`	`Dev`
`Production`	`Prod`	`QA`

Each Project will now store Knowledge Bases on its respective Git Branch in the repository, defined with Environment Branch Name. The Linked Environment Branch later allows importing Knowledge Bases from the linked Environment into the current Environment.

2.2.5 Importing Knowledge Bases from linked Environments

Now that the Projects are set up and configured, an empty Knowledge Base can be created and published in `Development`.

After publishing the Knowledge Base in `Development`, please open the Project `QA`. In Details under Git, it will show that there are differences in the linked Environment. Clicking on `Has Differences` will open a dialogue which shows the differences between the current Environment and the linked Environment.

By clicking on `Synchronize`, the merge process is started. `Replace` overwrites the current Environment with the incoming one.

The changes made in Development are now available in the `QA` Environment. After testing them, the changes can be imported into the `Production` Environment just like they were imported into the `QA` Environment.

It is possible to import a Knowledge Base from a linked Environment which uses External Connections in Data Links. Missing Data Link connections can be reconfigured during Synchronization.

2.3 External Connections

External Connections provide you with the capability to establish links to external database services of different types, including Azure Storage, Microsoft SQL Server, and Reference Data Service. These connections can be currently utilized to create Data Links or to import Members into a Domain from an external dataset.

A guide on where to find and how to navigate the External Connections page, as well as how to create a Connection within the HEDDA.IO user interface, is available in the External Connection Configuration Section.

2.4 Project Service Principal

The Project Service Principal gives you the possibility to define an Azure Service Principle which can be used within some External Connections for authentication.

2.5 Events

Events enable you to trigger external services when specific occurrences take place. This encompasses actions like initiating an HTTP Call, dispatching an Email, generating a work item within DevOps (such as Bugs, Tasks, or Issues), or even delivering a Teams notification.

The prescribed sequence for this process involves initially setting up an Event Sink, followed by establishing an Events layered on the foundation of the sink.

Use Case

The Events system offers a diverse range of applications, and we can explore several examples.

1. Generate a work item in Azure DevOps whenever there is a failure in a Dataset, Rulebook, or Business Rule.
2. Transmit a payload to Events Grid or Hub upon the publication of a new version of a Knowledge Base.
3. Dispatch an Email, Teams, or Slack message upon the completion of an execution.

A guide on where to find and how to navigate the Events page, as well as how to create Event Sinks and Events within the HEDDA.IO user interface, is available here.

2.6 User Management

Preventing unauthorized users from accessing Projects they shouldn't be a part of is always a priority.

That's why HEDDA.IO provides you with the ability to control the levels of permission users possess for various Projects within the platform, through the utilization of the Manage Users feature. Currently, there are three available permission levels at your disposal: Read, Write, and Manage.

A guide on where to find the User Management and how to correctly manage users within the HEDDA.IO user interface, is available here.

2.7 Knowledge Bases

Knowledge Bases are entities responsible for containing comprehensive information about the structure of your data, the regulations it needs to adhere to, Execution metrics, and a variety of additional details that we will cover further in the Concepts section.

Every Knowledge Base can be configured to facilitate a workflow that could systematically processes a particular dataset associated with the Project. The Knowledge Base encompasses Domains and their corresponding Members, Rulebooks along with their associated Business Rules, Dataset Rules, Data Links, Runs accompanied by their corresponding Execution Statistics, and Mappings. Each of these components will be explored in greater detail further into the Concepts section.

You have the option to manually create a Knowledge Base, which also involves establishing the associated Domains in a similar fashion. Alternatively, a Knowledge Base can be imported from a dataset. The dataset can take the form of a Parquet file or an SQL table, and can be imported from a local environment, Azure Storage, or a Microsoft SQL Database.

A guide on where to find, how to navigate, and how to add Knowledge Bases within the HEDDA.IO user interface, is available here

The Knowledge Base can also be exported and imported, enabling seamless transfer between various Projects and even different environments altogether. Click here to find out how this can be achieved within the HEDDA.IO user interface.

2.8 Domains

In HEDDA.IO, Domains refer to the distinct columns present in your dataset, and within a Knowledge Base. They can comprise a variety of categories, each encapsulating a comprehensive set of relevant values.

In addition to manually adding Domains to a Knowledge Base, you can also import them from a dataset, which is carried out when importing the entire Knowledge Base from a dataset source. During this process, you are presented with the choice to specify which columns from the dataset should be brought in as Domains within HEDDA.IO. In such cases, the configuration attributes of these Domains, such as their names and data types, will be inherited from the data types associated with the imported columns.

Furthermore, you always retain the flexibility to modify the Domains afterwards, allowing you to fine-tune other properties. These properties encompass a description, validation Algorithms (applicable to `string` data types), precision and scale adjustments (pertinent to `number` data types), and corresponding access modifiers, among other attributes.

Domains can be marked as Variables. Variable Domains can still be used in Business Rules and Actions but will not appear in Mappings as they are not read from the Source Data. This means that they do not need to be present in eg. the source data frame. On top of this Variable Domains can also have a Default Value configured.

On Preview Screen, Variable Domains are hidden by default but can be made visible through the Domain Selection.

A guide on where to find, how to navigate, and how to add Domains within the HEDDA.IO user interface, is available here.

2.9 Members and Member Search

In HEDDA.IO, Members represent potential values expected within a Domain in your dataset. These Members are categorized into two types: Main and Synonyms. Main Members encompass the specific values desired in your processed dataset, while Synonym Members, associated with Main Members, represent likely alternative values that HEDDA.IO might encounter in the corresponding Domain.

Each Domain can contain a number of Members, and each Member has a Validation Status property that can be set to Valid or Invalid, allowing the validator to make the distinction between the valid and invalid values.

Each Member can hold multiple Synonyms. These Synonyms might include misspelled or abbreviated names, or even IDs that require substitution with their corresponding full names or vice versa. Essentially, by having Members and their associated Synonyms, you gain the capability to link a value within a column to a Synonym and subsequently substitute that value with the Main Member associated with that Synonym. This procedure is a concept is what we call a **Member Search**.

Enabling the `Is External` switch within a Domain configuration permits you to import both the Main Members and their respective Synonyms from an external dataset.

Use Case

Let's consider a scenario where you have a column in your dataset containing payment type names, but for easier data handling, you prefer to represent them as IDs. In this setup, you designate the Main Member as the IDs and associate each Main Member with its corresponding Synonym, which is the name of the payment type.

Let's take a look at the following table.

ID	PaymentName
1	Cash
2	Credit Card
3	PayPal

In the context of our use case and based on the provided table, you would designate each `ID` as the primary Member and each `PaymentName` as the Synonym corresponding to its respective ID.

As a result, the values in this column will be transformed from `PaymentName` to their corresponding `ID` if a PaymentName matching the Synonym is encountered in your dataset's specified column.

A guide on where to find, and how to add Members to Domains within the HEDDA.IO user interface, is available here.

2.10 Algorithms

Depending on the data utilized within your Domain, the outcomes can be significantly enhanced by selecting the appropriate algorithm. Phonetic algorithms, for instance, can be effectively employed for product names.

When dealing with product numbers in your Domain, employing a distance algorithm like Levenshtein Distance can yield positive results.

However, when dealing with telephone numbers, even minor discrepancies can lead to entirely incorrect outcomes. Hence, in such cases, the optimal algorithm is an exact match.

2.10.1 How does HEDDA.IO uses Algorithms

When you validate data against a Domain, HEDDA.IO examines whether the value is already defined as a Member for this Domain. If a match is found, HEDDA.IO will continue using the found Member instead of the original value in all following processes. Additionally, if a match is found and this match is a Synonym of a different Member, the Main Member will be used instead.

If no direct match was found, HEDDA.IO will try again to find an existing Member, based on the selected algorithm. These Algorithms include distance algorithms, Phonetic Algorithms as well as keyboard-distance Algorithms.

An overview of all the Algorithms available in HEDDA.IO can be found here.

2.11 Rulebooks and Business Rules

To further enhance the validation process for the data that is ingested, HEDDA.IO offers a feature known as Rulebooks. This serves as an encapsulation for a collection of Business Rules, enabling you to organize and arrange them in a coherent order, based on basic logical rules.

Furthermore, it provides an overview and a visual representation of the sequence of Business Rules on a canvas, allowing you to easily visualize the flow and make modifications.

Conditions and subconditions can be incorporated into Business Rules, allowing the utilization of Data Links values or other Domain values to validate Domains.

Business Rules can be designated as Dataflow Rules, enabling HEDDA.IO to treat them as logical operations only. This configuration excludes them from the validation results and associated statistics, emphasizing their role in data manipulation or transformation without affecting validation outcomes.

It's essential to note that Business Rules can only be formulated within a Rulebook and must be part of the Business Rule sequence inside that Rulebook. A Rulebook can have just one Business Rule sequence, which is capable of forking any number of times.

Use Case

Let's consider a scenario where you have two distinct sets of Business Rules designed to validate and/or correct data for different purposes or departments within your company. An effective approach is to create one Rulebook for the first purpose, containing the corresponding set of Business

Rules. Simultaneously, you can create a second Rulebook for the business rules aligned with the other purpose. This methodology facilitates ease of modification to the Business Rules, streamlining their location. Additionally, it allows HEDDA.IO to organize statistics in a way that enables you to effortlessly extract statistics specifically related to each Rulebook.

Another use case is related to the Business Rules that are set as logical operators by enabling the `Is Dataflow Rule` switch.

By setting up a Rulebook with a Dataflow Rule based on the `Employment Duration` column, you can dynamically execute different sets of Business Rules depending on the value of that Domain. Enabling the `Is Dataflow Rule` switch ensures that the Dataflow Rule operates as a logical condition without impacting validation results.

For instance, in the case of employee data, having one set of rules for those with less than 10 years of employment and another set for those with 10 years and above is a practical way to tailor the data processing. The visual representation of the Rulebook, as you described, with the Dataflow Rule at the top, demonstrates a clear and effective structure for such conditional execution of Business Rules.

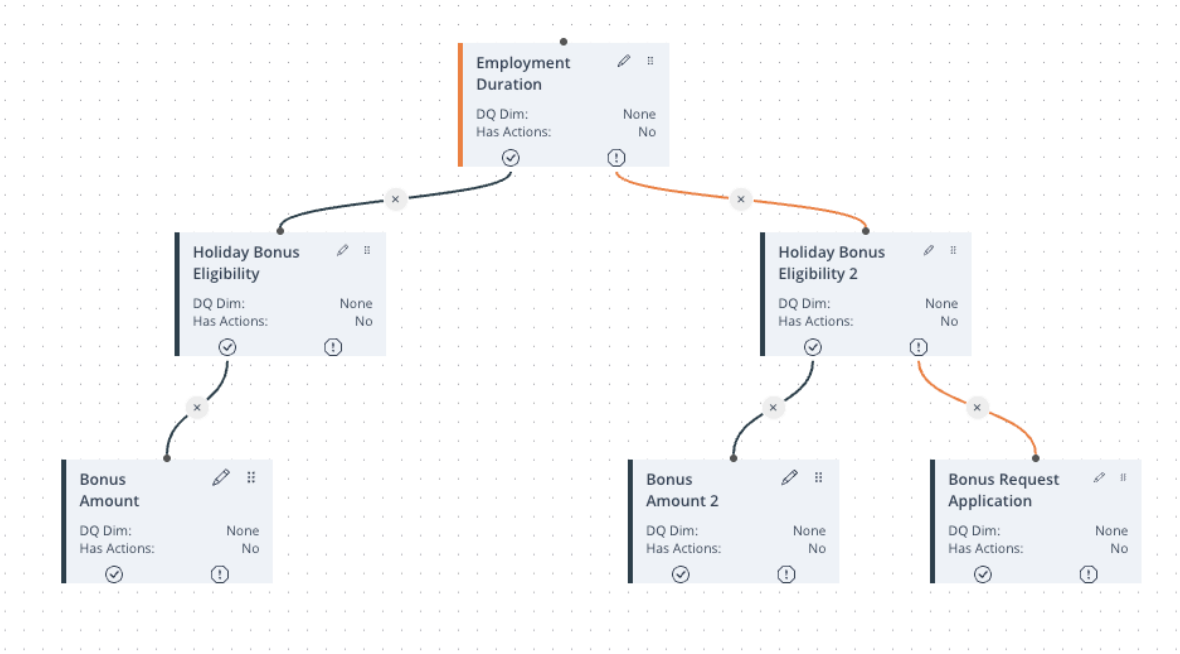


Figure 1: Business Rule Rulebook

A guide on where to find, and how to add Rulebooks and Business Rules within the HEDDA.IO user interface, is available here.

2.12 Business Rule Preparation

Sometimes Values need a bit of preparation to be correctly utilized in a Condition. This Preparation can take place in the Preparation of a Business Rule. An extensive Formula Editor helps here in creating clean data to be processed further.

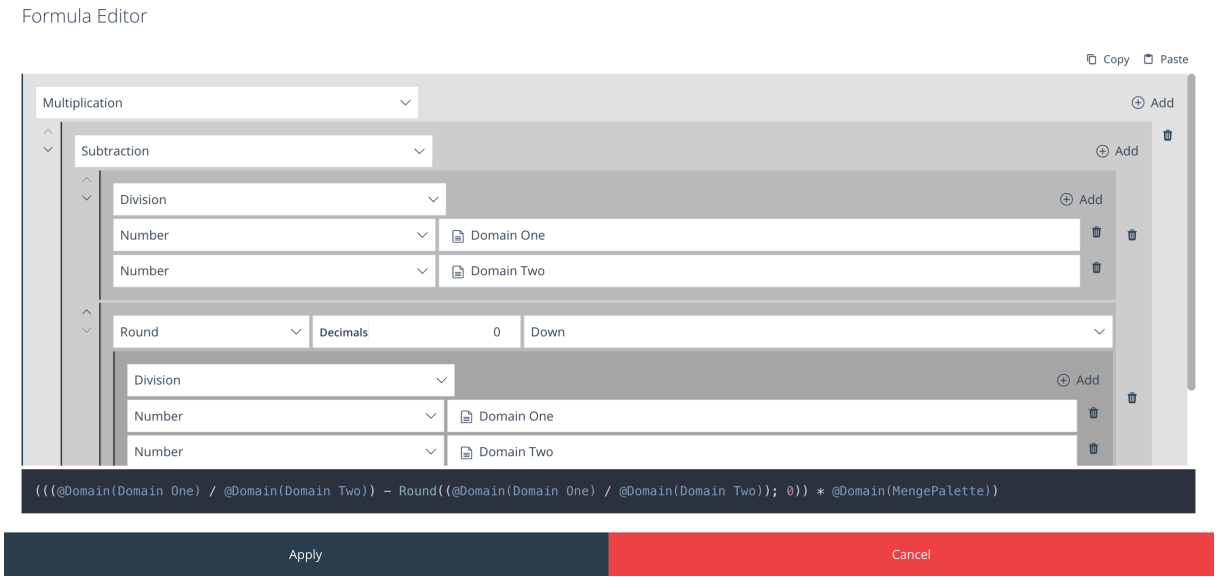


Figure 2: Business Rule Preparation Showcase

To learn more about which Formula Items are included, read Preparation.

Use Cases

Initially started as Formulas which should help utilize better data eg. fixed casing or trimmed strings, to be used in Conditions, it soon transformed into a new Step in the Processing of a Business Rule the Preparation. This approach gives directly the ability to clean the data before utilized in a Condition, it also can be used to have something like “calculated” Domains which can be directly utilized on the left and right hand side of a condition, which also includes setting those Domains from Data Links etc. So values form Data Links on the Left side of a condition are possible, too.

2.13 Business Rule Actions

In addition to incorporating conditions, users have the ability to integrate actions into Business Rules, which are executed when specific conditions, or sets of conditions, specified within a Business Rule are satisfied.

These action choices encompass functions like Set, Append, Prepend, Increment, Trim, and others. Domains and Data Links can also be applied in conjunction with these actions. For those with a knack

for creativity, it's possible to link multiple Actions together to mimic concatenation.

Use Case

In the context of concatenation for this use case, consider a dataset with two columns: `First Name` and `Middle Name`. The `First Name` column contains no null values, while the `Middle Name` column does have null values. If you prefer to consolidate both names into the `First Name` column rather than keeping them separate, you can configure a Business Rule.

This Business Rule would have one condition, checking whether the `Middle Name` column is not null, and it would include two actions. The first action would involve appending the value from the `Middle Name` column to the existing value in the `First Name` column, effectively combining them. This rule simplifies your dataset by consolidating names when a middle name is present.

1. Add condition that checks if column `Middle Name` is not null.
2. Add action that appends a white space to the `First Name` column.
3. Add action that appends the `Middle Name` value to the `First Name` value.

Further details on this are provided in the Rulebooks section for reference.

2.14 Dataset Rules

The Dataset Rules facilitate the validation of the input dataset at a broader, macro-level. Instead of individually inspecting every value within each row, these rules evaluate all the values present within a column, subjecting them to predefined conditions or a specified set of criteria.

Use Case

Let's suppose you have a column in your dataset that may contain null values. Having null values is generally acceptable, but you have a specific threshold for the percentage of rows within that column that can contain null values. You can set up a Dataset Rule to inform you when a column contains null values exceeding the specified percentage limit. This way, you can monitor and manage null values within your dataset more effectively.

A guide on where to find, and how to add Dataset Rules within the HEDDA.IO user interface, is available here.

2.15 Data Links

Data Links play a crucial role in HEDDA.IO. These are operations that involve retrieving relevant data from a reference data source based on shared identifiers to also supplement information in a dataset.

Use Case

Consider a scenario where you have a dataset featuring a list of products associated with various customers, and your focus is solely on the `client` and `product_name` columns. Imagine that one of these customers, referred to as `CustomerX`, decides to alter the names of all their products. Consequently, you find yourself in the need to update your dataset with the revised product names.

To achieve this, you initiate a Data Link pointing to a table containing the client, old product names, and new product names (let's denote these as client, old_product_name, and new_product_name). Subsequently, you formulate a Business Rule with a condition that verifies the client's name you intend to modify—in our case, it's `Client is equal to CustomerX`. Additionally, you introduce a second condition or subcondition to match the product_name in the source dataset with the old_product_name in the Data Link table.

Next, you establish an Action designed to substitute the value in the product_name column with the corresponding value from the new_product_name column in the Data Link table.

Consequently, whenever HEDDA.IO encounters a row with the value `CustomerX` in the client column and successfully aligns the product_name in the source dataset with the old_product_name in the Data Link table, it systematically replaces that value in the product_name column within the source dataset with the value derived from the new_product_name column in the Data Link table.

A guide on where to find, and how to add Data Links within the HEDDA.IO user interface, is available here.

2.16 Runs

Within the HEDDA.IO application, a Run is simply a container in which Executions and their respective Statistics are stored. Runs add context to executions so that you can then easily locate the statistics linked to respective executions, whenever you need.

Use Case

Suppose you are initiating HEDDA.IO executions from Databricks and Synapse Analytics pipelines, as well as from Polyglot Notebooks, each utilizing distinct datasets. In such a scenario, if all these executions share the same Run, it can lead to a entanglement of execution statistics, making it challenging to discern and locate specific metrics within the multitude of executions.

To address this, it is advisable to create a separate Run for each pipeline triggering a HEDDA.IO execution. This approach ensures that execution statistics for each pipeline remain organized and easily accessible, facilitating a more efficient analysis of the performance and outcomes specific to each dataset and pipeline.

A guide on where to find and how to add Runs within the HEDDA.IO user interface, is available here.

2.17 Preview

The Preview feature has been designed to provide you with the capability to Preview a sample or a full set of your data and see how it performs in the context of the Knowledge Base that it is executed from.

HEDDA.IO makes this possible without ever needing to leave the Frontend.

This functionality involves the execution of HEDDA.IO against a lookup table (or an execution) pointing at the dataset you want previewed, using a Mapping attached to a Run where the execution statistics will be saved. The results are then presented in a table format, with a toolbar that enables you to query the processed data by applying various filters.

You can also access statistical information for the processed dataset with a click of a button.

Use Case

Imagine you have an extensive dataset that you intend to process through HEDDA.IO, and you're faced with the task of configuring dozens or even hundreds of dataset and/or Business Rules. Periodically, you find it necessary to Preview the data processed against the configured Business Rules to check the performance of HEDDA.IO and to ensure accuracy and alignment with your expectations.

In such scenarios, a practical approach involves setting up a Data Link table that references a sample of your source dataset. Additionally, you can create a dedicated Run specifically for this purpose. By utilizing this setup, you can seamlessly review the performance of executing the data against the Knowledge Base, ensure that your Knowledge Base configurations is error-free and that the processed data aligns with your anticipated outcomes.

A guide on where to find and how to navigate the Preview menu within the HEDDA.IO user interface, is available here.

2.18 Preview Chaining

Preview Chaining is an extension of the Preview feature. Instead of selecting a lookup table as your data source, Preview Chaining lets you select the results of a previous HEDDA.IO Run execution, while giving you the option to filter for valid and invalid results.

2.18.1 Runner Data Upload

This functionality extends beyond the HEDDA.IO Frontend; within the HEDDA.IO Runner, you can also enable Data Upload.

A guide on how to use Runner Data Upload within the HEDDA.IO is available here.

2.19 Mappings

In HEDDA.IO, Mapping is an operation that associates Domains with columns from a dataset. It ensures that the data within different Domains and columns are correctly aligned, enhancing the overall consistency and accuracy of the dataset.

A guide on where to find and how to add Mappings within the HEDDA.IO user interface, is available here.

2.20 Knowledge Base Versioning

In HEDDA.IO, the initial step to utilize a Knowledge Base involves publishing it. This functionality serves a dual purpose, allowing the establishment of a version history as well as the ability to have the Knowledge Base reviewed before it being published. The primary page of the Knowledge Base is perpetually set in Read Only mode. This signifies that you can exclusively view or read values, without the capability to alter, delete or add entities.

To achieve any of the three actions, you must switch to the Edit Version mode, accessible by selecting `Edit Version` located at the upper-right corner.

Details on how to add, modify, or remove entities are outlined in the relevant documentation corresponding to each section within the Project Navigation Bar.

More information on the Edit Mode is available here

2.21 Tags

Tags can be used to differentiate between Rulebooks or sets of Rulebooks designated for specific purposes. For example, you might have a Tag called 'Financial' for Rulebooks related to financial data and another Tag called 'HR' for Rulebooks pertaining to human resources data. Tags can also be attached to Runs to ensure that the source data is executed only against Rulebooks associated with that Tag. This allows for targeted and efficient processing based on the tagged Rulebooks.

2.22 Knowledge Base Audit

The **Audit Portal** feature in the HEDDA.IO lets users see all changes made to the Knowledge Base since it was created. It shows what was modified, what type of component it is (like Domains, Rulebooks, and so on), when the change happened, who made the change, the Knowledge Base version at the time, whether it was added, removed, or changed, and if the change was published or still in draft. Users can filter by specific component types and states, or set a custom date range, making it easy to track and review all updates.

More information on the Audit Portal is available here

3 User Interface

3.1 Homepage

The initial point of interaction with the HEDDA.IO application is the Home page. Here you can either initiate the creation of a new Project or access existing ones.

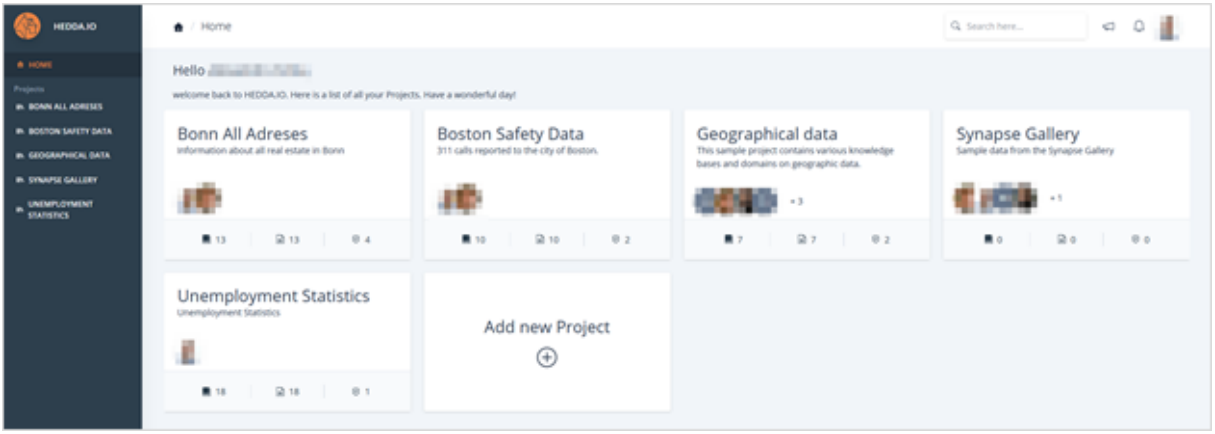


Figure 3: Homepage

On the left-hand side, the Projects Browsing Panel is clearly visible. Positioned at the upper part is the Header, and on the bottom side is the Footer. All three components are conveniently accessible from any location within the application.

3.2 Projects Browsing Panel

Situated in the upper part of the Projects Browsing Panel, just beneath the Logo, is the Home button. A simple click on this button will bring you back to the Homepage.

Directly under the Home button lies the list of already created Projects. Selecting a Project from the list will result in its expansion, revealing the dashboard button and a list of associated Knowledge Bases specific to the chosen Project.

Furthermore, clicking on a Project will also navigate you to the Project Dashboard, a point we will revisit shortly. The Project Dashboard can also be accessed by selecting the corresponding Project card located in the Projects Overview of the Home page.

At the bottom of the side menu bar, you will find the HELP button. 

Clicking this button will navigate you to the help page, where you can find a full documentation of HEDDA.IO.

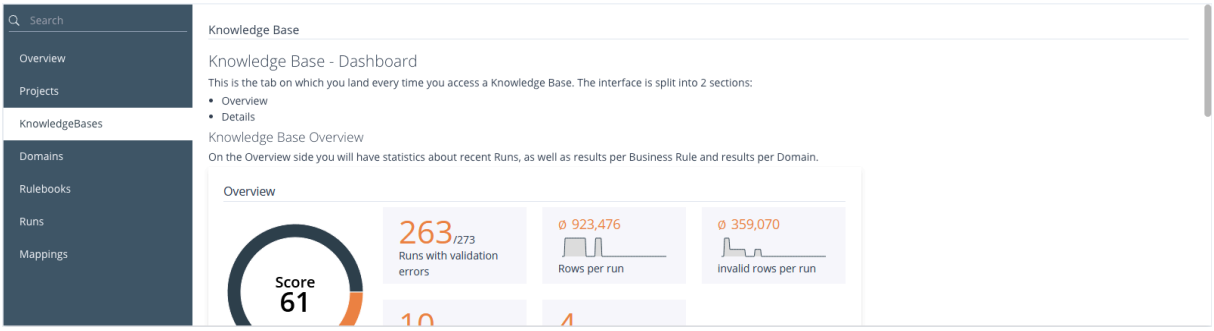


Figure 4: Help Button in Sidebar

3.3 Header

Towards the left side of the Header, you will find the navigational breadcrumbs, that will take you back to a particular page in the navigational path, including the Home page.

Figure 5: Header

Situated on the right side of the Header, you will find a search bar designed to aid you in locating any object within HEDDA.IO, provided you possess the appropriate access. This is a Full Text Search which searches through the Names and Descriptions of the items. In the Search Results you can further refine the list by selecting a specific type or including items found in Edit Versions.

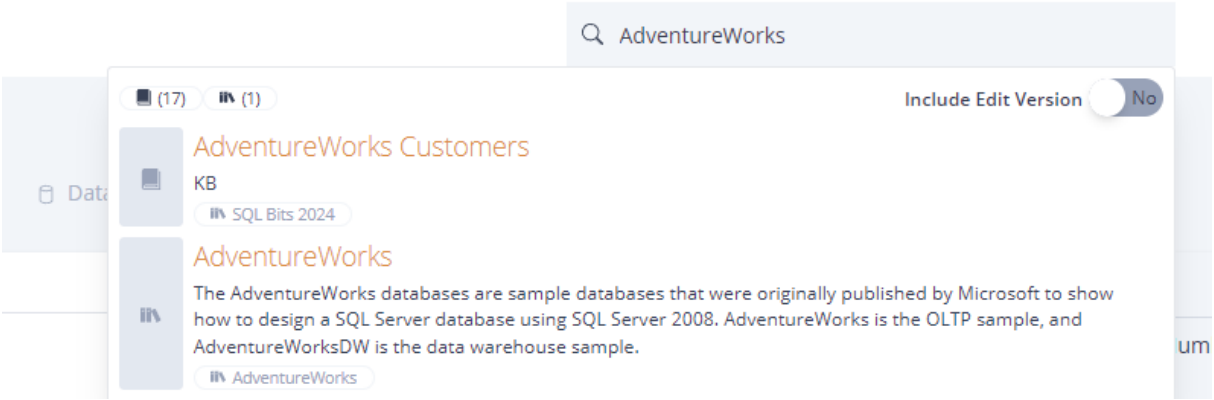


Figure 6: Global Search

Adjacent to it, the Announcements and Notifications buttons can be found; upon clicking, these will reveal the history of Announcements and Notifications.

At the far right of the Header, you have the option to click on the user's profile picture, thereby opening the user profile menu. This menu facilitates actions such as switching the application's language and managing API Keys.

3.3.1 API Keys Manager

Accessing the API Keys Manager involves selecting the API Keys button located within the user profile menu. This action opens a panel that allows the user to see existing API Keys and generate fresh ones.

> Previously created API Keys are not displayed fully.

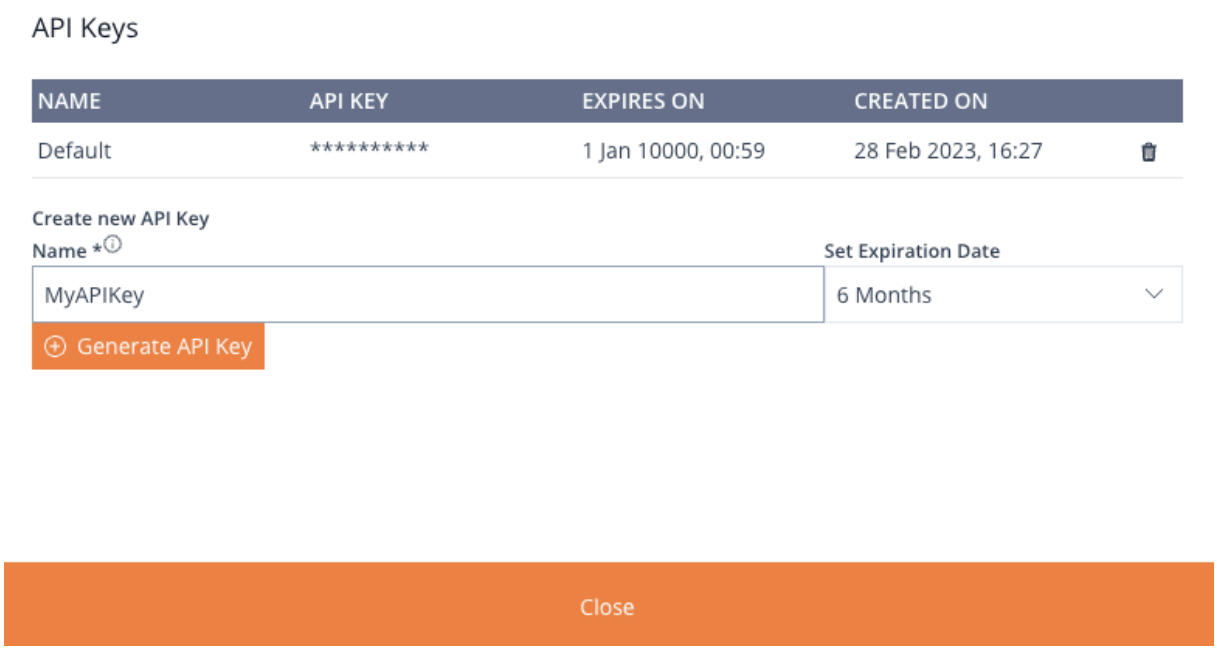


Figure 7: API Key Manager

Positioned at the upper section of the panel, you can see a table containing all generated API Keys. This table is divided into four columns: NAME, API KEY, EXPIRES ON, and CREATED ON. Situated at the far right of each API Key is a delete button.

3.3.1.1 Create New API Key The subsequent section within the API Keys Management panel facilitates the generation of a fresh API Key. Simply enter a name in the designated field and specify the

desired expiration date for the API Key. Conclude the process by clicking the `Generate API Key` button.

Upon completion, click the `Close` button located at the bottom , or the `X` icon button in the top-left corner of the panel.

3.3.2 Language Selection

The language selection is available under the user profile button, which you can always find in the utmost top-right corner of the application.

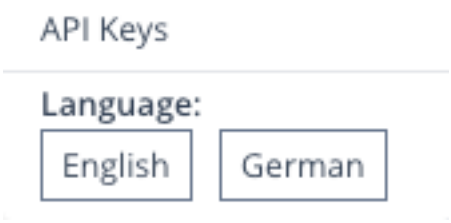


Figure 8: Language Selection

> Currently, HEDDA.IO supports two languages: English and German.

3.4 Footer

On the bottom side of HEDDA.IO, we have the footer, which contains three important pieces of information. The logo of the developer, the copyright marker, and the current version of HEDDA.IO.

Figure 9: Footer

3.5 Projects Overview

In the middle of the Homepage, the `Add New Project` button is positioned prominently.

Once you have already created Projects, each Project will be presented in distinct cards containing the Project Name, Description, and the Users with granted access. Additionally, these cards provide essential metadata, including the count of Knowledge Bases, Domains, and Business Rules housed within each Project.

Filter options allow for filtering of Project Names or User Permissions. The Permission Filters are as follows:

- Only Manage: Show all Projects where you have Manage permissions
- Only Write: Show all Projects where you have Write permissions
- Only Read: Show all Projects where you have Read permissions
- Owner: Only show Projects where you are the Project Owner
- Involved: For Superadmins only. Shows all Project where you are either Owner or a User

Additionally, the user is allowed to change the view of the Project list from Cards to List and vice versa, by clicking the 'Show List View' or 'Show Card View' button respectively.

To the very right, there is the 'Add new Project' button, which opens the 'Add Project' form. The same button exists in form of a Card, in the Card view, as the last item.

3.5.1 Add New Project

Clicking the `Add New Project` button will open up the drawer containing the `Add Project` form.

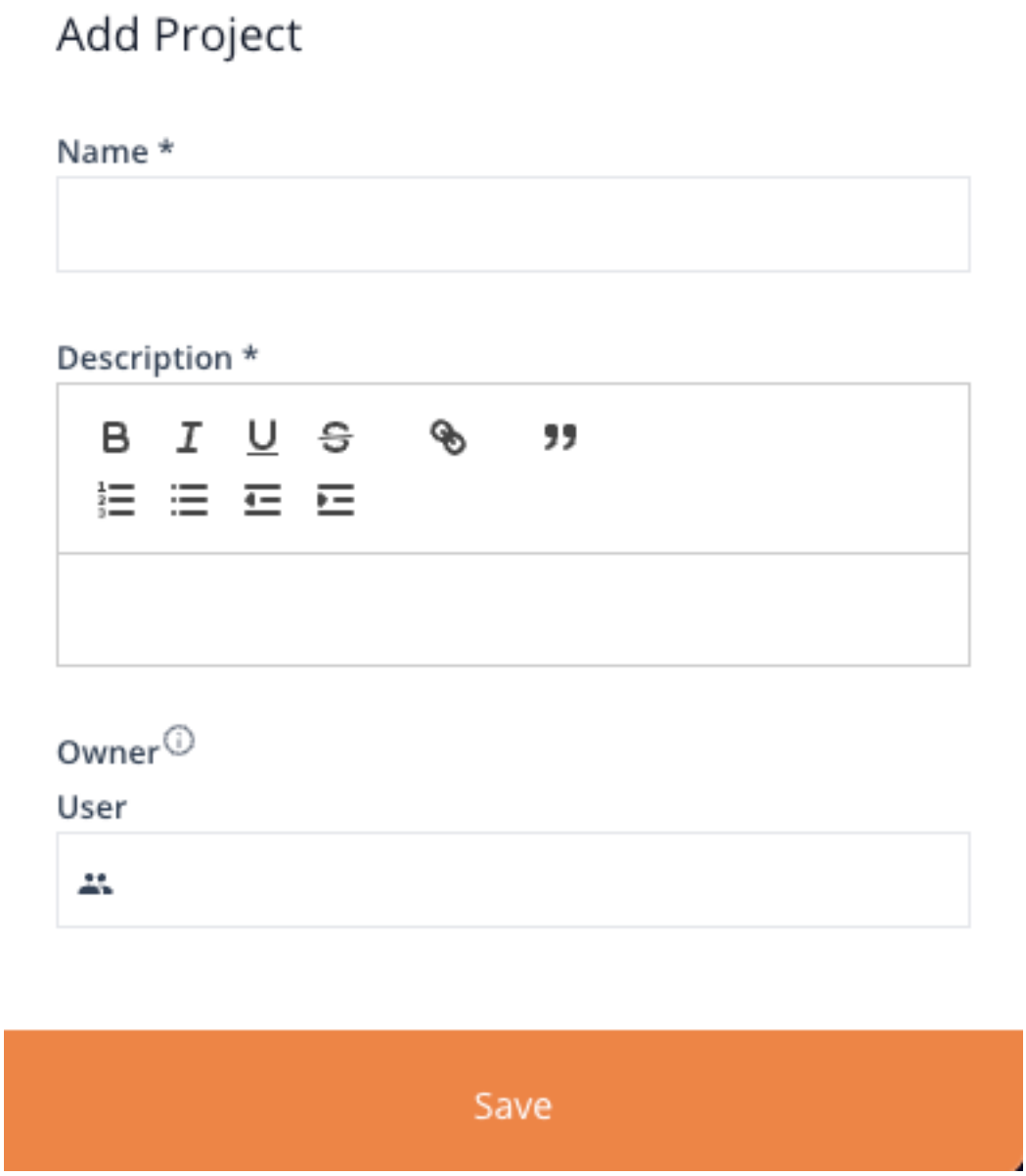


Figure 10: Add Project

The form allows you to set a Name, a Description for the Project, as well as select the Owner of the Project.

Once done, you can click the `Save` button at the bottom of the drawer to save and create the Project. The Project can now be accessed from the side bar menu or from the Project Overview.

3.6 Favorite Overview

By clicking on the tab `Favorites` next to the `Projects` tab, the Favorites Overview is shown.

This overview shows all favorited Projects, Knowledge Bases and their content as cards. Each card shows the type, name and description of the Favorite. When the mouse is moved over a Favorite Card, icons to edit and remove a Favorite are shown and can be used.

Clicking on a Favorite Card will open the favorited item. Clicking on the `Edit` button opens a form on the right side of the screen, where the name and description of the Favorite can be changed without changing favorited item. Changes are saved upon clicking on `Save` or can be discarded by clicking on the 'X' in the top left of the form. Clicking on the `Remove Favorite` Button will show a security prompt to avoid deleting a Favorite by mistake. The Favorite is removed by clicking on `Ok` or cancelled by clicking on `Cancel`. Removing a Favorite will not remove the favorited item.

Favorites can be created by clicking on the Favorite Icon ☆. It can be found in the details views of Projects, Knowledge Bases and their content. Clicking the Favorite Icon will open a form where the name and description of the Favorite can be changed. Upon clicking on `Save`, the Form will close and the Favorite is created and it can be found in the Favorite Overview. This will not change the item you want to favorite. If a Favorite already exists, the Favorite Icon will be filled and it will behave like the `Remove Favorite` Button on the Favorite Cards.

The Favorites can also be filtered by their name or by the type of the Favorite.

3.7 Monitoring Hub Overview

By clicking on the tab `Favorites`, the Favorites Overview is shown.

The Monitoring hub shows recently Run Executions of monitored Runs. A Run can be added to be monitored or removed from monitoring on the Run Detail page.

The Monitoring Page lists recent Executions starting with the most recent Run. Preview Executions are hidden by default, but can be shown using the Flter button. Clicking on a Run opens the detailed Run Statistics of that Execution.

Monitored Executions can be filtered by the name of their Project, Knowledge Base or Run. By default, Run Executions of the last 7 days are shown. The range can be changed using the datepicker. Similarly, Preview Run Executions are hidden by default, but they can be shown using the respective filter option.

Hovering over the name of a Project, Knowledge Base or Run will show a search icon to the right of the name. By clicking on it, the search text will be set to that name.

3.8 Project Page

Projects are essentially workspaces within HEDDA.IO, designed to cater to your distinct needs for utilizing the HEDDA.IO platform. This includes, Project-wide, necessary External Connections, Project Service Principal and Events as well as user management.

Upon accessing a Project, the first element we see is the Project Name. Directly underneath resides the Project Navigation Bar.

3.8.1 Project Navigation Bar

The Navigation Bar comprises four tabs:

- Dashboard
- External Connections
- Events
- Event Log

The canvas situated below the navigation bar adapts its content according to tabs selected on the bar.

3.8.2 Project Dashboard Page

Upon selecting a Project, the user is directed to the Project Dashboard as their initial landing point.

The Project Dashboard is divided into two distinct sections: the left-middle panel, which is the Dashboard Panel, encompasses the Statistics and the Knowledge Base Overview, and the Project Info panel located on the right-hand side, which encompasses the Details, Owner, the Users with access to the Project, and a list of Service Principals.

3.8.2.1 Statistics Positioned at the upper side of the Dashboard Panel is the Project statistics section.

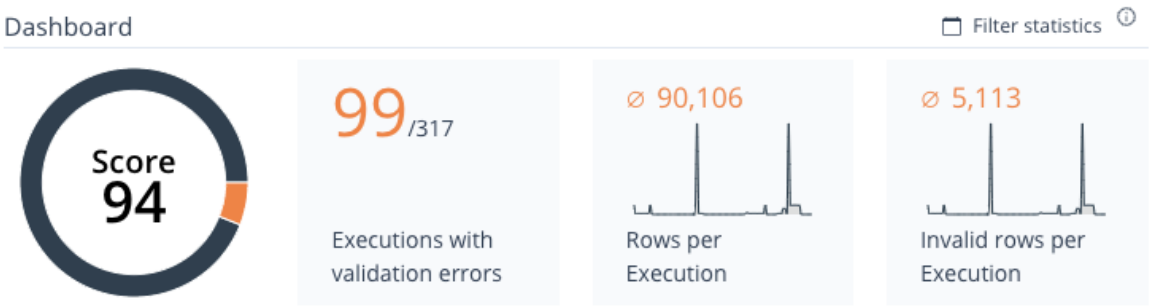


Figure 11: Project Statistics

This section contains statistics that include:

- The overall score ranging from 0 to 100.
- Count of Executions with validation errors.
- Count of rows per Execution.
- Count of invalid rows per Execution.

At the upper-right corner is a statistics filter that allows you to refine your statistical queries by specifying a date range.

By clicking the `Filter statistics` button, you can access the `Filter statistics by Date Range` panel.

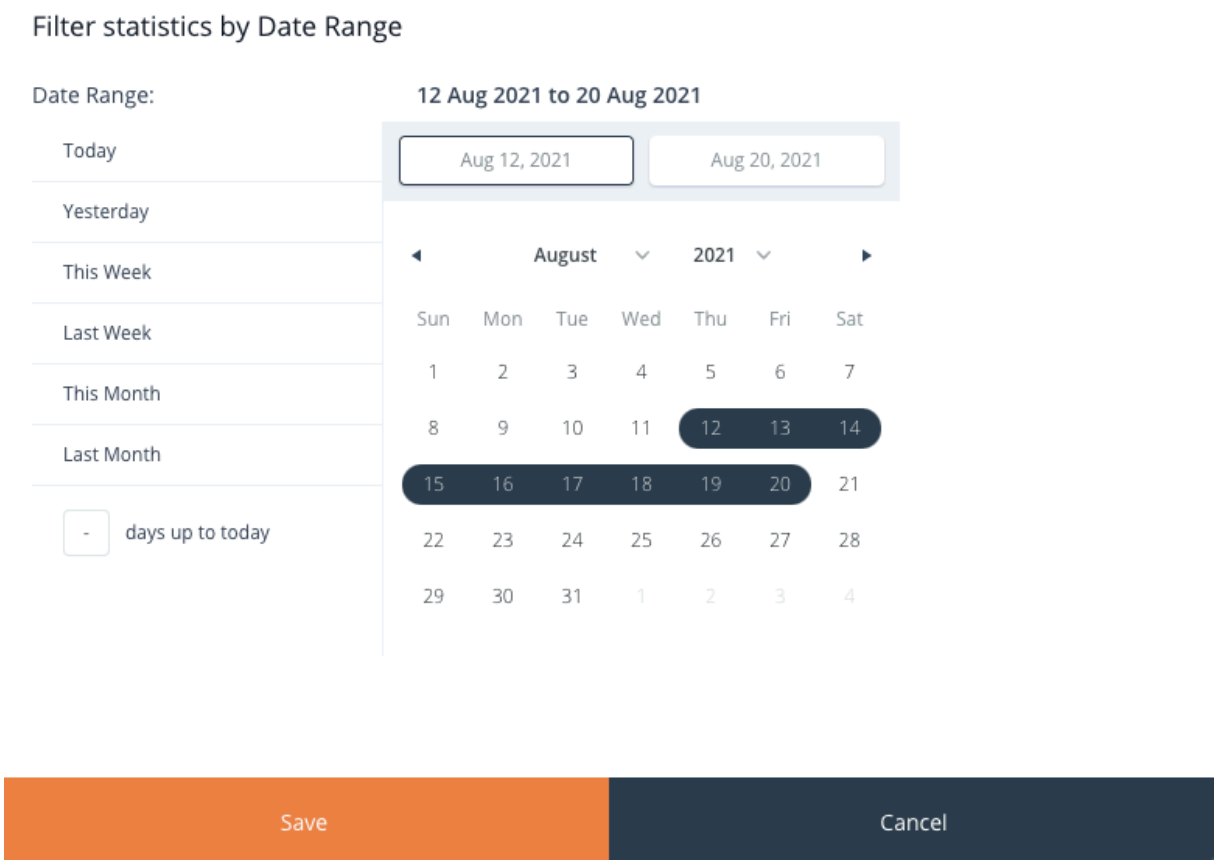


Figure 12: Filter Project Statistics

Within this panel, on the left-hand side, you have the option to choose from a range of predefined date periods, including:

- Today
- Yesterday
- This Week
- Last Week

- This Month
- Last Month

Directly below, you can specify a certain number of days leading up to the current date.

Alternatively, you also have the flexibility to define a custom date range using the date picker tool located on the right side.

3.8.3 Knowledge Bases Overview

Within the corresponding Project, this section showcases the roster of Knowledge Bases. It is located right below the Statistics section.

You have the capability to vertically navigate through the list, opt for the specific Knowledge Base you wish to access, or create a new Knowledge Base, by clicking the `Add Knowledge Base` button located in the top-right corner of the section.

Figure 13: Project Knowledge Bases

Each entry in the Knowledge Base list includes the following details:

- Latest published Knowledge Base version
- Name
- Description
- Last edit date and time
- Category
- Data Responsibility Office
- Domain count
- Business Rules count
- Runs count

By engaging the Arrow icon positioned on the right-hand side of a given Knowledge Base entry, you can navigate to the respective Knowledge Base page.

The buttons to the left of the Arrow, tagged as Domains, Rules, and Runs, will also navigate you to the respective Domains, Rules, and Runs, when clicked.

Furthermore, in case you have unpublished Knowledge Bases, they will be displayed accordingly in this section, under `Not Yet Published`.

For an in-depth look at the Knowledge Bases, click here.

3.8.4 Project Info Panel

On the right-hand side of the Project's dashboard, you will find the Project Info Panel. In the upper right corner of this panel, an `Edit` button is positioned, granting you the capability to revise the Project's name, description, and Owner, or delete the Project.

In the Details section, you will see the description of the Project.

Figure 14: Project Details

Under the Details section, the Project Owner's name is displayed, followed by a roster of users with varying degrees of permissions to the Project. Adjacent to this, is a Manage User button, designed to facilitate the administration of user permissions. Lastly, the Service Principal displays the configured Service Principle, and the Add button allows you to add the Service Principal if none is configured.

3.8.4.1 Edit/Delete Project

Locate and click the Edit button in the top-right corner of the Project Info Panel. This will open the Edit Project drawer, allowing you to change the name, the description and the Owner of the Project.

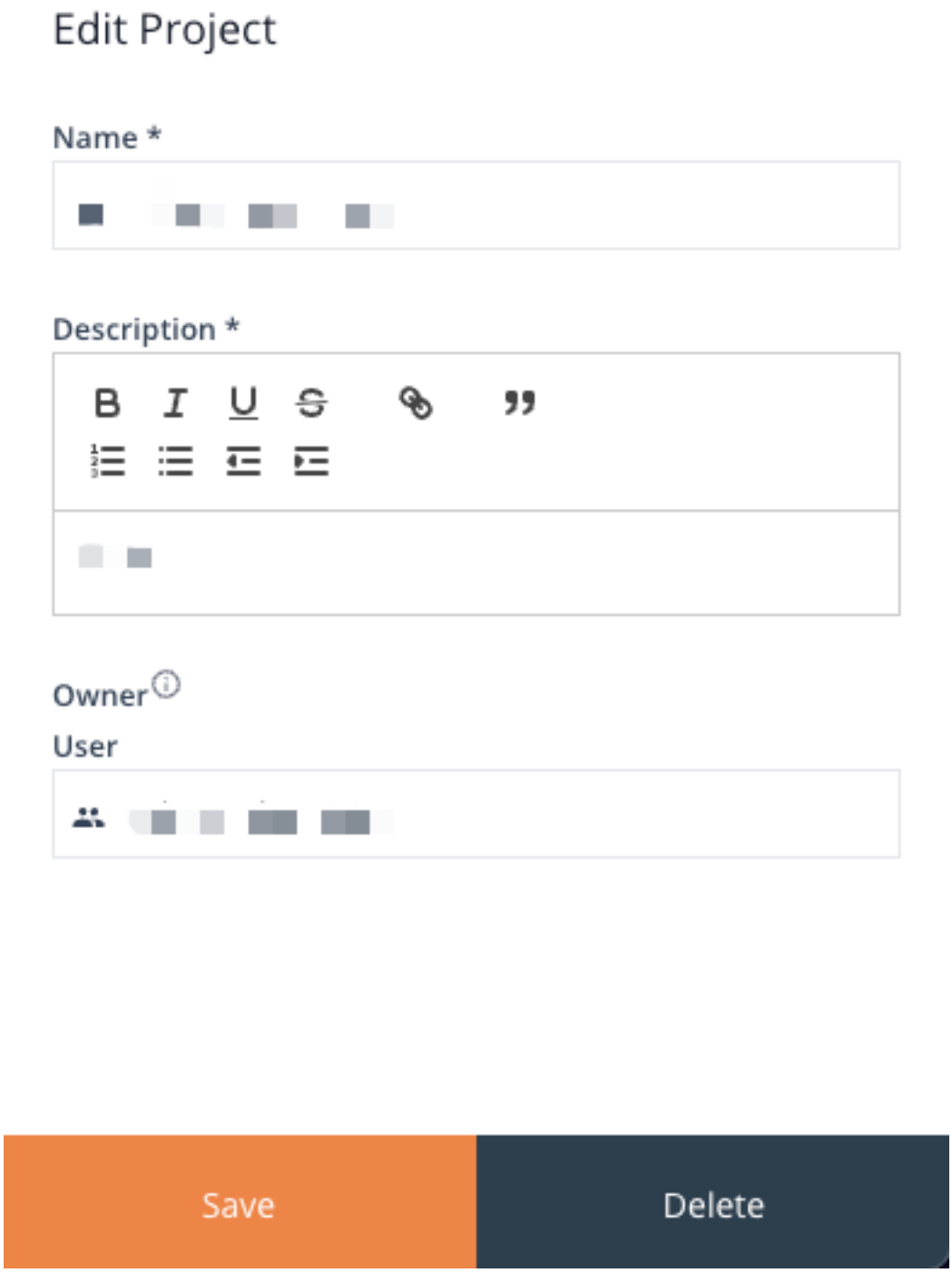


Figure 15: Edit Project

To Delete the Project, locate the `Delete` button in the bottom-right corner of the drawer.

3.8.4.2 Manage Users Upon clicking the `Manage Users` button, the User Permissions drawer will open, allowing you to adjust the access level for users who were already given permissions, as well as give permissions to new users.

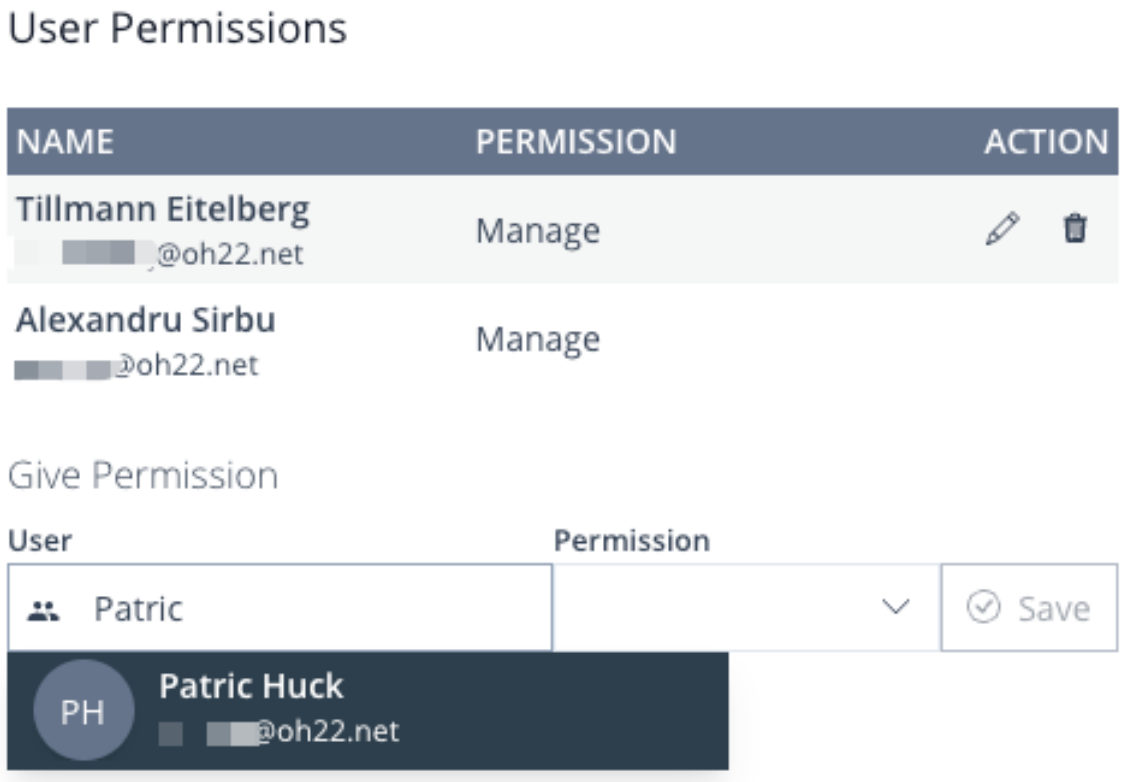


Figure 16: Manage Users

In the upper section, you can find a list of all the users that were already given permissions.

Below that, you have the Give Permissions section, which allows you to select Members and Permissions you want to grant them.

You can simply start typing the name of the user and a dropdown will present you with suggestions. The suggestions are fed from the Azure Active Directory Tenant associated with the environment to which HEDDA.IO has been deployed.

3.8.4.3 Add/Edit/Delete User Permissions To the right of each user, you can see the `Edit` and the `Delete` buttons.

Clicking the `Edit` button will allow you to edit the permissions for the selected user. Click `Save` to save changes.

Clicking the `Delete` button will remove all permission from the user.

3.8.4.4 Project Service Principal Configuration The Project Service Principal can be configured in the Info panel on the Project Dashboard page. In the Info panel, you can see whether a Project Service Principal is configured. To add or edit the Service Principal information, the user must have Manage privileges for this Project.

The following information will be required:

- **Name**: The name of The Service Principal
- **Tenant Id**: The Tenant Id where The Principal REsides
- **Client Id**: The Client Id of the Service Principal
- **Client Secret**: The Client Secret of the Service Principal

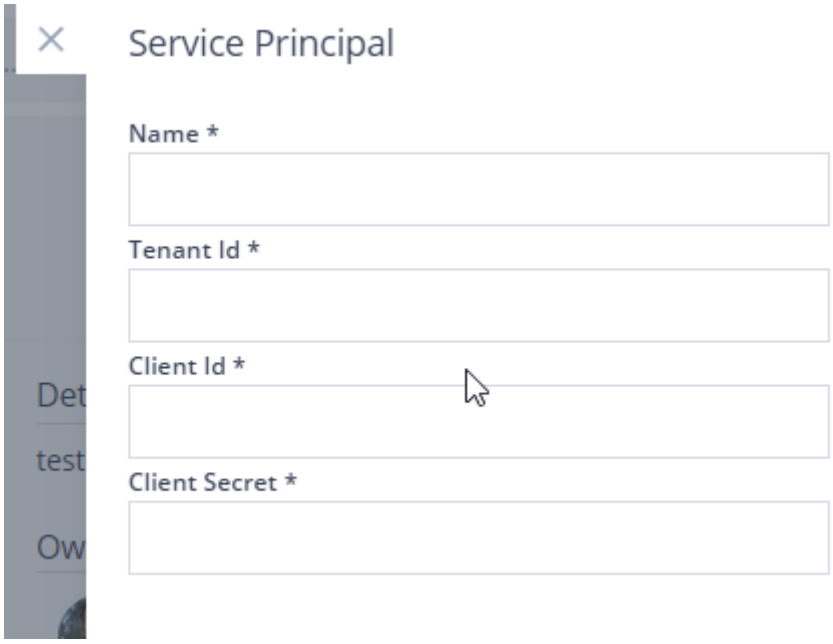


Figure 17: Project Service Principal Form

3.8.4.5 Project API Key The Project API Keys Manager can be found in the Details Panel of the Project Dashboard. Pressing the Edit button opens a panel that allows the user to see existing Project API Keys and generate fresh ones. This requires Manage permissions on the Project.

Positioned at the upper section of the panel, you will encounter a table containing all generated API Keys. This table is divided into four columns: NAME, API KEY, EXPIRES ON, and CREATED ON. Situated at the far right of each API Key is a delete button.

3.8.4.5.1 Create New Project API Key The subsequent section within the API Keys Management panel facilitates the generation of a fresh API Key. Simply enter a name in the designated field and

specify the desired expiration date for the API Key. Conclude the process by clicking the `Generate API Key` button.

Upon completion, click the `Close` button located at the bottom , or the 'X" icon button in the top-left corner of the panel.

Project API Keys are limited to Read and Write permissions for the Project.

3.8.4.6 Git Configuration

"Git" with a line "Git is not configured" should be visible in the Details to the right of the page if Git support is enabled in HEDDA.IO. A form to configure your remote Git connection opens by clicking on "Add".

The following information is required:

- **Remote Url**: Remote Url of the remote Git Repository
- **Username**: Username used to access the remote Git Server
- **Password/Access Token**: Access Token used to access the remote Git Server
- **Environment Branch Name**: Name of the branch used for the Project. Default configured under `EnvironmentBranchName`. It is best practice to use one Repository per Project.
- **Linked Environment Branch**: Optional. Name of the Environment Branch Name of an Environment saved in the same Repository. Used to import changes from Projects from other Environments

> **Note:** Saving this form for the first time will move all Knowledge Bases to Git. This cannot be reverted.

3.9 External Connections Configuration

The External Connections page can be accessed by clicking the `External Connections` tab in the Project navigation bar.

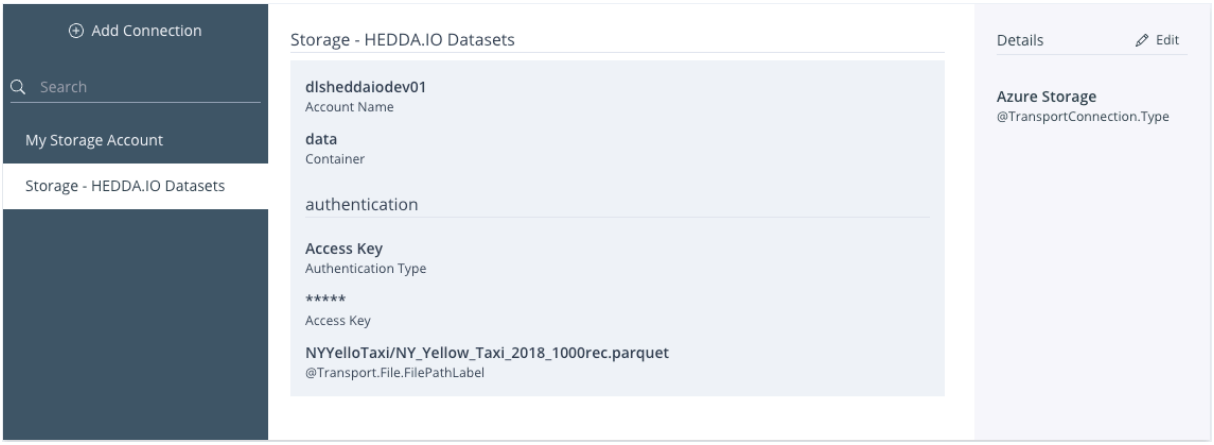


Figure 18: External Connections Overview

The page is split into three distinct panels:

- Connections Browsing Panel
- Connection Overview
- Connection Info Panel

3.9.1 Connection Browsing Panel

The Browsing Panel, situated on the left side of the External Connections overview, contains the Add Connection button at the top. Below this, there is a list of External Connections accompanied by a search option. The search function allows you to quickly filter and find External Connections by matching the sequence of characters you enter with items in the list.

3.9.2 Add Connection

Locate and click the `Add Connection` button located at the top of the Browsing Panel. This will open the `Add Connection` drawer containing a form that will allow you to establish a connection with one of the currently supported providers:

- Azure Data Lake Storage
- Azure Blob Storage
- Databricks
- Microsoft Fabric Lakehouse
- Microsoft OneLake
- Microsoft SQL Server
- Reference Data Service

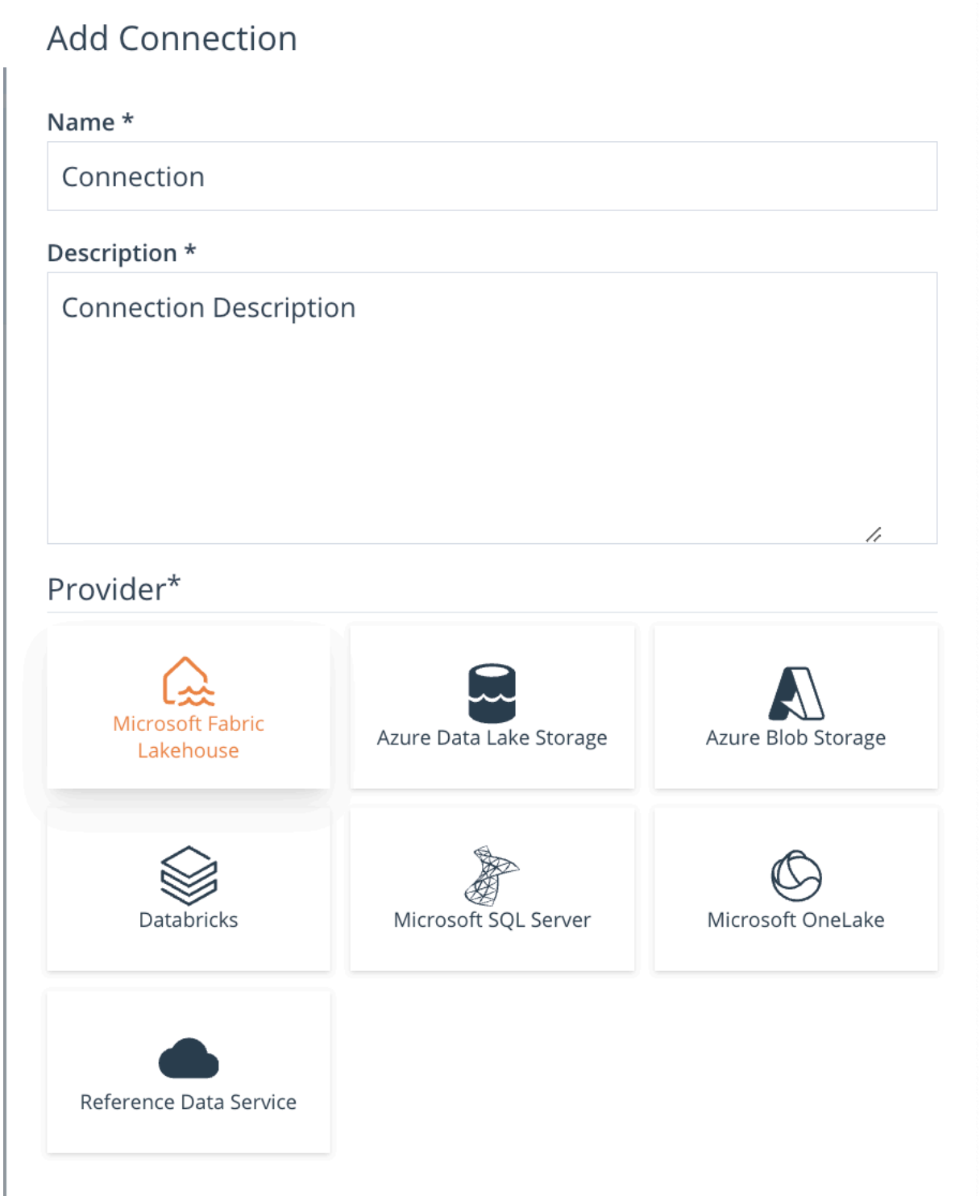


Figure 19: Add External Connection

The first two input fields are shared and will allow you to set a Name and a Description for the Connection. Selecting one of the providers will reveal a set of inputs which are required in order to suc-

cessfully configure a connection with the respective provider in HEDDA.IO.

3.9.2.1 Connection Types

Let's run through all the supported providers and the requirements associated with configuring one.

3.9.2.1.1 Microsoft Fabric Lakehouse

- **Server**: The network address or hostname of the Microsoft Fabric Lakehouse instance.
- **Database**: The name of the specific database within the Microsoft Fabric Lakehouse instance to connect to.
- **Connection Type**:
  - **Custom Service Principal**:
    - **Client ID**: A unique identifier for the service principal within your Azure Active Directory (AAD) tenant.
    - **Client Secret**: A password or key used to authenticate the service principal when connecting to the Microsoft Fabric Lakehouse.
  - **Project Service Principle**: Will select the configured Project wide Service Principle.
  - **Service Principle**: Will allow you to copy the name and id of the HEDDA.IO service principle.

3.9.2.1.2 Azure Data Lake Storage

- **Account Name**: The unique name of your Azure Storage account.
- **Container**: The specific container within your Azure Storage account where data is stored.
- **Authentication Type**:
  - **Custom Service Principle**:
    - **Tenant ID**: The directory (tenant) ID from your Azure Active Directory.
    - **Client ID**: The application (client) ID of your service principal.
    - **Client Secret**: The secret key associated with your service principal for authentication.
  - **Project Service Principle**: Will select the configured Project wide Service Principle.
  - **Service Principle**: Will allow you to copy the name and id of the HEDDA.IO service principle.
  - **Access Key**:
    - **Access Key**: The primary or secondary access key for your Azure Storage account, used for authentication.

3.9.2.1.3 Azure Blob Storage

- **Account Name**: The unique name of your Azure Storage account.
- **Container**: The specific container within your Azure Storage account where data is stored.
- **Authentication Type**:
  - **Custom Service Principle**:
    - **Tenant ID**: The directory (tenant) ID from your Azure Active Directory.
    - **Client ID**: The application (client) ID of your service principal.
    - **Client Secret**: The secret key associated with your service principal for authentication.
  - **Project Service Principle**: Will select the configured Project wide Service Principle.
  - **Service Principle**: Will allow you to copy the name and id of the HEDDA.IO service principle.
  - **Access Key**:
    - **Access Key**: The primary or secondary access key for your Azure Storage account, used for authentication.
  - **Anonymous** (for anonymous authentication): Allows access to public containers without any authentication requirements. You need to have Anonymous authentication enable on your Blob Storage and Blob container.

3.9.2.1.4 Databricks

- **Databricks API**:
  - **All Purpose Cluster**: a versatile Databricks cluster for running various workloads, including data engineering, data science, and data analytics tasks.
  - **SQL Warehouse**: a Databricks managed compute resource optimized for running SQL queries and BI workloads.
- **Host**: The URL of the Databricks workspace where the cluster is located.
- **Cluster ID**: A unique identifier for the specific Databricks cluster to which you want to connect.
- **Schema**: The database schema within the catalog that specifies the organization of data tables. It is an optional value in the following format:<catalog>.<schema> (i.e.hive_metastore.default)
- **Personal Access Token**: A secure token used for authenticating and authorizing access to the Databricks API.
- **Request Timeout** (seconds): The maximum duration, in seconds, allowed for a request to the Databricks server before it times out.

3.9.2.1.5 Microsoft SQL Server

- **Connection Type**:
  - **Custom Service Principal**:
    - **Server**: The network address or hostname of the SQL Server instance.
    - **Database**: The name of the specific database on the SQL Server to which you want to connect.
    - **Client ID**: A unique identifier for the service principal within your Azure Active Directory (AAD) tenant.
    - **Client Secret**: A password or key used to authenticate the service principal when connecting to the MS SQL Server.
  - **Project Service Principle**: Will select the configured Project wide Service Principle.
    - **Server**: The network address or hostname of the SQL Server instance.
    - **Database**: The name of the specific database on the SQL Server to which you want to connect.
  - **Service Principle**: Will allow you to copy the name and id of the HEDDA.IO service principle.
    - **Server**: The network address or hostname of the SQL Server instance.
    - **Database**: The name of the specific database on the SQL Server to which you want to connect.
  - **Credentials**
    - **Server**: The network address or hostname of the SQL Server instance.
    - **Database**: The name of the specific database on the SQL Server to which you want to connect.
    - **User**: The username required to authenticate and access the SQL Server.
    - **Password**: The password associated with the user account for accessing the SQL Server.
  - **Connection String**: A string that contains all the necessary information to establish a connection to the SQL Server, including server, database, user, and password details.

3.9.2.1.6 Microsoft One Lake

- **Workspace**: The name of the Microsoft OneLake workspace where your data items are stored.
- **Item**: A specific data object or file within the OneLake workspace that you want to access.
- **Authentication Type**:
  - **Custom Service Principle**:
    - **Tenant ID**: The directory (tenant) ID from your Azure Active Directory.
    - **Client ID**: The application (client) ID of your service principal.
    - **Client Secret**: The secret key associated with your service principal for authentication.

- **Project Service Principle**: Will select the configured Project wide Service Principle.
- **Service Principle**: Will allow you to copy the name and id of the HEDDA.IO service principle.

3.9.2.1.7 Reference Data Service

- **URL**: The web address of the Reference Data Service endpoint that you want to connect to.

3.9.3 Connection Overview

In the central panel, you can locate the configuration settings for the selected connection.

3.9.4 Connection Info Panel

On the right-hand side, you will find the Details section, showing the Description and the connection type.

3.9.4.1 Edit/Delete Connection In the top-right corner of the Info Panel, you will notice the `Edit` button. This button allows you to make edits to connections or delete them.

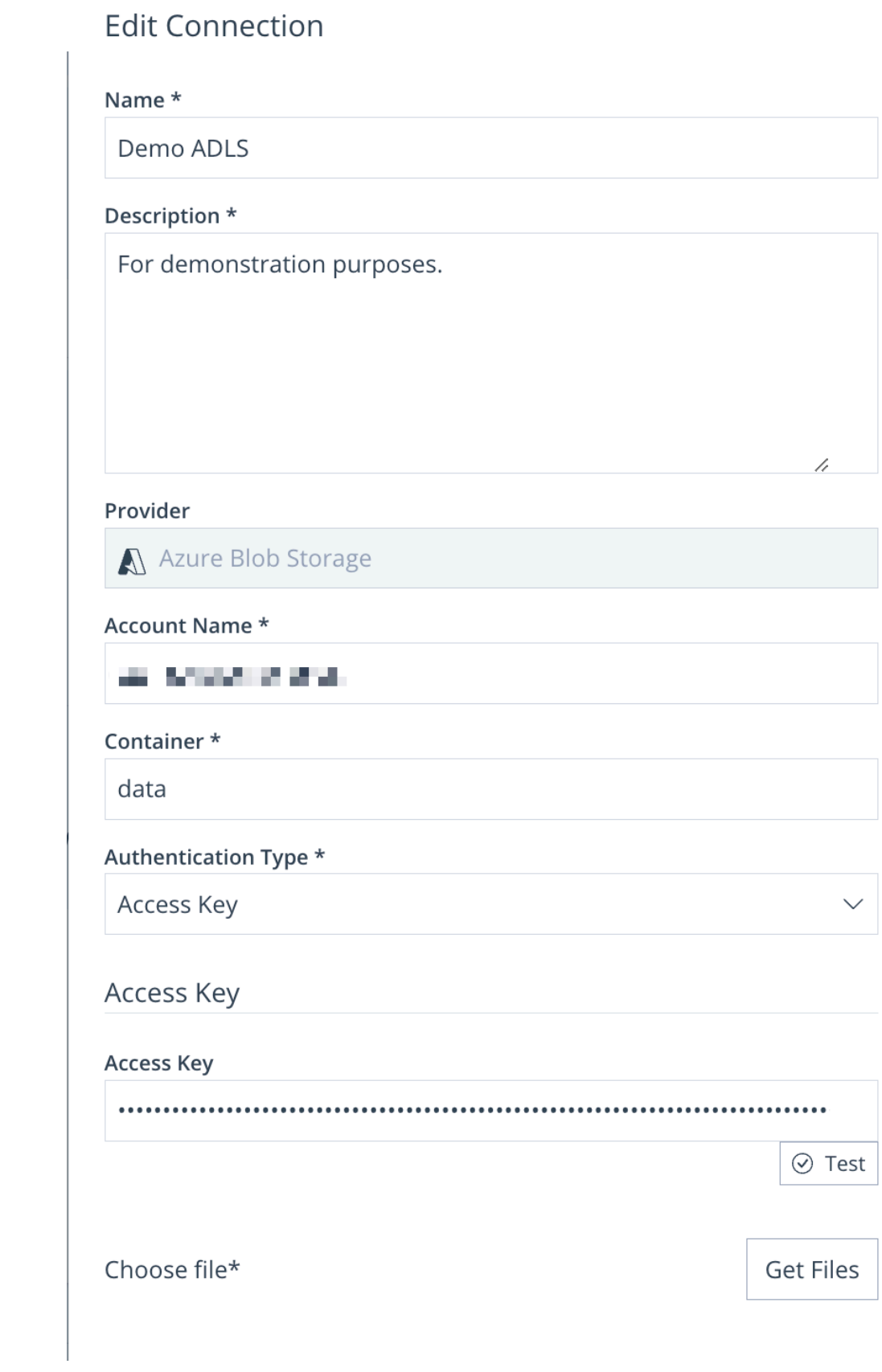


Figure 20: Edit External Connection

Here's an example of what the Azure Storage connection form looks like. Within this form, you can see a `Test` button that is designed to assess the connection. Once the configuration is accurate, clicking the `Get Files` button should unveil all the contents within the designated container. This will enable you to navigate through directories and opt for the specific file you intend to choose.

Click `Save` or `Delete` to save or delete the connection.

3.10 Events Page

Using Events in HEDDA.IO, you can send a Notification via Teams, Email, or other, when a new Execution has finished. You can also create work item in DevOps when a Dataset Rule has failed.

To access the Events page, navigate to the `Events` tab located in the top Project Navigation Bar.

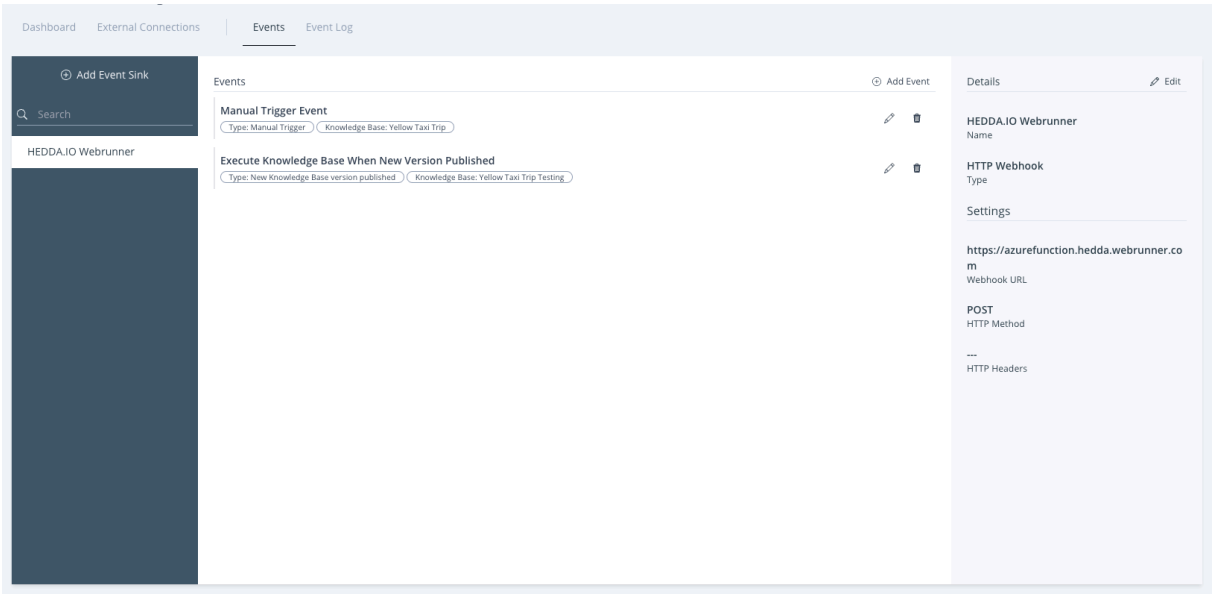


Figure 21: Events Overview

The page is split into three panels:

- Event Sinks Browsing Panel
- Events Overview
- Event Sink Info Panel

3.10.1 Event Sink Browsing Panel

This panel is located on the left side of the page.

At the top, you can see the `Add Event Sink` button. Just below that, there is a search input field designed to assist you in locating sinks, which are listed right below the search bar.

3.10.1.1 Add Event Sink

Before creating an Events, you must first create an Event Sink. In simpler terms, this entails forming a link to the service that should respond to our Events.

HEDDA.IO currently supports the following Event Sink types:

- Azure Event Hub
- Azure DevOps
- SMTP Email
- Azure Event Grid
- Slack
- Teams
- Twillio
- Webex
- HTTP Webhook

In order to add an Event Sink, locate and click the `Add Event Sink` button. This will open the `Add Event Sink` drawer that contains a form for configuring an Event Sink.

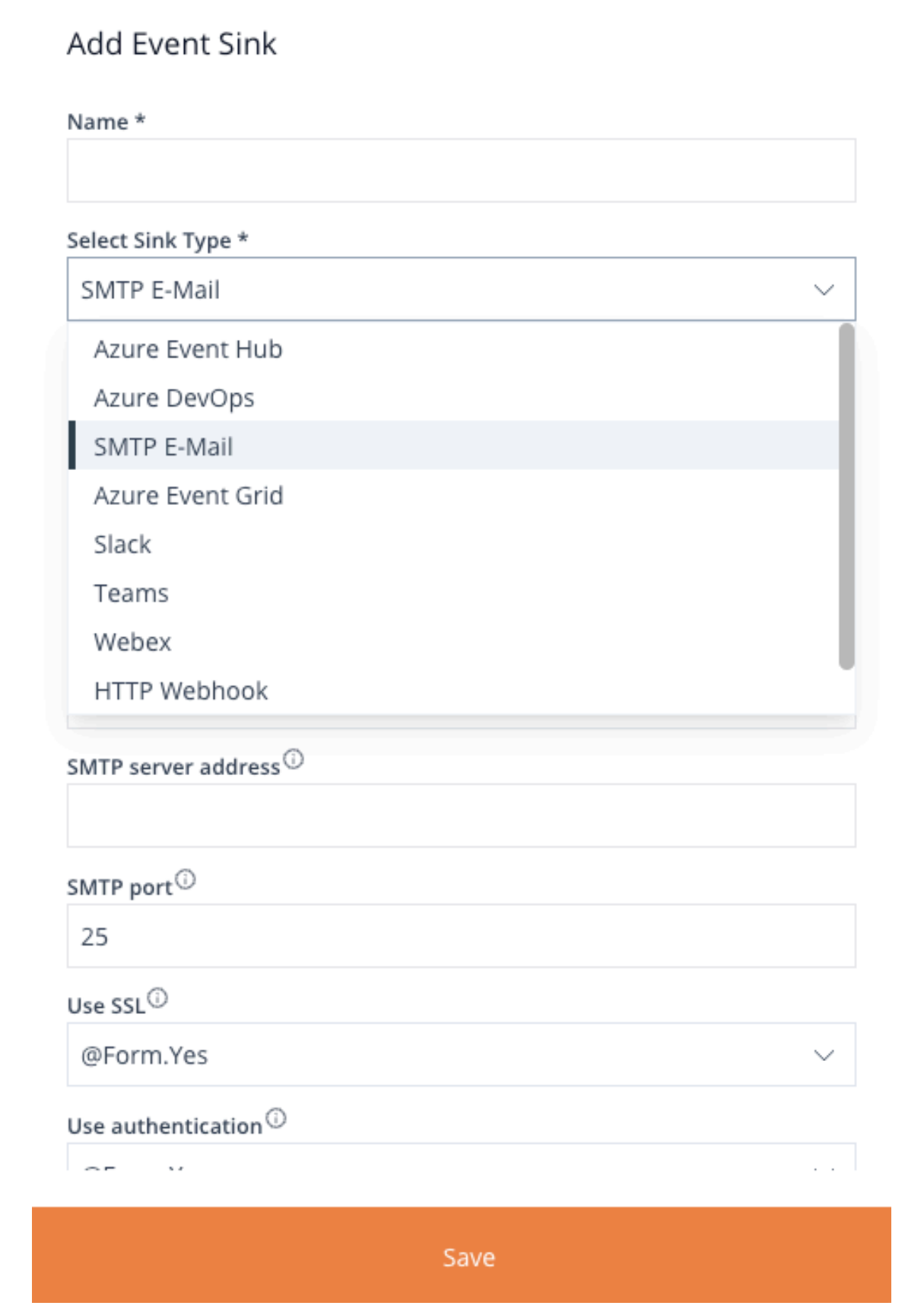


Figure 22: Add Event Sink

The form is comprised of two primary fields, Name and Sink Type, as well as a collection of inputs determined by the chosen Sink Type, as shown in the example above.

3.10.1.1.1 Azure Event Hub The Azure Event Hub event sink has three fields that require user input:

- **Connection String**: The unique key and endpoint URL used to authenticate and establish a secure connection to the Azure Event Hub namespace.
- **Event Hub Name**: The specific identifier of the Event Hub within the namespace where events are sent and received.
- **Consumer Group Name**: The label for a logical group of consumers that allows multiple applications to independently read from the same Event Hub without interfering with each other.

A guide on how to create an Event Hub in Azure Portal is available here.

3.10.1.1.2 Azure DevOps The Azure DevOps event sink has four fields that require user input:

- **Azure DevOps Organization Name**: The unique name of the Azure DevOps organization where projects and repositories are managed.
- **Project Name**: The specific Project within the Azure DevOps organization that contains the relevant work items and pipelines.
- **Personal Access Token**: A secure token used to authenticate and authorize access to Azure DevOps resources and actions within the organization. A guide on how to create a PAT in Azure DevOps is available here.
- **Work Item Type**: The classification of a work item (such as bug, task, or user story) to define the kind of work being tracked within the Project.

3.10.1.1.3 SMTP Email The SMTP Email event sink has nine fields that require user input:

- **Recipient E-Mail**: The email address where the notification will be sent.
- **SMTP Server Config**: The overall configuration settings that define how the SMTP server will be used for sending emails.
- **Sender E-Mail**: The email address that will appear as the sender of the notification.
- **SMTP Server Address**: The specific address of the SMTP server used to send emails.
- **SMTP Port**: The network port through which the SMTP server connects.
- **Encryption Method**: The protocol (e.g., SSL/TLS) used to secure the connection to the SMTP server.
- **Use Authentication**: Specifies if authentication is required to access the SMTP server.
- **SMTP User Name**: The username for authenticating with the SMTP server.

- **SMTP Password**: The password associated with the SMTP username for secure access to the SMTP server.

3.10.1.1.4 Azure Event Grid The Azure Event Grid event sink has four fields that require user input:

- **Access Key**: The security key used to authenticate and authorize access to the Azure Event Grid.
- **Azure Event Grid Endpoint**: The specific URL where events are sent to trigger notifications within Azure Event Grid.
- **Event Subject**: A description or identifier for the event, typically indicating the source or context of the notification. This field has 'heddaio/alert' set as default, but is allowed to be adjusted.
- **Event Type**: A classification that specifies the nature of the event, such as "new file uploaded" or "status update". This field has 'HEDDAIO.Alert' set as default, but is allowed to be adjusted.
- **Event Data Version**: The version number that defines the schema or structure of the data included with the event. This field has '1.0' set as default, but is allowed to be adjusted.

3.10.1.1.5 Slack The Slack event sink has only one field that requires user input:

- **Incoming Webhook URL**: The unique URL endpoint that receives and processes incoming HTTP requests to trigger notifications or actions in a specific application or service, in this case, it allows sending messages directly to a specified Slack channel.

A guide on how to set it up on Slack's side is available here.

3.10.1.1.6 Teams The Teams event sink form has only one field that requires user input:

- **Incoming Webhook URL**: The unique URL endpoint that receives and processes incoming HTTP requests to trigger notifications or actions in a specific application or service, in this case, it allows sending messages directly into a Microsoft Teams channel via the webhook integration.

The Webhook can be obtained from Teams and depending on which version of Teams you are currently using, a guide on how to create an Incoming Webhook is available here.

Unfortunately, when using the New Teams, the default workflow that is automatically created, will not work with HEDDA.IO event trigger payload out of the box. Adjustments are required.

In order to make the adjustments to the workflow, go to Teams and and click on the `...` in the left menu, then look for "Workflows" then click on it. The Workflows app should open and a list of available workflows should be presented. Locate the one you created for this purpose and click on it to open.

On the page for the selected Workflow, locate and click the 'Edit' button. Once inside the Editing canvas, remove all except the first action called 'When a Teams webhook request is received'. Once

done, add a new step with an action of type 'Parse JSON' as a followup to the inital action, and then add a 'Post message in a chat or channel' type of action.

Parse JSON (action)

In the Content input, add Dynamic content 'Body' from the initial action. In the following Schema in the input below:

```
{
  "properties": {
    "Title": {
      "type": "string"
    },
    "Text": {
      "type": "string"
    }
  }
}
```

Post message in a chat or channel (action)

There are a number of inputs, in this action, which should not require explanation.

In the 'Message' input, add the 'Title' and the 'Text' dynamic content, from the 'Parse JSON' action.

You are good to go. Now, every time the event is triggered in HEDDA.IO a message with the specified Title and Text will be posted on the selected Teams channel.

Please note, that this is the simplest way to set the workflow up. The Workflows in Teams are based on Power Automate, and a person with more expertise can set the workflow up in very creative ways.

3.10.1.1.7 Twillio

The Twillio event sink has five fields that are require user input:

- **Mode**: The operational mode for sending the notification, typically specifying whether it's a test or live environment.
- **Account SID**: The unique identifier for your Twilio account used to authenticate and send messages.
- **Auth Token**: The secure token paired with the Account SID for authenticating API requests to Twilio.
- **From**: The Twilio phone number or sender ID from which the notification message will be sent.
- **To**: The recipient's phone number where the notification message will be delivered.

3.10.1.1.8 Webex

The Webex event sink has two fields that require user input:

- **Bot Access Token**: A unique authentication token that allows the bot to interact with the Webex API and send notifications on behalf of the user.

- **Room Name**: The specific name of the Webex room (or space) where the bot will send the notifications or messages.

3.10.1.1.9 HTTP Webhook The HTTP Webhook has three fields that require user input:

- **Webhook URL**: The endpoint URL where the HTTP request should be sent to trigger the webhook notification.
- **HTTP Method**: The type of HTTP request to be used, such as POST, PUT, PATCH, or GET, defining how data is sent or retrieved from the webhook URL.
- **HTTP Headers**: The key-value pairs that provide additional context for the request, such as content type or authorization, sent along with the HTTP request (can also use @ for dynamic values).

Click the Save button at the bottom of the form to save the Event Sink and close the drawer.

3.10.2 Events Overview

The middle section of the page will list all created Events.

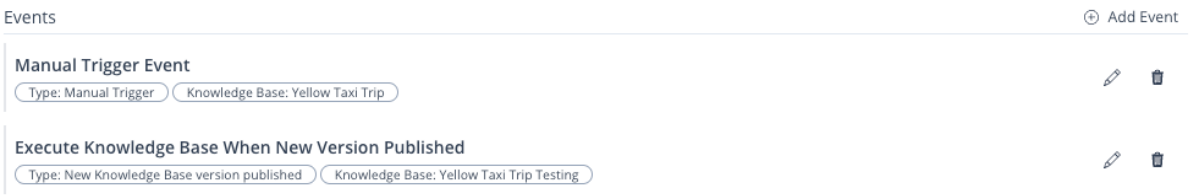


Figure 23: Events Overview

Each item in the list is represented by the Name, as well as the inputs that were configured. On the far-right side, each item has an edit and a delete button.

3.10.2.1 Add/Edit/Delete Events After setting up the Event Sink, you are ready to create an Events on top of it.

To initiate the Events creation process, you can locate the Events creation button situated at the upper-right corner of the overview, just slightly above the roster of Events.

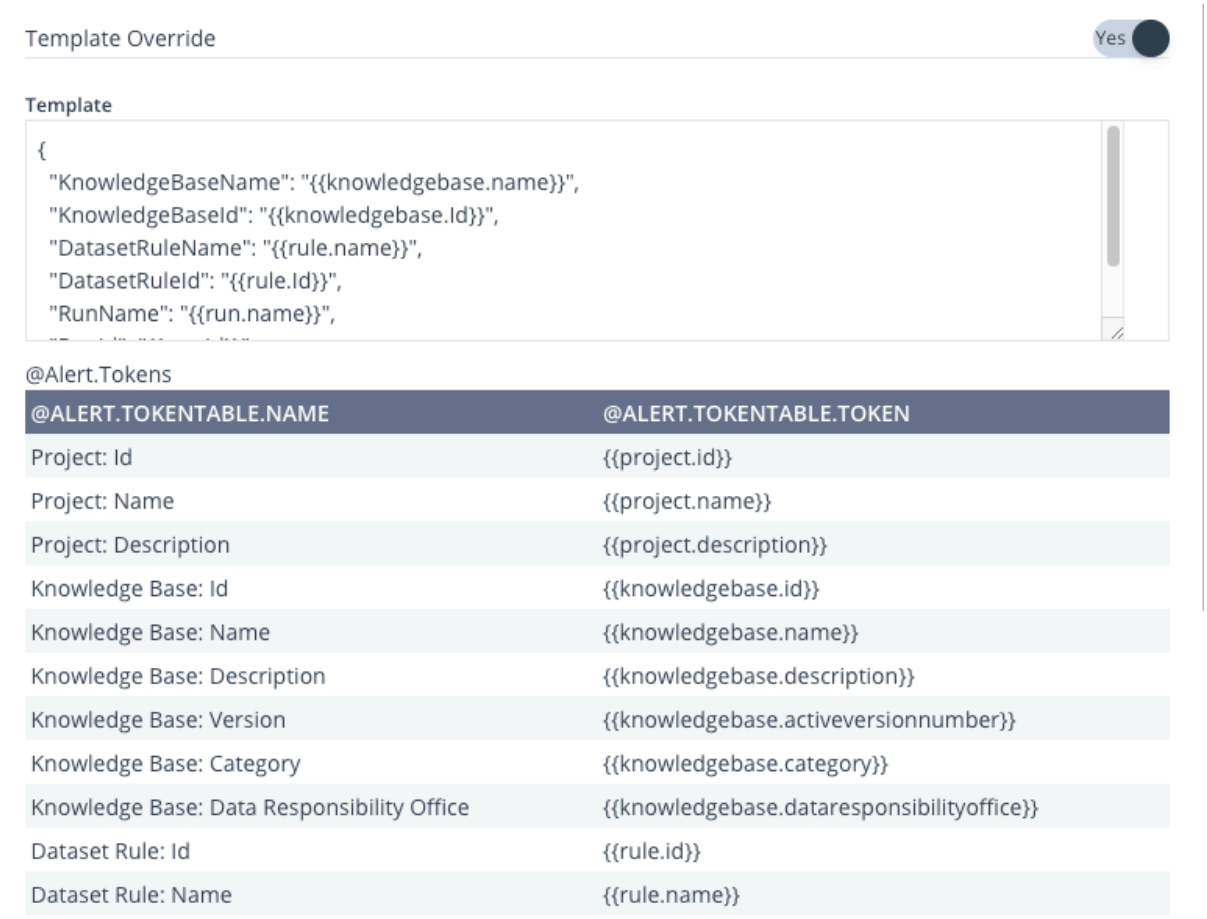
Upon clicking the Add Event button, a drawer will unfold presenting a form designed for crafting an Events. This form encompasses the following input fields:

- Name
- Description

- Events Type
- An array of inputs corresponding to the chosen Events category.
- Template

The template stands as the core content of the message, which includes essential metadata from the Knowledge Base. This content will be passed to the sink as soon as the Events is triggered.

Every kind of Events comes with a pre-established template message content. Nevertheless, you have the liberty to personalize the message content by switching to `Template Override`. This action grants you the ability to modify the message using tokens. For your convenience, a guide Mapping out these tokens is supplied right beneath the Template input box.



@ALERT.TOKENTABLE.NAME	@ALERT.TOKENTABLE.TOKEN
Project: Id	{{project.id}}
Project: Name	{{project.name}}
Project: Description	{{project.description}}
Knowledge Base: Id	{{knowledgebase.id}}
Knowledge Base: Name	{{knowledgebase.name}}
Knowledge Base: Description	{{knowledgebase.description}}
Knowledge Base: Version	{{knowledgebase.activeversionnumber}}
Knowledge Base: Category	{{knowledgebase.category}}
Knowledge Base: Data Responsibility Office	{{knowledgebase.dataresponsibilityoffice}}
Dataset Rule: Id	{{rule.id}}
Dataset Rule: Name	{{rule.name}}

Figure 24: Template Override

> Based on Events Type, the available tokens might differ.

Click the `Save` button at the bottom of the form to save the Events.

3.10.3 Event Sink Info Panel

On the right side of the page, as always, is the Info Panel section, which includes metadata about the selected Event Sink, and an `Edit` button in the top-right corner to edit or delete the Event Sink.

3.10.3.1 Edit/Delete Event Sink Clicking the `Edit` button in the top-right corner of the Info Panel, will open up the `Edit Sink` drawer, allowing you to edit the pre-existing Sink, by adjusting the inputs and then clicking the `Save` button at the bottom of the form. Click the `Delete` button to the right, if you wish to delete the Sink.

3.11 Event Log Page

The Events Log page will display a record of all activities triggered within this Project's Events.

This page is composed of a table located at the lower section and a filtering segment positioned at the upper section.

The table includes the subsequent columns:

- Events Type
- Sink
- Knowledge Base
- Run
- Timestamp
- Sent

Within the filtering segment, there are three buttons:

- **Refresh**: This button refreshes the rows within the table.
- **Edit Log Filters**: This button enables the filtering of rows within the table.
- **Reset Filters**: This button resets any applied filters.

By clicking the `Edit Log Filters` button, a drawer will emerge, presenting a form that allows you to apply filters to the Events table in the log.

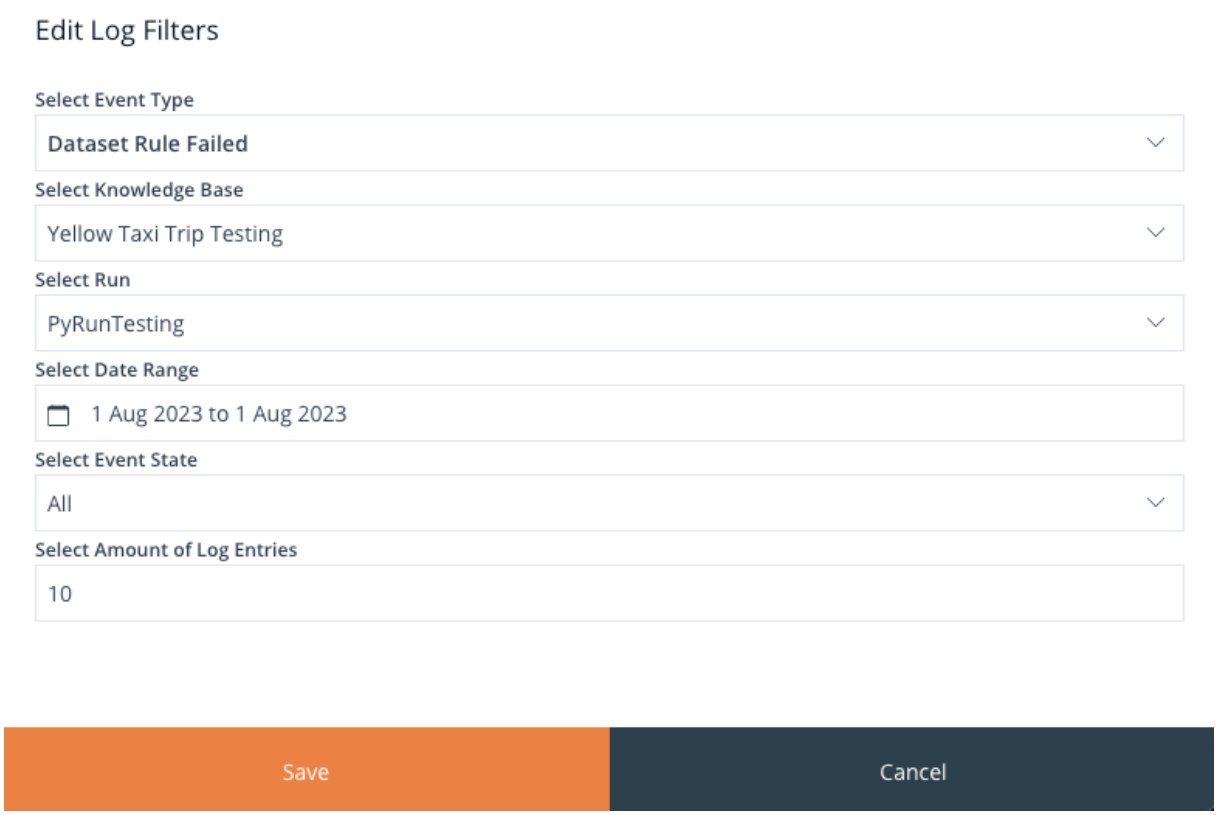


Figure 25: Filter Events logs

Clicking Save will apply the filter to the table.

3.12 Knowledge Base

This part of documentation will get you familiar with the Knowledge Base page as well as guide you through the process of creating a Knowledge Base.

3.12.1 Add Knowledge Base

Selecting the Add Knowledge Base button positioned right above the Knowledge Bases Overview on the Project Dashboard will trigger the opening of the Add Knowledge Base drawer.

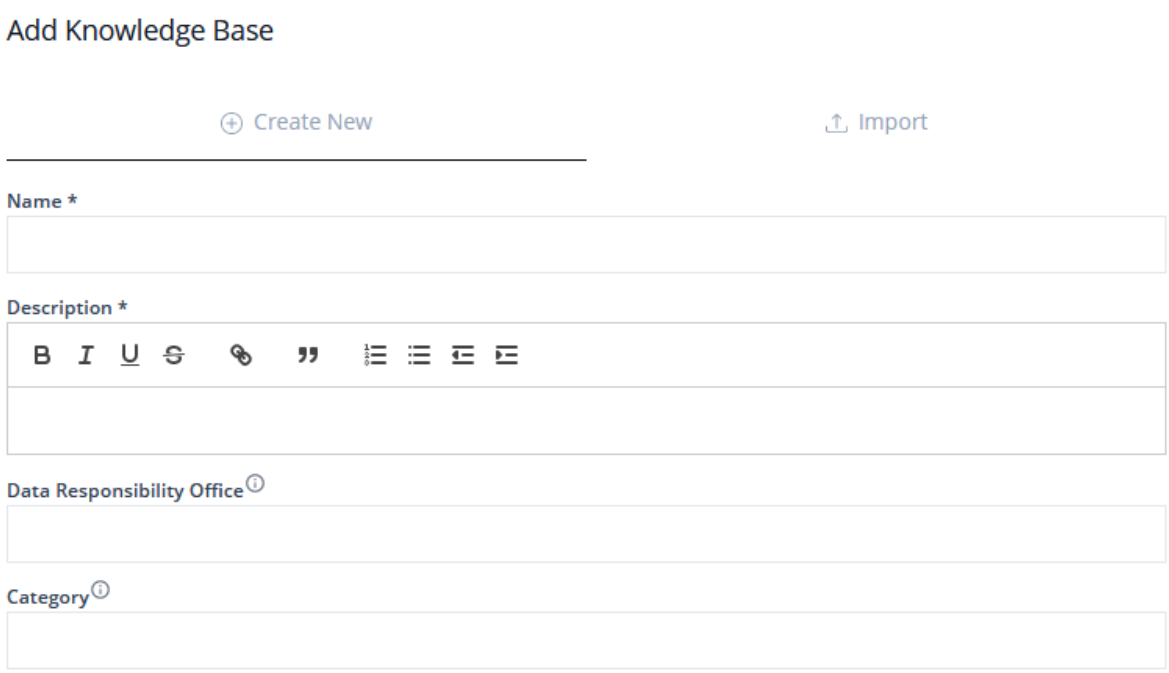


Figure 26: Add Knowledge Base

This particular drawer has two tabs, each offering two forms representing two distinct methods for incorporating a Knowledge Base:

- Create New
- Import

3.12.1.1 Create New In the 'Create New' form, you can create a Knowledge Base from scratch. The only mandatory field is the Name and the Description, while additional information, like Data Responsibility Office and Category, are optional. This flexibility allows you to add extra context and categorization either as needed or at a later point.

3.12.1.2 Import The import procedure unfolds in three distinct stages: `Select Provider`, `Select Knowledge Bases`, and `Finalize`. This comprehensive process empowers you to seamlessly import Knowledge Bases from a variety of distinct sources.

3.12.1.2.1 Select Provider

- Microsoft Fabric Lakehouse
- Azure Data Lake Storage
- Azure Blob Storage

- Databricks
- Local File
- Microsoft SQL Server
- Microsoft OneLake

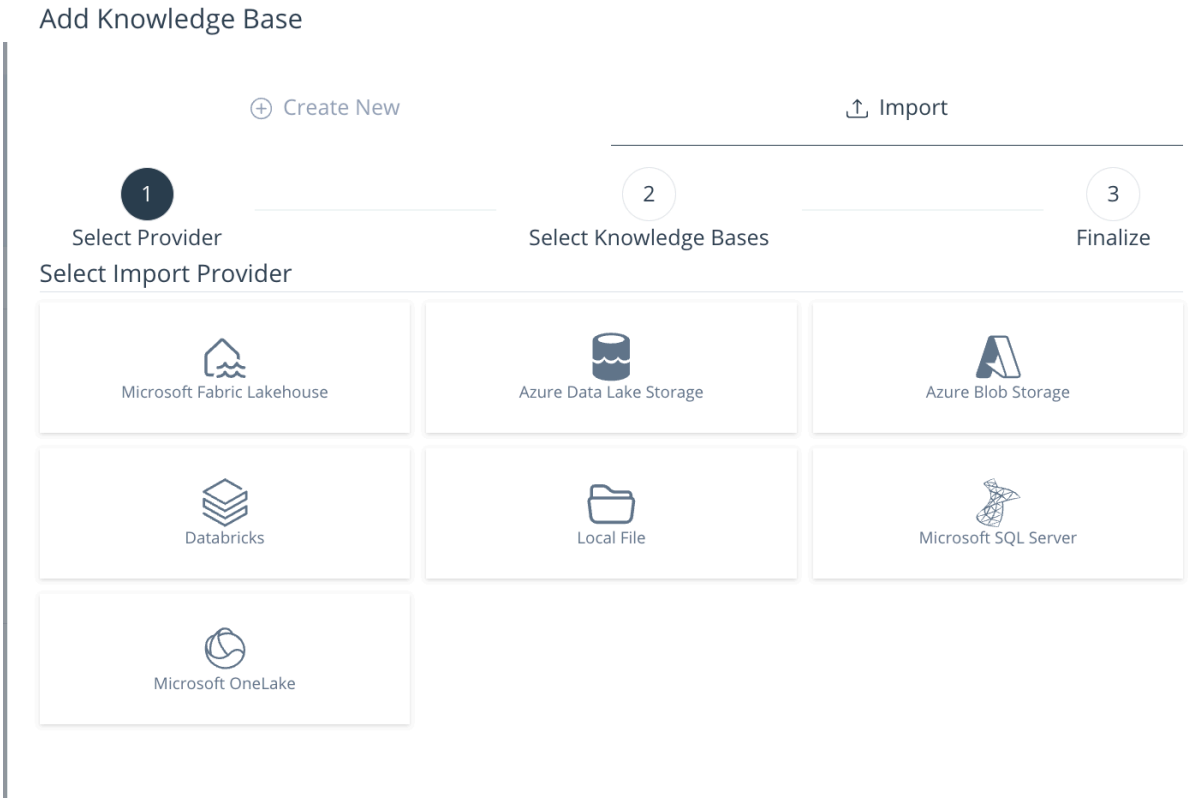


Figure 27: Knowledge Base Import Source List

Local File

Load a Knowledge Base from a local .parquet, .XLSX, or a .heddaexport file.

To use a local data file or a hedda export, simply drag the file in the drop box or click on the drop box to browse the local files.

3.12.1.2.2 Select Knowledge Base No matter which provider you choose to use, when connection is successfull, you will be allowed to select a data file, table, or schema you want to use as a Knowledge Base and select which columns to import as Domains. Click Next, once you have selected the desired data file, table, or schema.

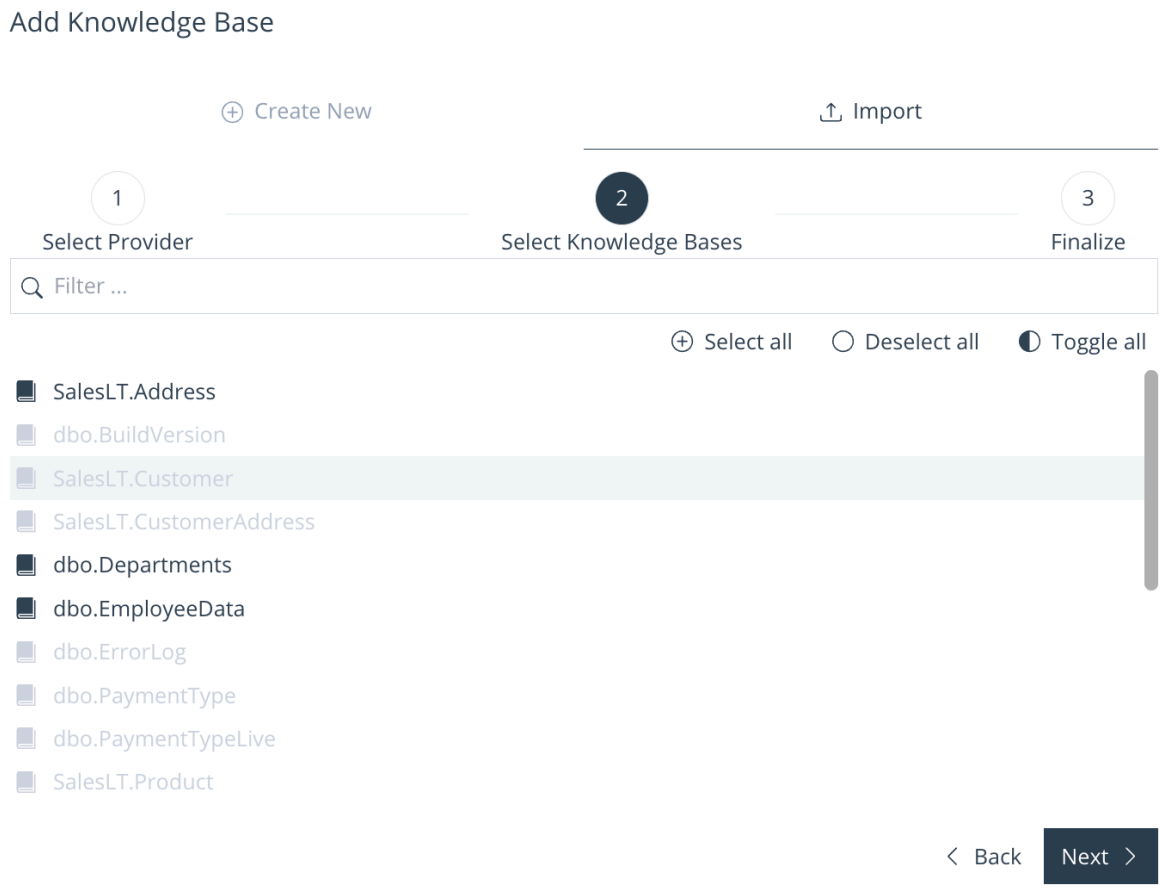


Figure 28: Import Knowledge Base - Select Table/Data File

3.12.1.2.3 Finalize This step will reveal a number of imputs that will allow you to configure the Knowledge Base about to be imported.

- **Name**: a unique name for the knowlege base in the context of the Project. When imported from an external source, the name will automatically be set to the name of the data file, table or schema that is being imported.
- **Description**: the description of the Knowledge Base.
- **Data Responsability Office**: the name of the responsible office or person
- or the contact person.
- **Category**: helps with the classification of the Knowledge Base.
- **Domain Selection**: allows you to select which of the columns in the data file, table, or schema should be imported as Domains for the Knowledge Base.

By default, all available columns will be pre-selected for import as Domains into the Knowledge Base. Should you wish to exclude any, a simple click on the respective column/Domain will un-select it, effectively excluding it from the import.

The `Toggle Domains` toggle is to either mark previously unchecked items as checked or vice versa.

You can always click the `Back` button to return back to the data file, table, or schema selection menu.

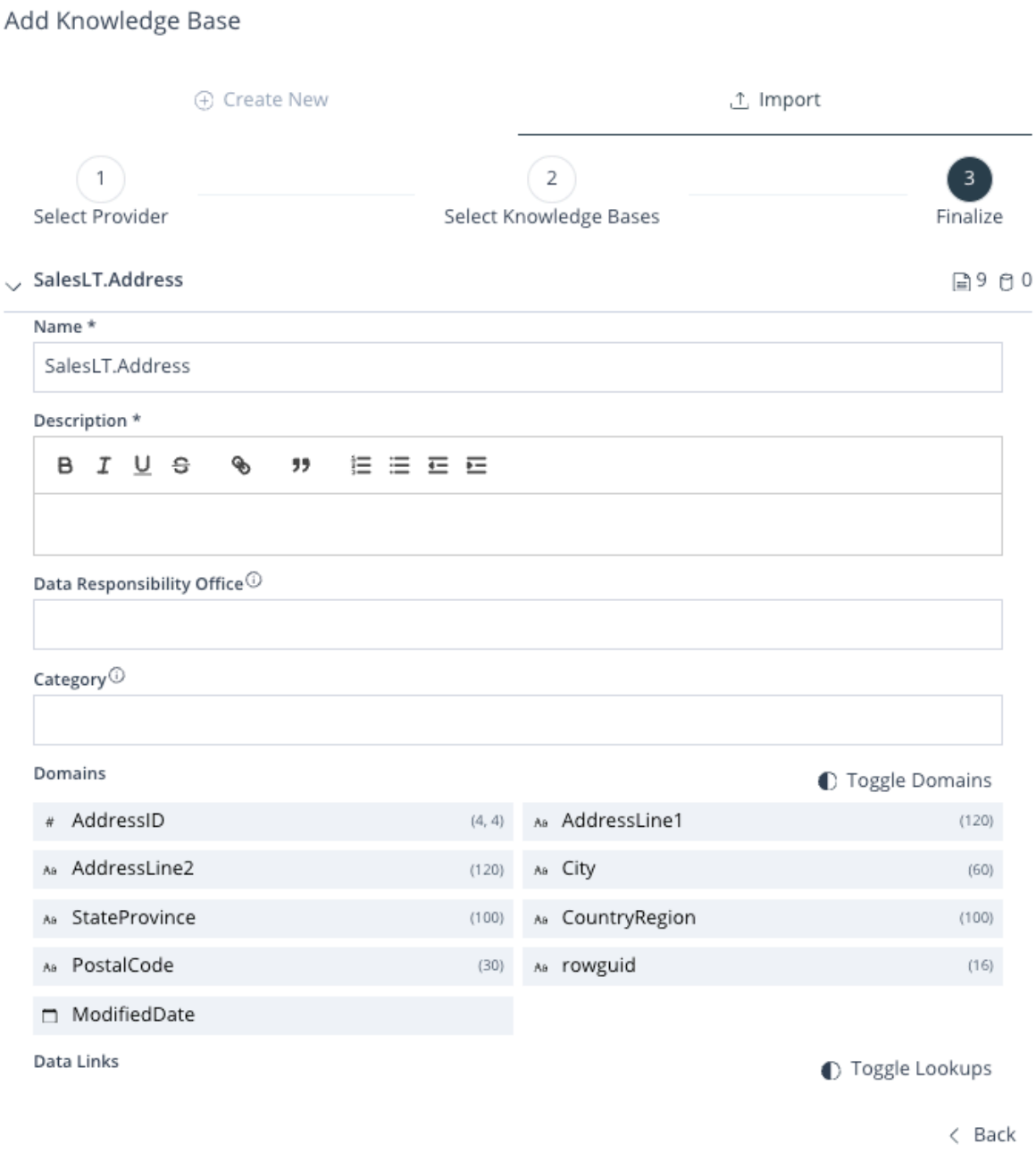


Figure 29: Finalize Knowledge Base Import

After finishing this step using any of the providers, move ahead and click the `Save` button positioned

at the lower part of the menu. This action will finalize the import process. Following that, the list will display your imported Knowledge Base with the name you assigned to it.

3.13 Knowledge Base Page

3.13.1 Knowledge Base Navigation Bar

Upon accessing a Knowledge Base, your initial focal point will be the Navigation Bar situated at the top. This bar provides access to various functionalities within the Knowledge Base.

This bar is equipped with 8 distinct tabs, each providing access to different segments of the Knowledge Base:

- Dashboard
- Domains
- Rulebooks
- Dataset Rules
- Data Links
- Runs
- Mappings
- Tags

Positioned to the right of the Navigation Bar are two buttons: `Export to File` and `Edit Version`.

Figure 30: Knowledge Base navigation

We'll take a closer look at each of these tabs to see what's inside them and what they offer.

Later, we'll also talk about how to export a Knowledge Base and how the Edit Version mode works.

3.13.2 Knowledge Base Export/Import

3.13.2.1 Exporting Within HEDDA.IO, you have the capability to extract Knowledge Bases for the purpose of importing them into different Projects or even into HEDDA.IO instances in alternate environments.

This process is straightforward and involves clicking the `Export to File` button located at the upper-right corner of the Knowledge Base page. Upon clicking this button, the `Export Knowledge Base` drawer will be unveiled, offering you the option to designate which components of the Knowledge Base you intend to export.

It makes sense to consider exporting the elements that pose the greatest setup complexity. This primarily encompasses the Domains, along with their corresponding Members, the Rulebooks, and the Dataset Rules.

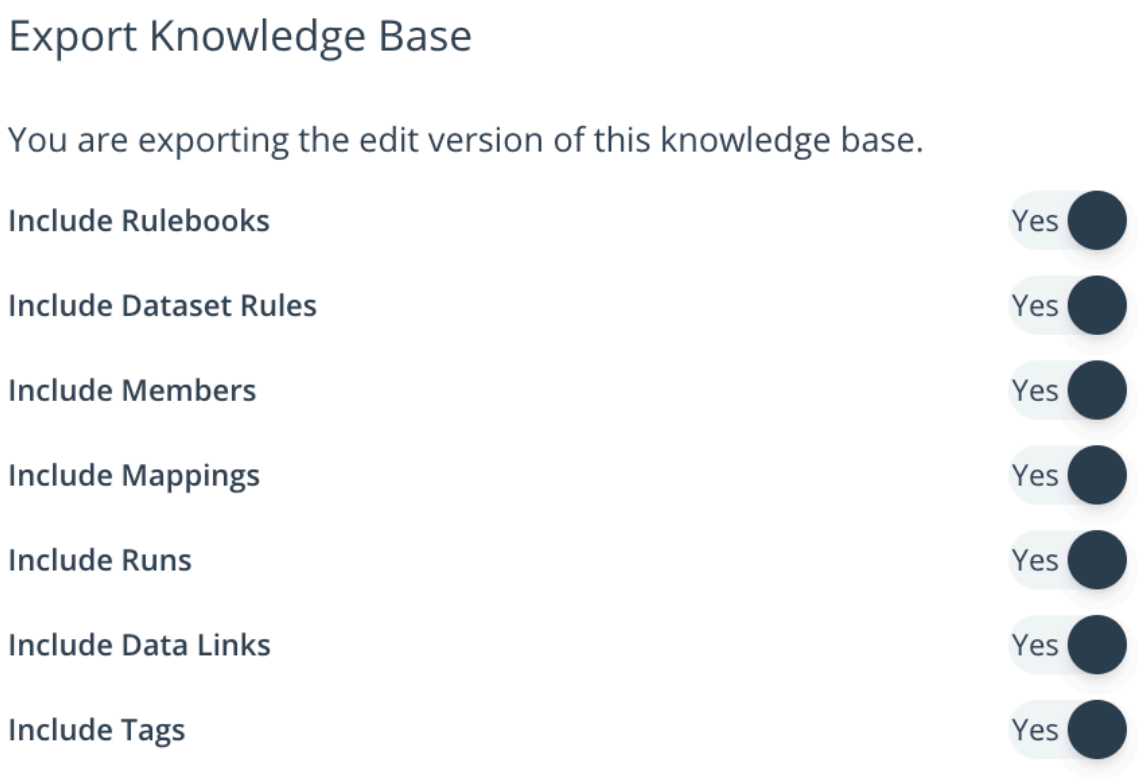


Figure 31: Export Knowledge Base

The Domains will be automatically included in the export. You are then provided with the choice to determine which among the remaining three components you wish to include in the export.

When you are prepared to proceed with exporting the Knowledge Base, simply click the `Export` button located at the base of the drawer.

3.13.2.2 Importing The import of the Knowledge Base export occurs upon the creation of the Knowledge Base. To initiate this process, you simply switch to the `Import` tab within the `Add Knowledge Base` drawer and utilize the `Local File` button.

Read the Local File Import for more details.

3.13.3 Knowledge Base Edit Version

In HEDDA.IO, the initial step to utilize a Knowledge Base involves publishing it. This functionality serves a dual purpose, allowing the establishment of a version history as well. The primary page of the Knowledge Base is perpetually set in Read Only mode. This signifies that you can exclusively view or read values, without the capability to alter, erase existing entities, or introduce new ones.

To achieve any of the three actions, you must switch to the `Edit Version` mode, accessible by selecting `Edit Version` located at the upper-right corner.

When using Git, you have the possibiliy to use several edit versions at the same time and the form `Knowledge Base Version` pops up where an existing version can be selected in the tab `Select version` and opened by clicking `Ok`. To create a new Edit Version, select the tab `Add Version`, enter a name and confirm by clicking `Ok`.

Details on how to add, modify, or remove entities are outlined in the relevant documentation corresponding to each section within the Navigation Bar.

3.13.3.1 Knowledge Base Version Publishing The `Edit Version` mode, as explained before, allows you to add, edit, and delete entities. However, these changes will not apply until they are published, which is done by pressing the `Publish` button when all the needed adjustments are complete.

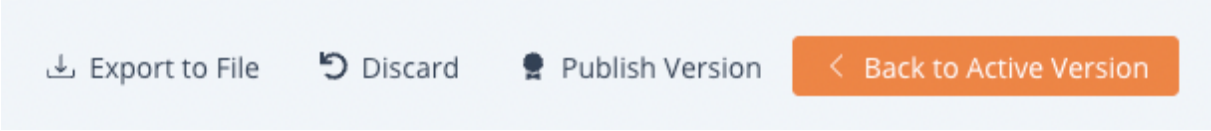


Figure 32: Knowledge Base Publish

Once you click the `Publish Version` button, a small pop-up panel in the middle of the screen will display all the changes that will be applied to the Knowledge Base with the new version, allowing you to complete the publishing process or cancel it. When using a Git persisted Project, the current Edit Version will be removed after Publishing successfully. If a conflict is detected between changes in the live Version coming from other Edit Versions and the current Edit versions, the Publishing process is stopped until the merge conflict is resolved. This can be done under `Git` on the Project Page which will show a “Has Conflict” indicator if conflicts are detected.

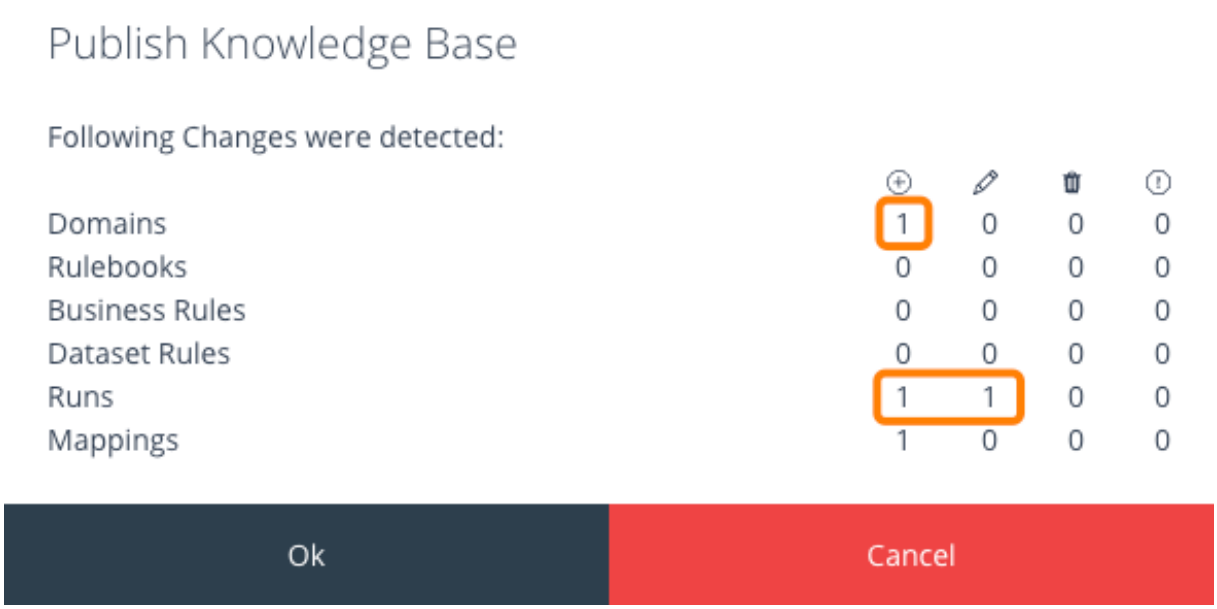


Figure 33: Knowledge Base Publish validation

You can always go back to the active version by clicking `Back to Active Version`.

You also have the option to discard all changes in the `Edit Mode` by clicking the `Discard` button next to the `Publish Version` button. A Git persisted Knowledge Base Edit Version can also be deleted without publishing using `Discard`.

3.13.3.1.1 Resolving a Merge Conflict After a Merge Conflict is detected during publishing, a message will open announcing a conflict was detected. Confirming the prompt will open a drawer to resolve this issue. This drawer can also be opened under Git on the Project Page by clicking `Fix` next to `Has Conflicts`.

Conflicting Items can be selected from the List shown in the drawer. On selecting a conflict item by clicking `Show Differences`, a table opens showing the differences between the current version on the left side and the active version on the right side. Selecting a version will use all changes of a side and all changes from the other side are discarded.

`Resolve Conflict` will then complete the merge and open the merged Knowledge Base page. `Cancel Conflict Resolution` will stop the publishing process so changes can be made for publishing at a later time.

3.13.3.2 Synchronize An Edit Version can be changed by two Users at the same time. This will cause a conflict within an Edit Version. In order to resolve the issue, a `Synchronize` Button will

appear. This button will overwrite your current changes within the Edit version with changes of the remote Edit version.

3.13.3.3 Edit/Delete Knowledge Base

In order to edit or delete the Knowledge Base, you need switch to the Edit Version mode, which can be done by clicking the `Edit Version` button at the top-right corner of the page.

Once in the Edit Version mode, the `Edit` button is now present in the top-right corner of the Knowledge Base Info Panel.

Clicking the button will open the Edit Knowledge Base drawer, allowing you to edit the Name, Description, Data Responsibility Office, and the Category.

Click the `Save` button to save the changes and close the drawer.

3.13.4 Knowledge Base Dashboard

This is the page that you are directed to whenever you access a Knowledge Base.

The dashboard page is split into two sections:

- Knowledge Base Overview
- Knowledge Base Info Panel

3.13.4.1 Knowledge Base Overview

On the Overview side you will have statistics about recent Runs, as well as results per Business Rule, results per Domain and results per Dimension.

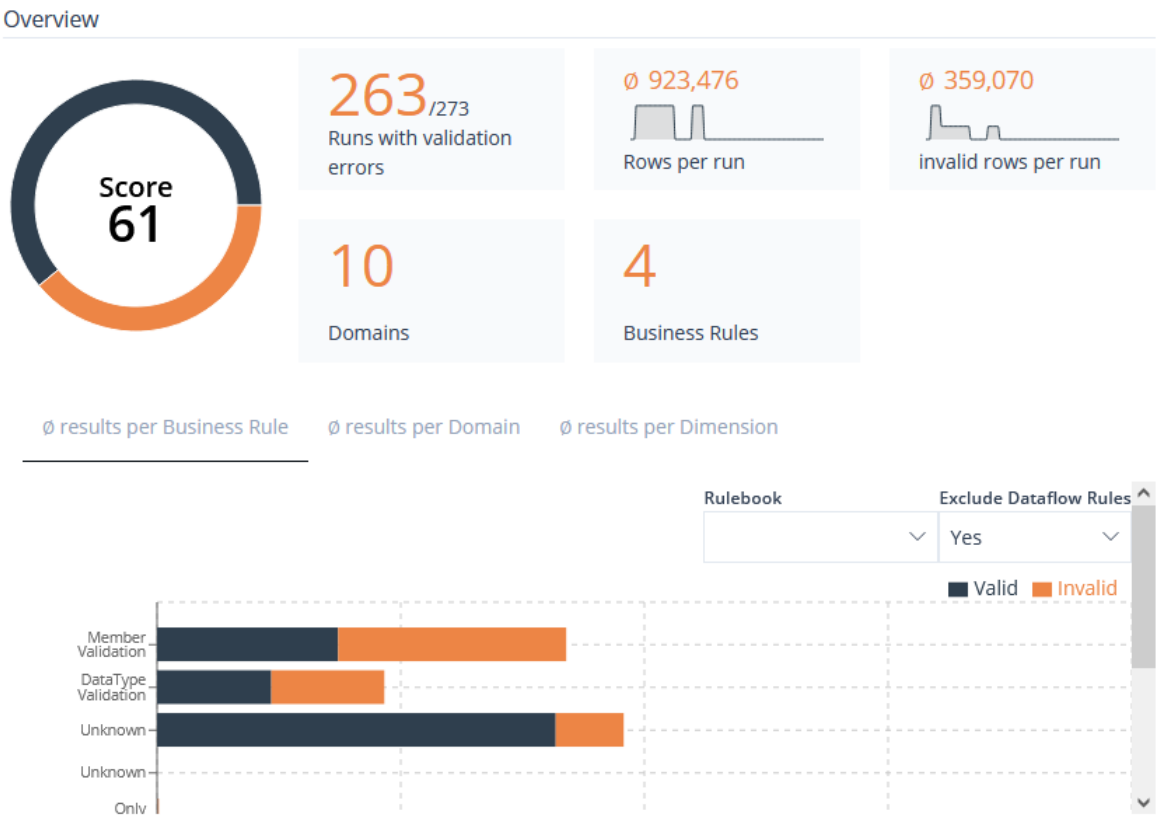


Figure 34: Knowledge Base Overview

On the top-right side of the overview, we have a button called `Filter Statistics`.

By clicking this button, you can access the `Filter statistics by Date Range and Version` drawer.

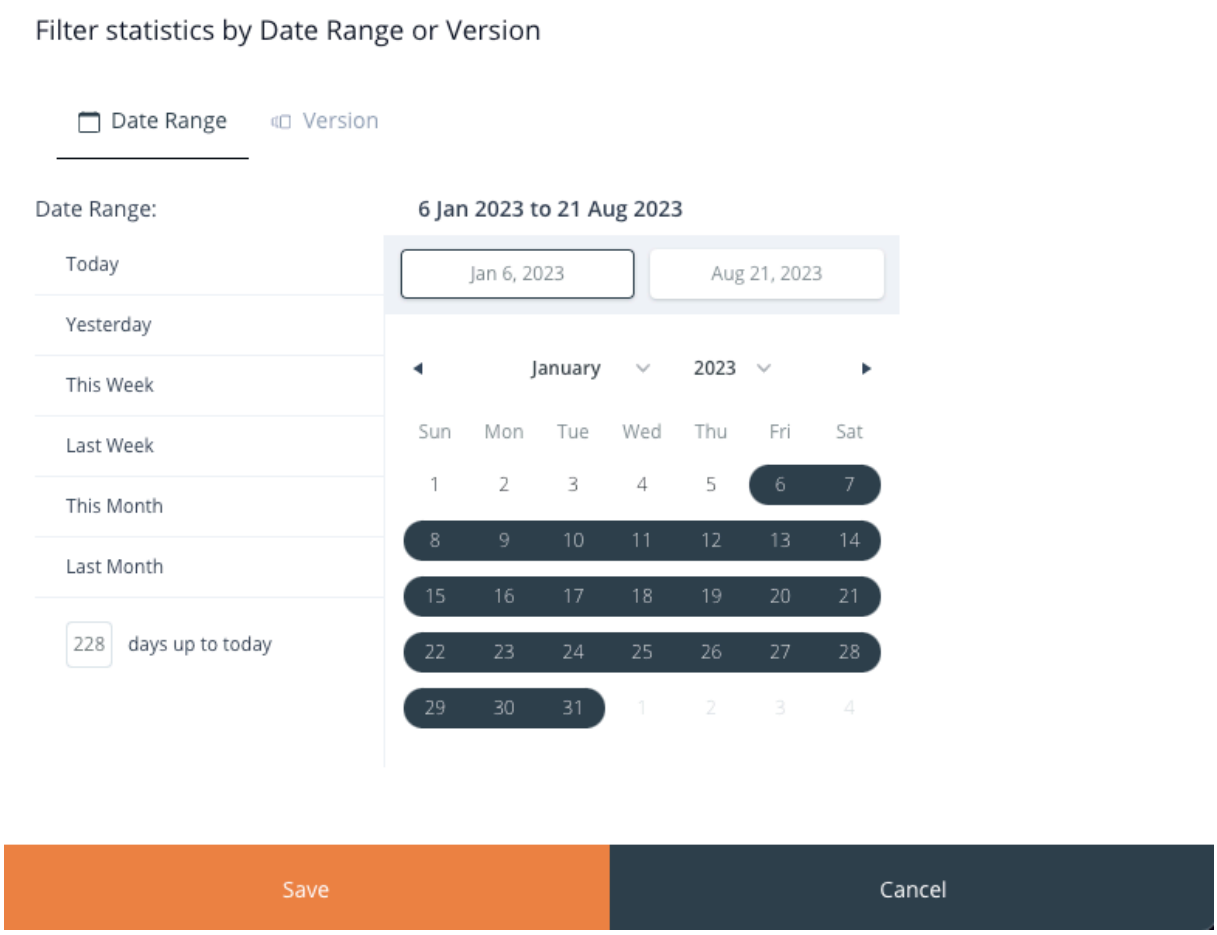


Figure 35: Filter Knowledge Base Statistics

Within this drawer, on the top side, you can see two tabs:

- Date Range
- Version

When you click on the `Date Range` tab, on the left-hand side, you have the option to choose from a range of predefined date periods, including:

- Today
- Yesterday
- This Week
- Last Week
- This Month
- Last Month

Directly beneath, you can specify a certain number of days leading up to the current date.

Alternatively, you also have the flexibility to define a custom date range using the date picker tool located on the right side.

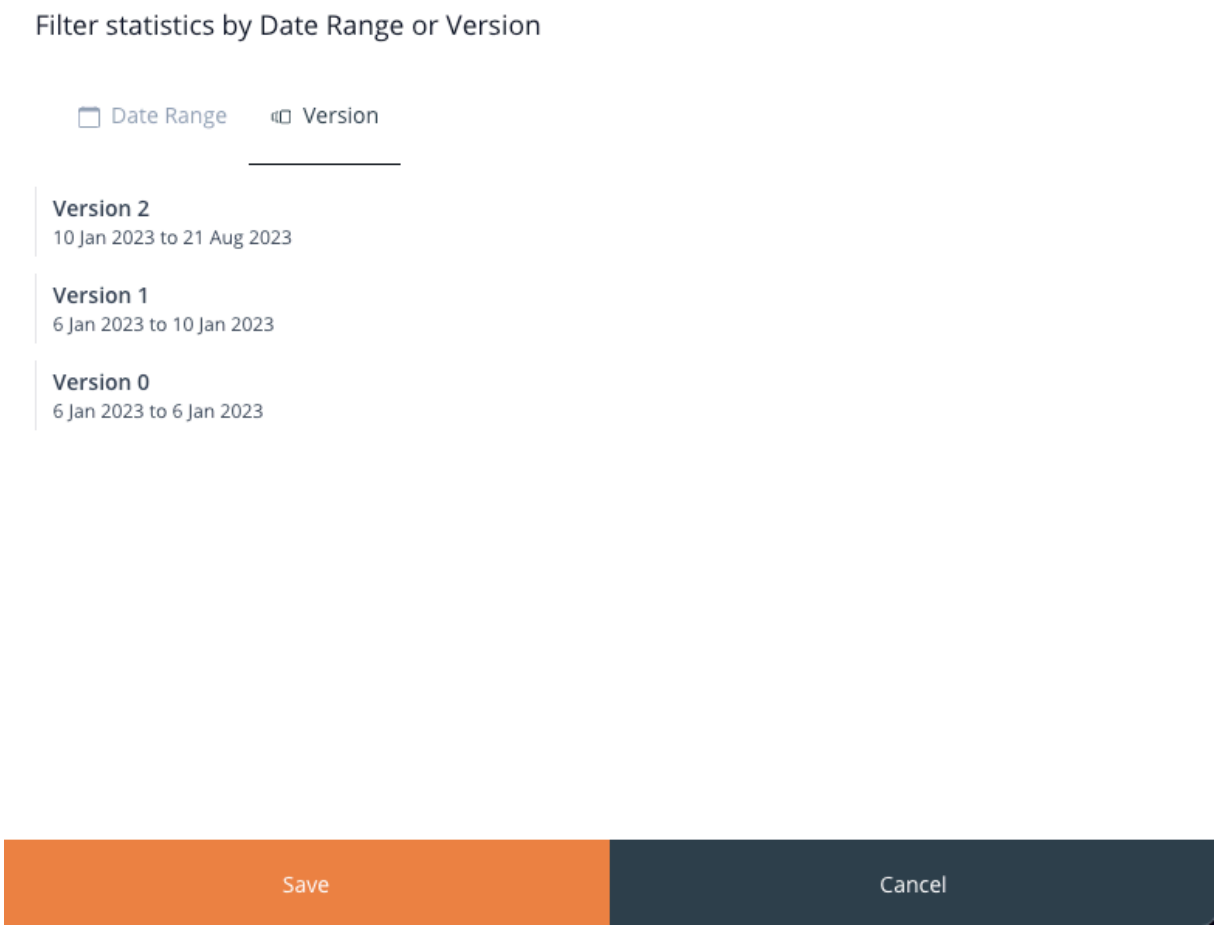


Figure 36: Filter Knowledge Base Statistics by Version

Switching to the Version tab, you have a list of previously published Knowledge Base versions. This will allow you to filter the overview based on all previously published Knowledge Bases and their respective active times.

3.13.5 Knowledge Base Audit Portal Page

Audit Portal: Knowledge Base Change History Feature

The Audit Portal provides a comprehensive overview of all modifications made to the Knowledge Base and its components since its creation. To access the Audit Portal, users must enter the Knowledge Base edit mode. Then, in the Dashboard tab, locate the 'Audit Portal' button on the bottom of the left-side panel.

Here, users can view detailed information on each change, including:

- **Name**: The name of the Knowledge Base component that was added, changed, or removed.
- **Type**: The specific part of the Knowledge Base that was modified (e.g., Domains, Rulebooks).
- **Modified On**: The exact date and time the change occurred.
- **Modified By**: The individual who made the change.
- **Knowledge Base Version**: The version of the Knowledge Base at the time of the modification.
- **State**: Whether the component was added, removed, or changed.
- **Publish Status**: Whether the modified component was published or remains in draft.

The feature also includes filters located at the top, allowing users to:

- Filter by specific **Knowledge Base Component Types** and/or **State**.
- Select a **Date Range**, defaulting to the last 7 days for quick access to recent changes.

Additionally, users are allowed to search for an item in the table by 'Name', and select the number of rows they want to be displayed on every page, with the minimum being 25 and the maximum 100.

The Audit Portal ensures full traceability and easy navigation through the Knowledge Base history.

3.13.6 Knowledge Base Info Panel

As per usual, the Info Panel is located on the right side of the page. It includes:

- Preview button
- The Description of the Knowledge Base
- The Data Responsibility Office
- The Category

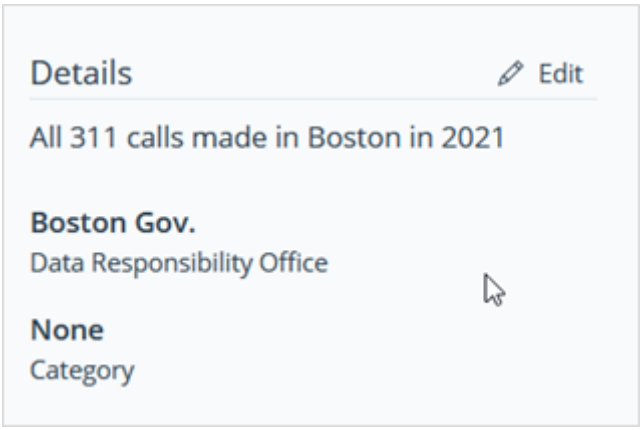


Figure 37: Knowledge Base details

3.13.7 Preview Page

The Knowledge Base Dashboard provides access to the Preview button. In the top-right corner of the Info Panel on the right, you'll locate the `Preview` button.

3.13.7.1 Start Preview Upon clicking the button, a drawer will be revealed, prompting you to select a Run and a Data Source.

The Data Source section has two tabs:

- Data Link - allows the execution of a Preview using a Data Link, which points to a data file, table, or a schema, depending on the connection it is using.
- Execution - allow the exeution of a Preview on result data from a previous execution performed on the current or different Knowledge Base.

3.13.7.1.1 Preview from Data Link Has only one dropdown field, that allows you to select the Data Link that you want to use as data source for the Preview.

3.13.7.1.2 Preview from Execution Has the following dropdown fields:

- **Select Knowledge Base**: the Knowledge Base in which the execution was performed.
- **Use Version**: the version of the selected Knowledge Base in which the execution was performed.
- **Select Run**: the Run that was used to perform the execution.
- **Select Execution**: a list of executions belonging to the selected Run. Comes with a Date range filter to the right.
- **Select Filter**: comes with predefined filter buttons at the bottom:
  - **All**: will use all the records.
  - **Valid**: will use only the valid records.
  - **Invalid**: will use only the invalid records.

After selecting these, click `Start` to initiate the Preview.

This action triggers a HEDDA.IO execution on the chosen Data Link or Execution, followed by the opening of the Preview menu. Here, you can examine the data processed by HEDDA.IO.

3.13.7.2 View Data Within the Preview menu, you have the ability to navigate and filter the dataset using various useful filters. Additionally, you can view the statistics associated with the execution.

Figure 38: Knowledge Base Preview

The toolbar situated just above the data table contains a series of buttons, each serving a distinct purpose. Let's outline these from left to right.

- **Execution** - Will unveil the statistics panel, showcasing the statistical information linked to the execution carried out for the current Preview. More info on the Statistics panel can be found here.
- **`Valid`** - will display only the valid data rows.
- **`Invalid`**- will display only the Invalid data rows.
- **`Corrected`** - will display only the corrected data (Original differs from Actual) rows for currently selected Domains. > **Note**: The displayed data will not be reevaluated if Domain Selection is > changed while Corrected is active.
  - **`Check all Domains without Variables`** - will select all Domains except variables prior to making the Corrected check.
  - **`Check all Domains include Variables`** - will select all Domains prior to making the Corrected check.
- **`Domain Selection`** - Will reveal a compact menu, enabling you to choose the Domains you wish to have presented in the Preview data table. Click on a Domain entry to either select or deselect it, then press `Apply`. Alternatively, utilize the `All` button to choose all the Domains, or the `Toggle` button to select the currently unselected Domains or vice versa. Eg. Show only `Domain A` and `Domain B`.
- **Domain Filter** - Will open a compact menu that enables you to filter the Preview data table based on Domain value. Within this window, you can choose a Domain from the list and enter

a corresponding value. Click `Apply` once you are content with your selection. Eg. Show Rows where `Domain A` equals `Value`.

  - Booleans
    * Booleans filter with **`true`** or **`false`**
  - Strings
    * Strings have optional wildcards % and _. % matches any optional sequence of characters. _ matches a required single character. Eg. `Auto%` will match `Auto` and `Automat`. `Auto_` will match `Autos` but not `Auto` nor `Automat`.
  - Numbers
    * Numbers can be prefixed with >, <, >= and <= to filter respectively eg. `> 1.2` will filter for numbers greater 1.2. If the prefix is omitted the default operator = will be used.
  - Dates
    * The general Format is ISO8601. Eg. `2007-12-24T18:21`.
    * The format has the possibility to consist only of the date portion eg. `2024-01-01` this will then filter all the rows which have the same date portion. Dates containing the time portion `2024-01-01T00:00:00` will respect the time portion, so this would only find dates which actually represent midnight on January 1st, 2024. Milliseconds will be ignored.
    * Similar to Numbers dates respect the prefixes>, <, >= and <=. Eg. `< 2024-01-01` will filter for dates smaller than January 1st, 2024.
    * The token `$NOW` is able to filter for dates in comparison to the current timestamp. Eg. `>= $NOW`.
    * The token `$TODAY` is able to filter for dates in comparison to the current date. Eg. `<= $TODAY`.
  - Filtering for NULL Values is possible with the Token: `$NULL`.

- **Rule Filter** - Will open a compact menu enabling you to filter the Preview data table based on Rulebook and/or Business Rule state. You have the option to apply only the Rulebook filter by clicking the `Apply Rulebook` button or apply only the `Business Rule` filter by pressing the `Apply Business Rule` button. Clicking the `Apply` button will simultaneously apply both the Rulebook and Business Rule filters, as well as the selected states. Eg. Show Rows where `Business Rule A` was Invalid.
- **Browsing Arrows** - A set of arrows designed to assist you in navigating the Preview data table. The central arrows facilitate movement between adjacent pages, while the outer arrows navigate to either the first or the last page of the Preview data table.
- **Displayed Rows** - The rightmost dropdown, located at the far end, permits you to select the number of data rows per page you desire to be displayed in the table below.

Every row within the Preview data table features either a green or red button on the far left side. Beyond signifying the validation status, clicking on it will reveal the results per row in a drawer referred to as `Row Result`.

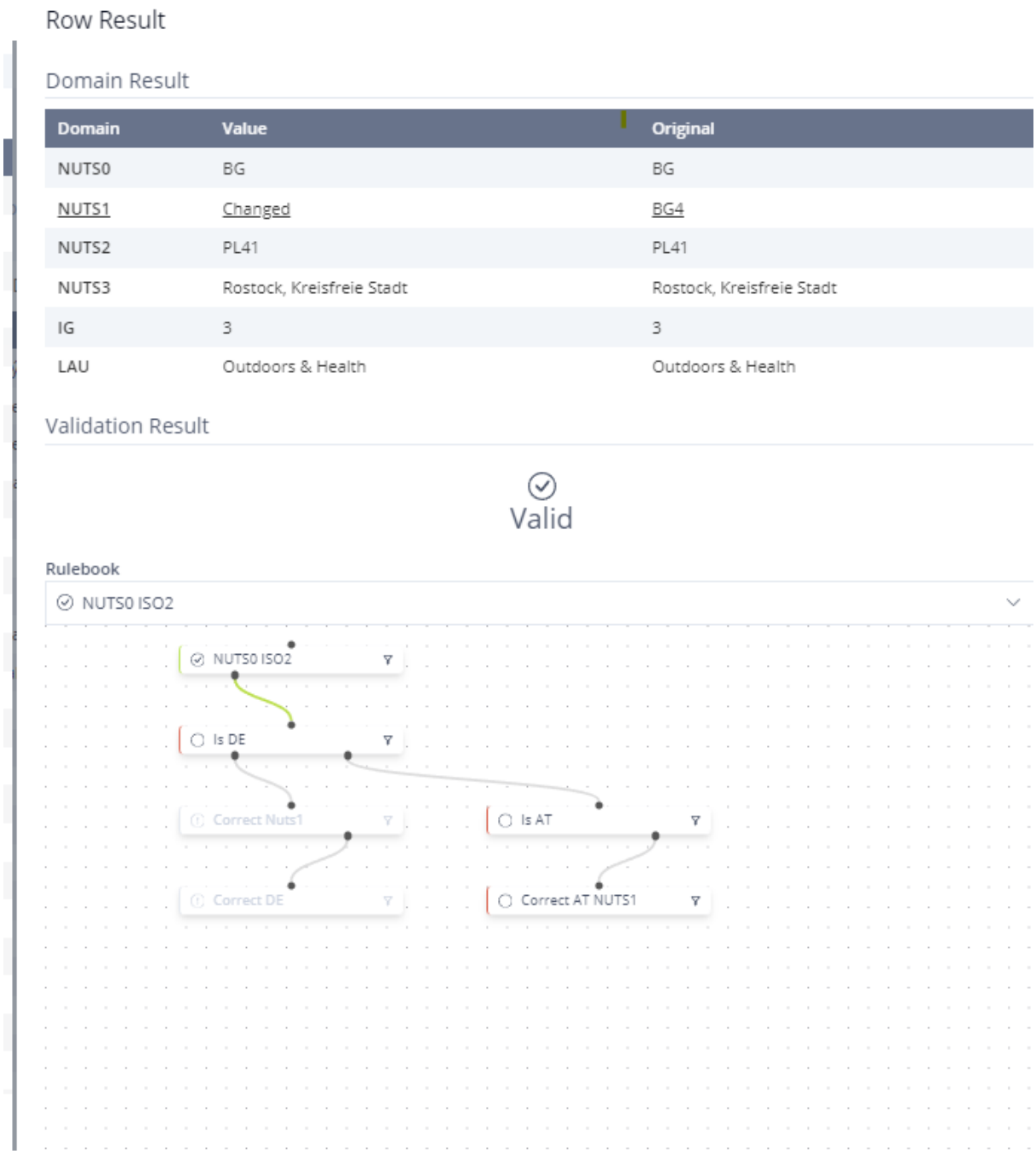


Figure 39: Preview Row Result

The drawer comprises a Domain Results section, presenting each Domain in a list with two additional

columns indicating the current value in that Domain and the original value.

The second section provides an overview of the overall validation status.

In the third section, you can navigate through each Rulebook against which the data rows have been checked. Here, you can view the business rules flow under the selected Rulebook on the canvas. Each Business Rule is accompanied by a filter button on the right side. Clicking on this button will filter the entire Preview data table based on the business rule and its validation status.

3.13.7.3 Runner Data Upload - In Detail

Enabling Data Upload when executing HEDDA.IO Runner is accomplished differently based on the specific HEDDA.IO Runner you are utilizing.

Dotnet Runner:

```
Hedda.Create(url, apiKey)
  .EnableDataUpload();
```

PySparkRunner:

```
hedda = Hedda.create(url, apiKey)
hedda.enable_data_upload()
```

The data generated by this process can subsequently be examined. On the Run Screen, the presence of the `Has Data` flag alongside Executions signifies the availability of data for review.

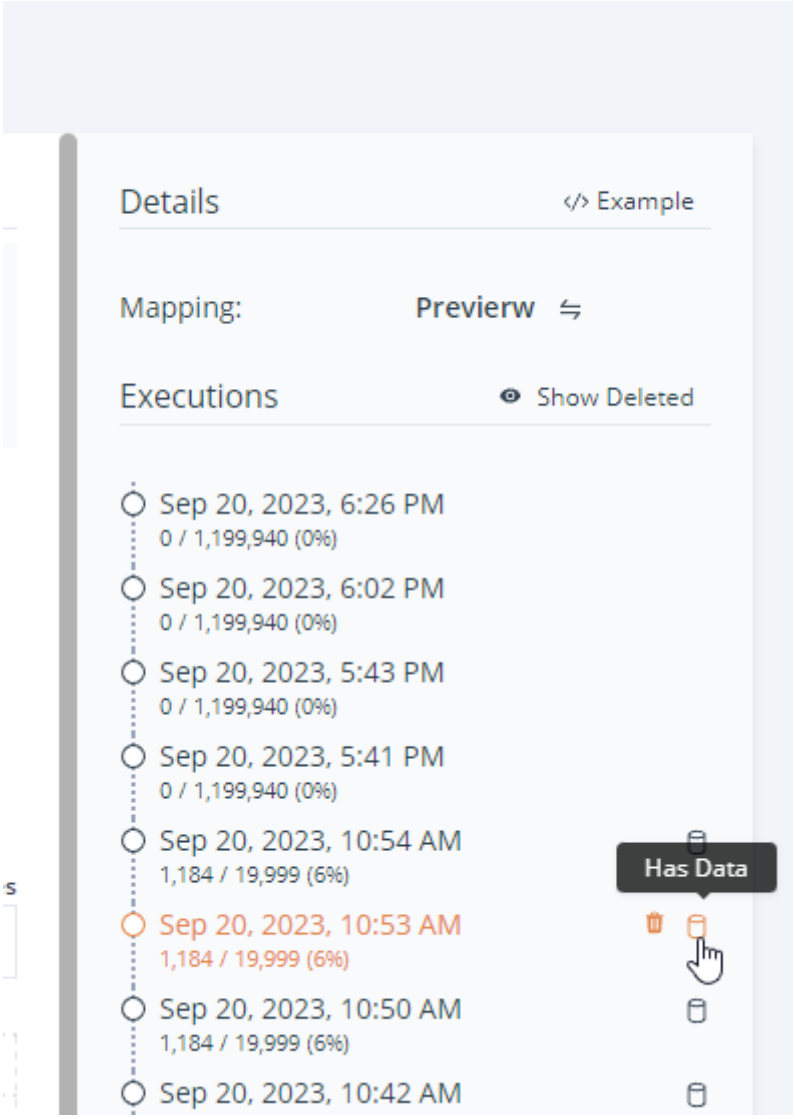


Figure 40: Run overview - Has Data

When inspecting the Execution, you can click on `Analyze Result set` to examine the corresponding data. Similarly, you can review past Previews in a similar manner. On the Data View Page, Executions initiated directly by a Runner need to be registered; this can be accomplished by clicking on the `Register` button.

3.14 Domains Page

It is essential to create a Domain for each data column in your dataset, that you want to execute against HEDDA.IO.

The Domains tab is split into 3 panels:

- Domains Browsing Panel
- Members Panel
- Domain Info Panel

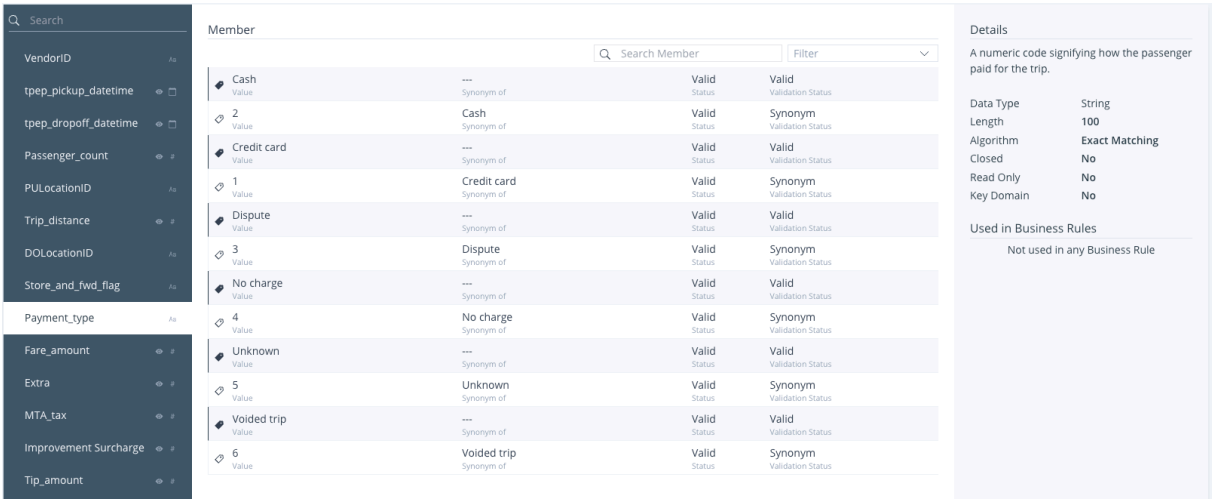


Figure 41: Domains

Let's take a look at these sections more closely, starting from the Domains Browsing Panel on the left.

3.14.1 Domains Browsing Panel

Presented here is a compilation of already created Domains.

Adjacent to each Domain on the right-hand side, you'll find small icons denoting the Domain's data type, along with indicators for whether it is in a read-only state, closed, or falls into neither category.

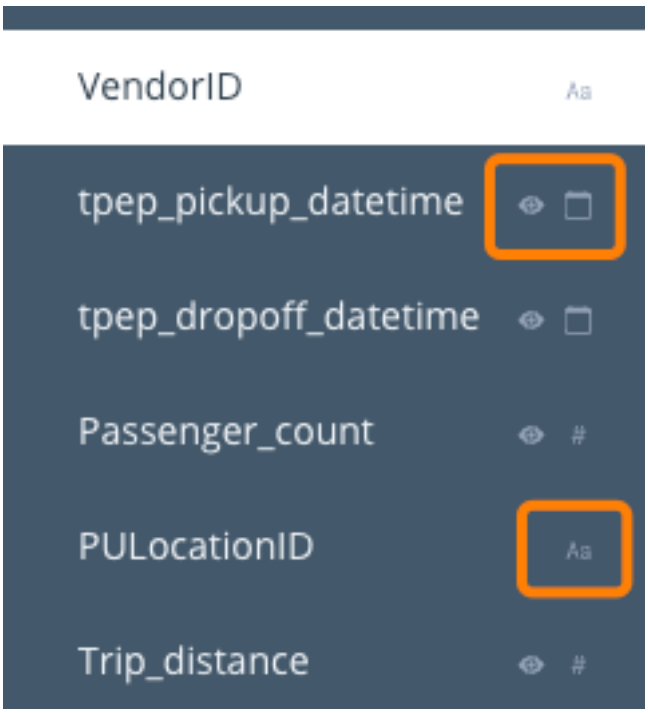


Figure 42: Domain Data Types

In `Edit Version` mode, you can also change the position of the Domains in the browsing panel, by grabbing the handle on the left side of the Domain, and moving it vertically along the list of Domains.

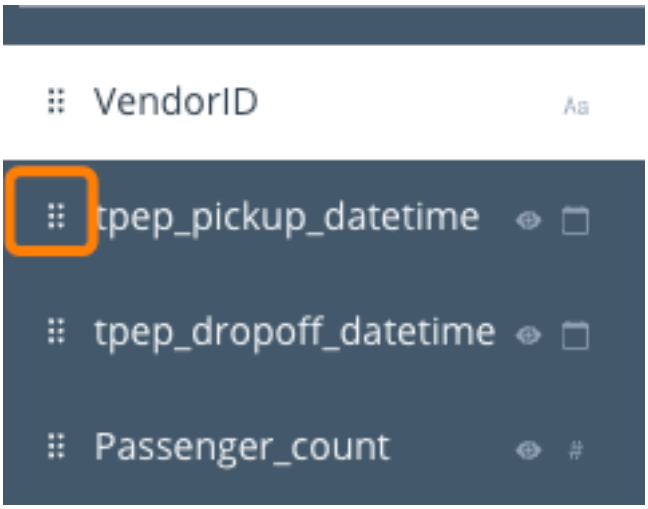


Figure 43: Domain position

Let's take a look at how you can add a Domain.

3.14.1.1 Add Domain

In order to add Domains in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page. Once there, click on the Domains tab in the navigation bar.

The `Add Domain` button is then located at the top of the Domains Browsing Panel. The button is, of course, only visible in Edit Version mode.

Clicking on the “Add Domain” button, opens the `Add Domain` drawer and form.

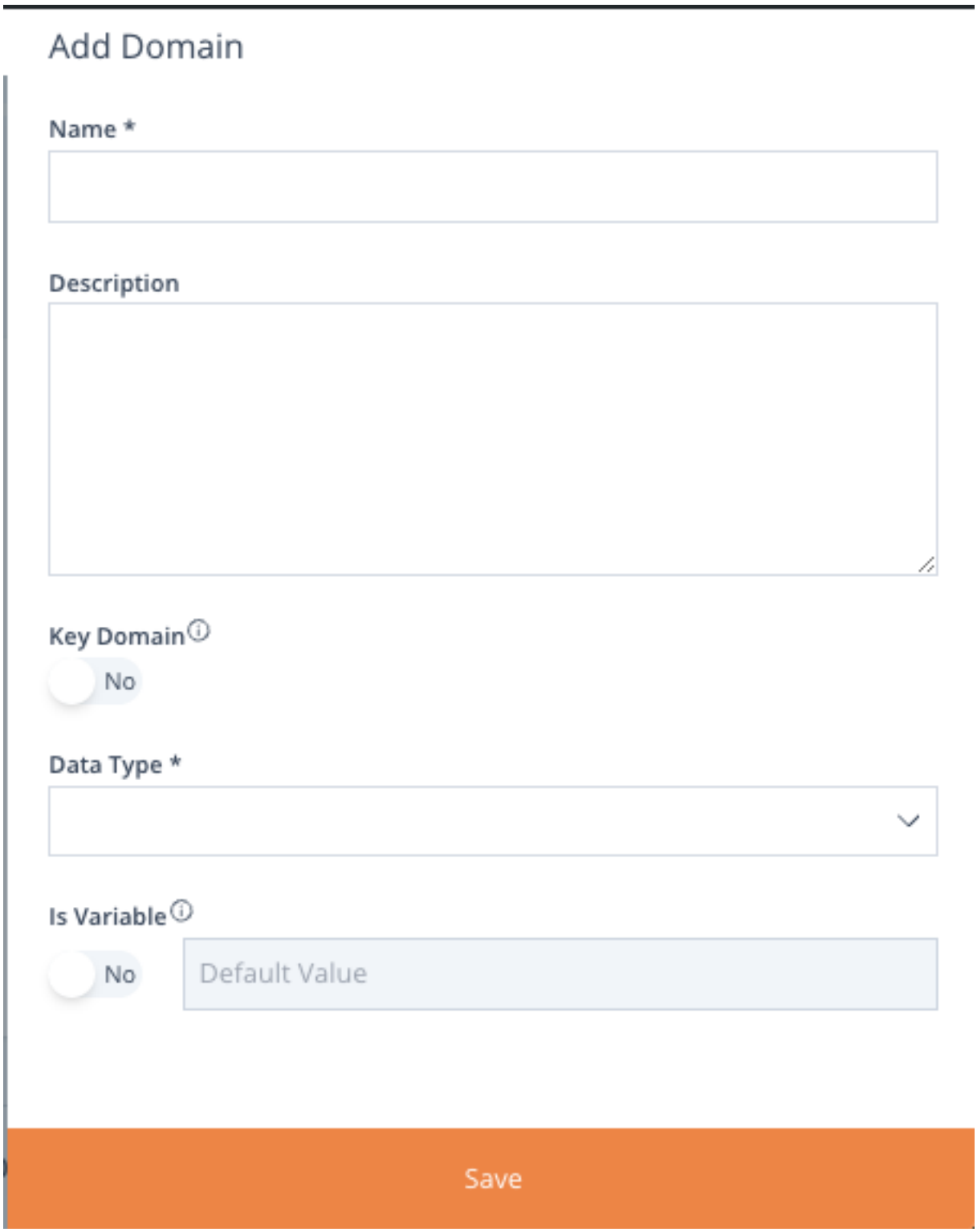


Figure 44: Add Domain

The form has three text inputs and four switches:

- **Name**: the name of the Domain.
- **Description**: the description for the Domain.
- **Data Type**: the type of data of the Domain.
- **Key Domain** (slide): means the Domain is an identifier or part of a combined identifier for a tuple. Only meta information.
- **Is Variable** (slide): allow you to use this Domain as a variable. This means that the Domain can be used to assign values dynamically during an execution or a default value can be set. The variable can be used in conjunction with other functionalities, such as Business Rules, in Actions or Preparation Stages.

There are currently four data types supported for the Domains, and each of them will reveal additional input fields and/or options:

- String
- Number
- Boolean
- Date

3.14.1.1.1 String When selecting data type `String`, several additional inputs will be revealed allowing you to enter the Length of the value and the option of enabling Member Search.

Figure 45: Add Domain - Strings

Member Search will allow you to add main and synonym members, which in conjuction with Algorithms, will help correct and/or standardise your data. Read more about Members and Member Search

Enabling Member Search reveals the dropdown for the Algorithm selection, and two slides:

- **Closed**: this means that only Members already defined for this Domain can be valid.
- **Is External**: means the Members are loaded from an external source.

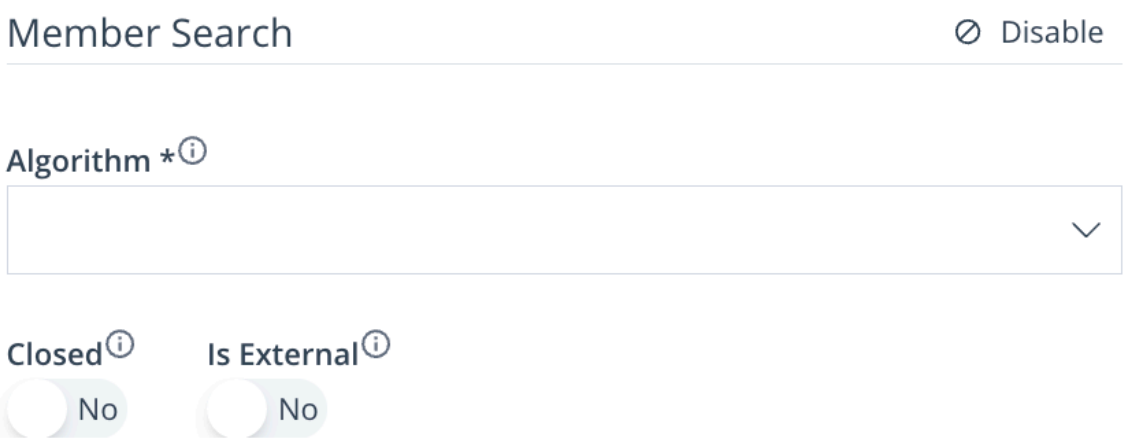


Figure 46: Add Domain - Member Search

Let's take a look first at the Algorithms and then at how "Is External" works.

3.14.1.1.2 Algorithm List Here's an overview of all the Algorithms currently available in HEDDA.IO.

Algorithm	Description
Cologne Phonetics	The Cologne Phonetic (also known as Kölner Phonetik) is a Phonetic Algorithm that assigns a sequence of digits to words according to their sound, the phonetic code. This procedure aims to assign the same code to words with the same sound to implement a similarity search for search functions.
Levenshtein	Levenshtein Distance is a string comparison method between two strings or sequences. Unlike other methods such as Cologne Phonetic, it measures the distance between two strings and not the phonetic sound.

Algorithm	Description
Double Metaphone	Contrary to the original algorithm “Metaphone” (Wikipedia) whose application is limited to English only, this version takes into account spelling peculiarities of several other languages.
Jaro	Jaro or Jaro-Winkler is a string method that compares two strings or sequences. Unlike other methods such as Cologne Phonetic, it measures the distance between two strings and not the phonetic sound between them.
Longest Common Substring	The Longest Common Substring algorithm is to find the longest string (or strings) that is a substring (or are substrings) of two or more strings.
Caverphone	The Caverphone algorithm is based on a similar concept as the Metaphone algorithm and can be used to match english names but also general english words.
NYSIIS	The New York State Identification and Intelligence System Phonetic Code, commonly known as NYSIIS, is a Phonetic Algorithm devised in 1970 as part of the New York State Identification and Intelligence System. It features an accuracy increase of 2.7% over the traditional Soundex algorithm.
Phonem	The Phonem algorithm mainly targets german names.
Phonetex	The Phonetex algorithm works best in matching words that sound alike but are spelled differently in the english language. It is an improvement over the Soundex algorithm.
Phonex	Phonex is an adaption of the Soundex Algorithm. It has an increased performance when used for the english language.
Phonix	Phonix is an extension of the Soundex Algorithm. It has been found to be particularly useful when applied to personal names in a multi-lingual environment.
Exact Match Case Insensitive	Checks if two strings are identical to each other without case sensitivity.
Exact Match	Checks if two strings are identical to each other with case sensitivity.
Levenshtein Damerau	Extends the Levenshtein Algorithm by allowing more operations to calculate the string distance. It will be less than with the Levenshtein Algorithm and can be useful, e.g., for human misspellings.

Algorithm	Description
Keyboard Distance	Calculates the distance between two strings based on the chosen Keyboard layout. The closer keys are on the chosen Keyboard Layout, the smaller is their distance.

In the Appendix you can find a compairison of the Return values for different combinations.

3.14.1.1.3 Is External Switching the `Is External` switch on, will allow you to connect to an already configured External Connection.

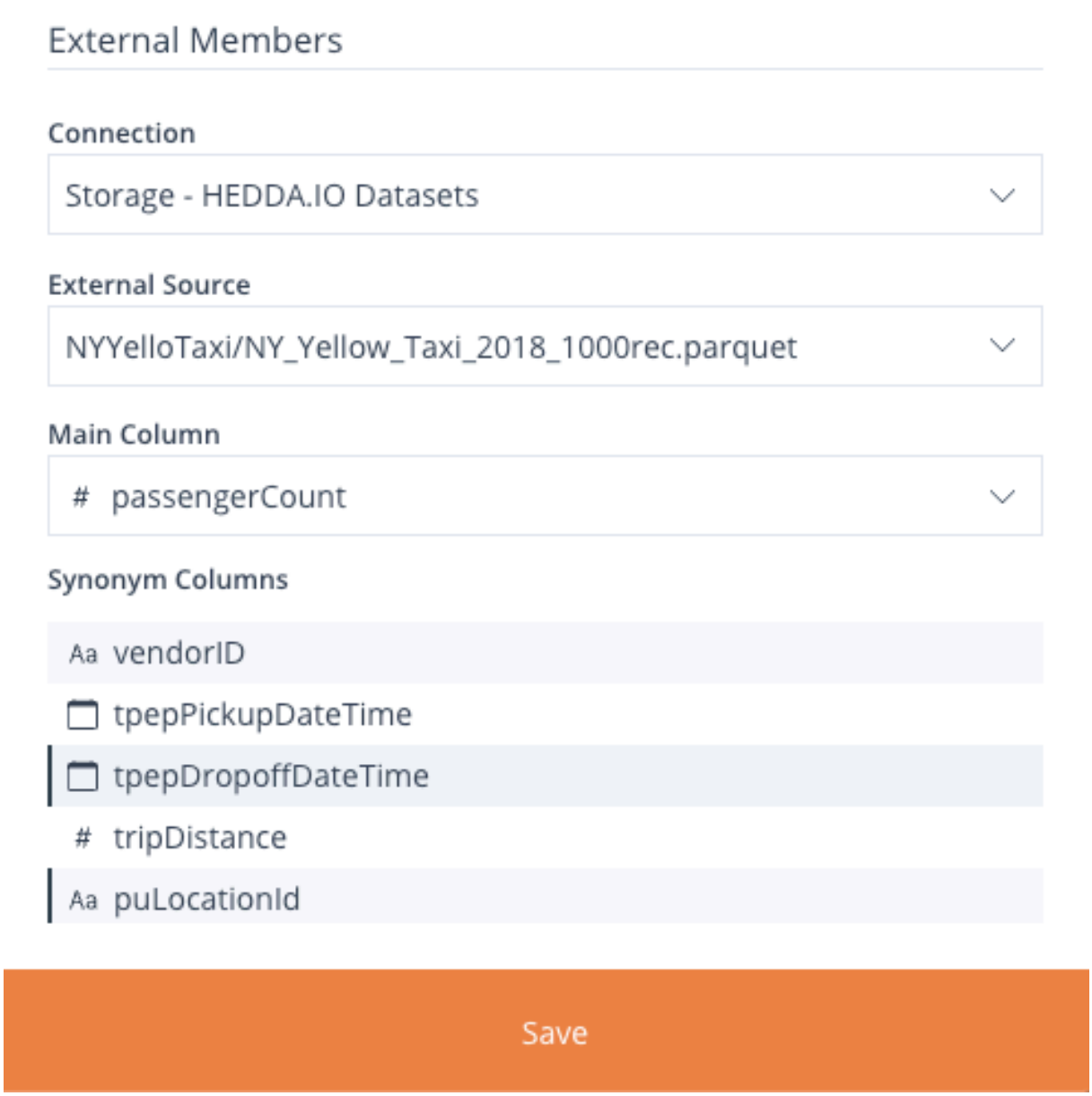


Figure 47: External Members

This will make the `External Members` section visible, which contains 3 dropdown inputs and a list input:

- **Connection**: select an existing External Connection.
- **External Source**: select the path to the dataset within the connection.
- **Main Column**: select the column you need it to act as the source for main Members.
- **Synonym Columns** (list): select the column/s you need to act as Synonym Members for the main column selected above. A simple click will select the column, and a second click will uns-

elect it.

Click the Save button once you finish the setup.

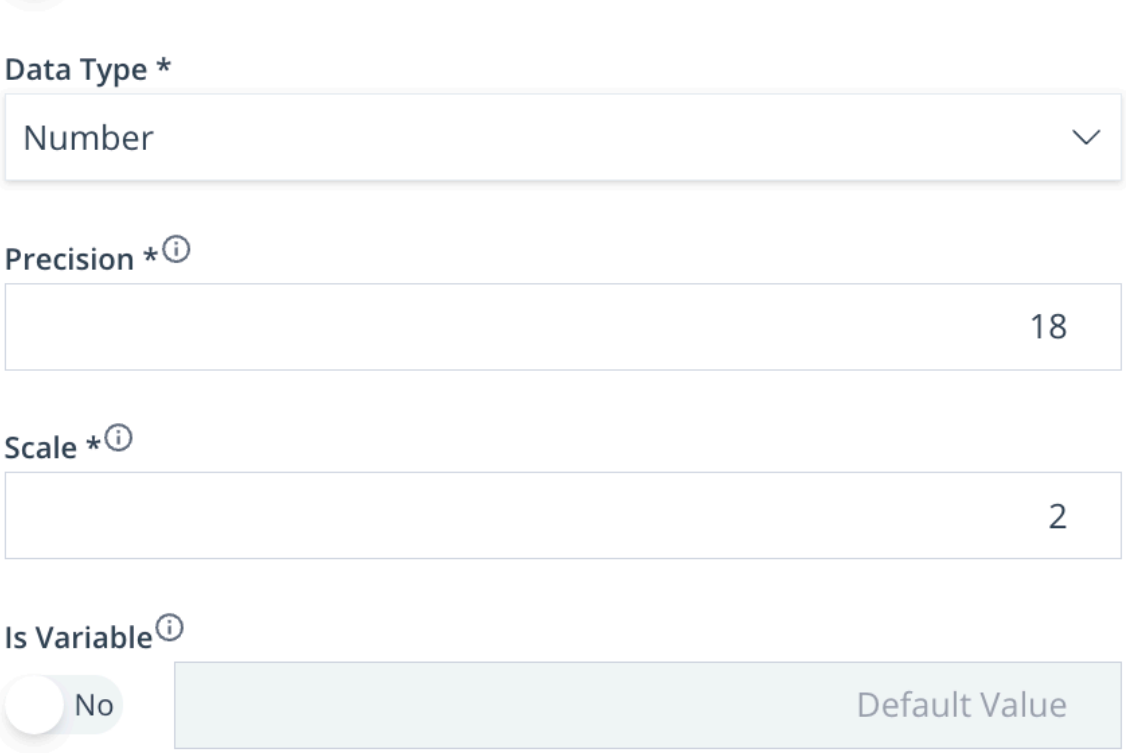


Figure 48: Add Domain - Precision, Scale and Read Only

3.14.1.1.4 Number Selecting the Number data type, will make 2 additional inputs visible, allowing you to set:

- **Precision**: the number of digits in the number.
- **Scale**: the number of digits stored to the right of the decimal point.

3.14.1.1.5 Boolean and Date For Boolean and Date no additional inputs are revealed.

3.14.2 Members Panel

The roster of Members in the Members List encompasses the Main Member's value as well as their corresponding Synonyms, accompanied by selected attributes such as Status and Validation Status.

Directly above the list, you'll encounter a search input labeled Search Member, enabling you to locate a Member by its specific name.

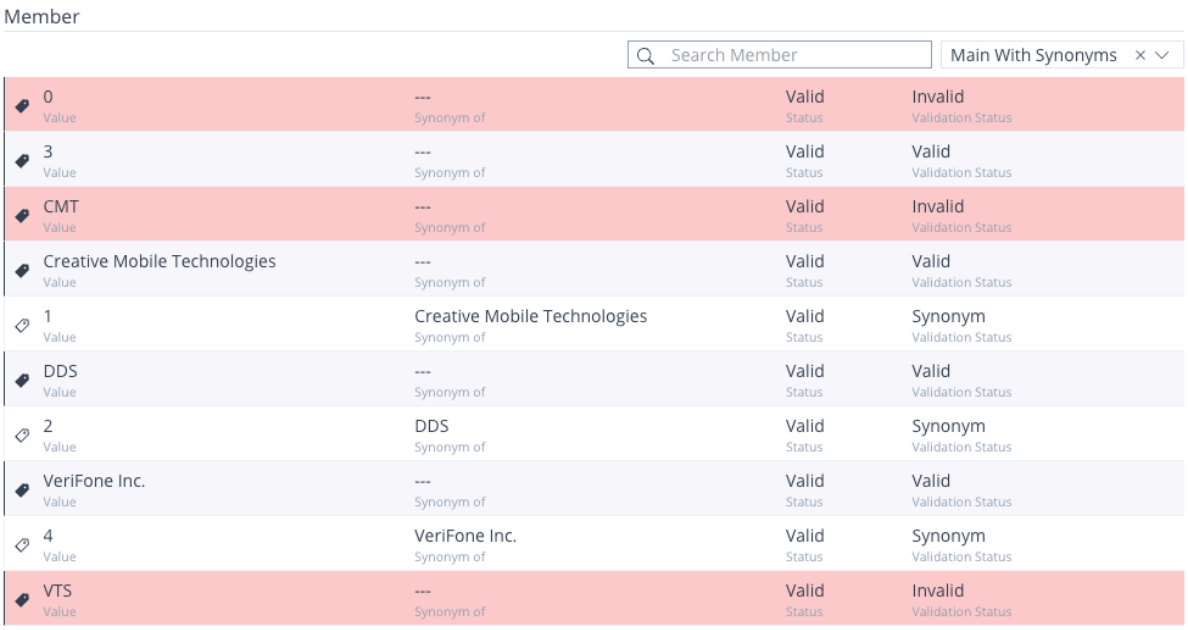


Figure 49: Member list

Adjacent to this input, there is a filter dropdown, providing the option to sort the Members list based on:

- Is Main
- Main with Synonyms
- Valid
- Invalid
- New

To remove the filter, simply hover over the filter dropdown and click the small `x` icon that emerges to the right of the chosen value.

Each Member is accompanied by a small tag icon on the far right side. The black tag signifies Main Members, while the white tag indicates Synonyms.

3.14.3 Add/Edit/Delete Members

In order to add Members to Domains in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once in the Edit Version mode, click the Domains tab and select the Domain you want to add Members to. You'll notice the `Add Member` button located at the upper-right corner of the Members Panel.

Selecting this button will prompt the appearance of the `Add Member` panel, which offers 3 methods for adding Members to the Domain:

- Add Member Manually
- Paste from Excel
- Import Stage

In the following sections, we will elaborate on each of these options, explaining their respective functionalities.

3.14.3.1 Add Member Manually

This functionality enables you to incorporate Members individually, along with any associated Synonyms if applicable.

All these actions can be easily performed within a user-friendly dialog interface.

The `Value` input signifies the value within the corresponding Domain.

Within the center of the panel, you'll find two tabs labeled `Main` and `Synonym`. By clicking these, you can transition to distinct input sets. These buttons are utilized to specify the type of Member being added.

When in the `Main` tab, you have the capability to select a validation status and input Synonyms to the value that was entered in the `Value` input.

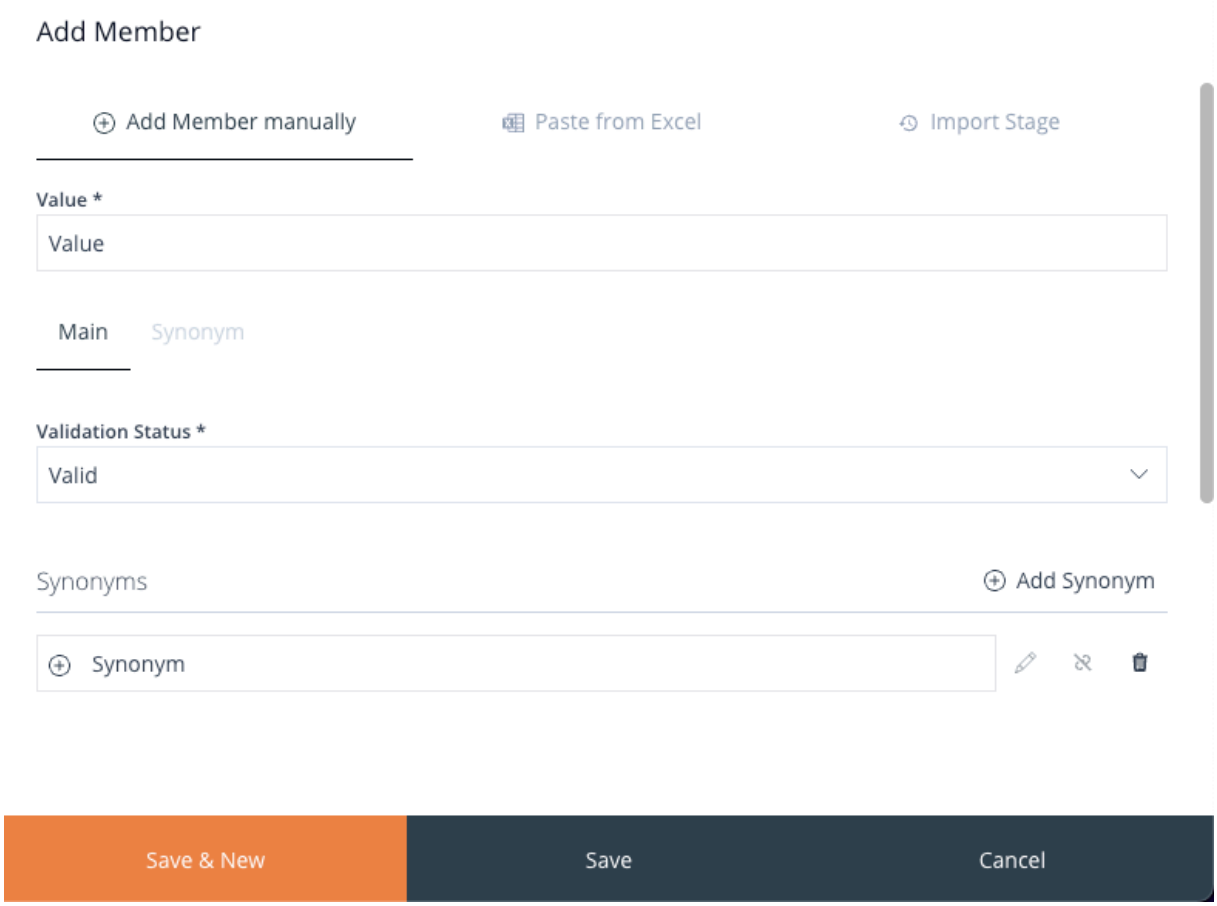


Figure 50: Added Member manually

To add Synonyms, simply click the Add Synonyms button and input the value for the Synonym. It is possible to also delete and edit Synonyms as needed.

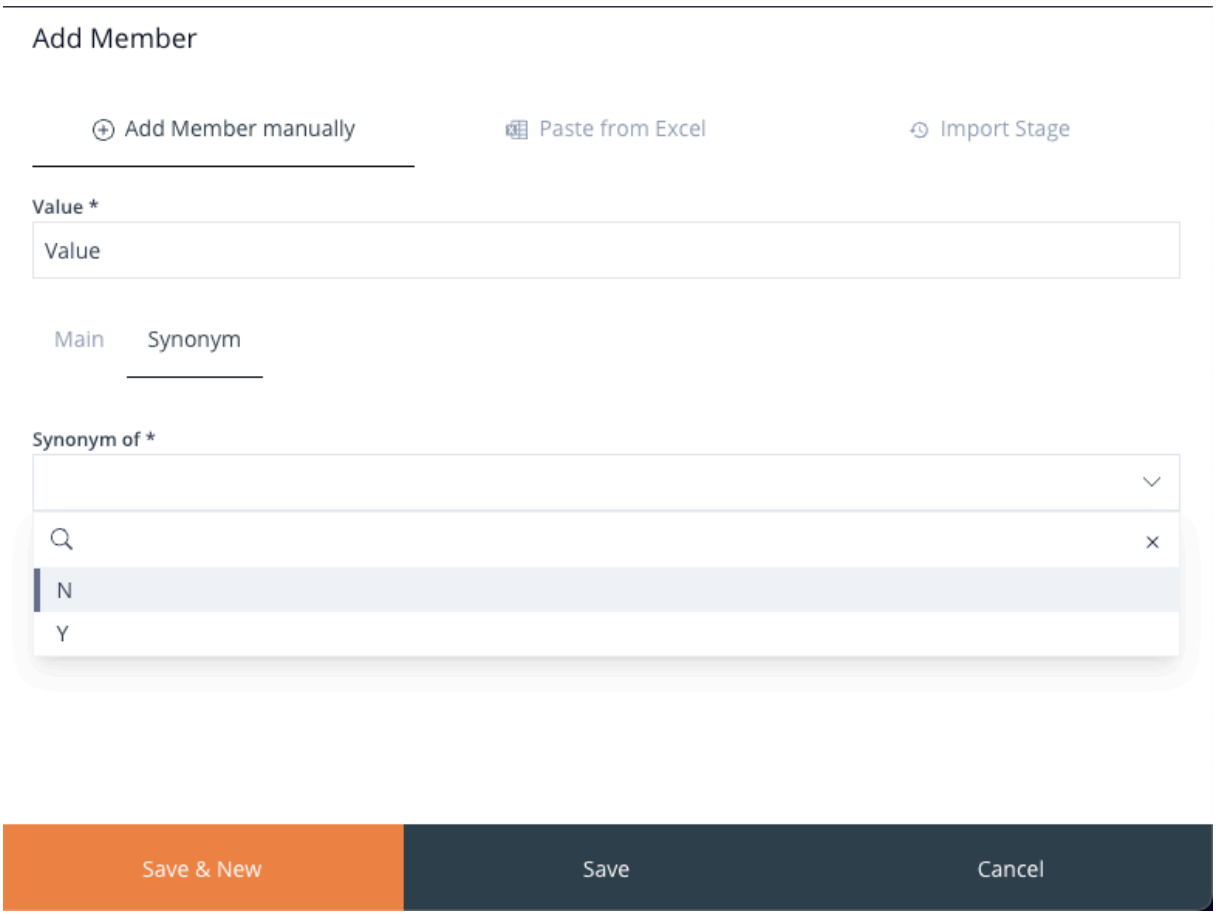


Figure 51: Add Member with Synonym manually

Conversely, when on the `Synonym` tab, you will encounter a dropdown labeled `Synonym Of`. This dropdown enables you to designate the Main Member to which the particular Synonym is associated. The dropdown also has a search input on top of the list of Members, to help you search through longer list of Members.

3.14.3.1.1 Paste From Excel This option allows you to paste entire columns of values from an Excel file.

Once you've accessed the "Paste from Excel" tab, simply click within the space shown on the left, then press Ctrl + V on your keyboard.

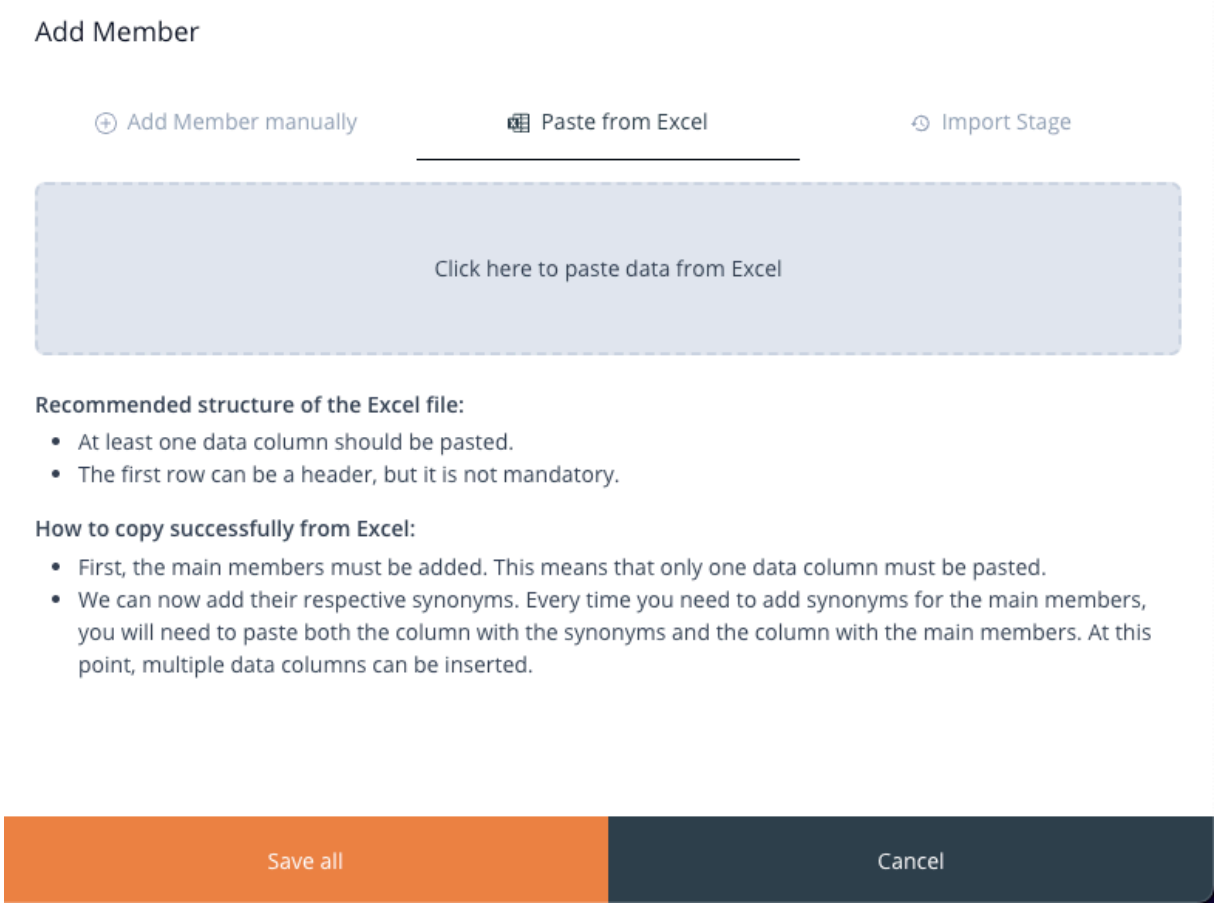


Figure 52: Add Member by pasting form Excel

The values should be exhibited in the manner illustrated in the image example below.

The `Value Column` dropdown permits the user to choose from the various data columns that were copied, if applicable. For this instance, only one column was copied, resulting in a solitary column available for display, which is what the dropdown reflects.

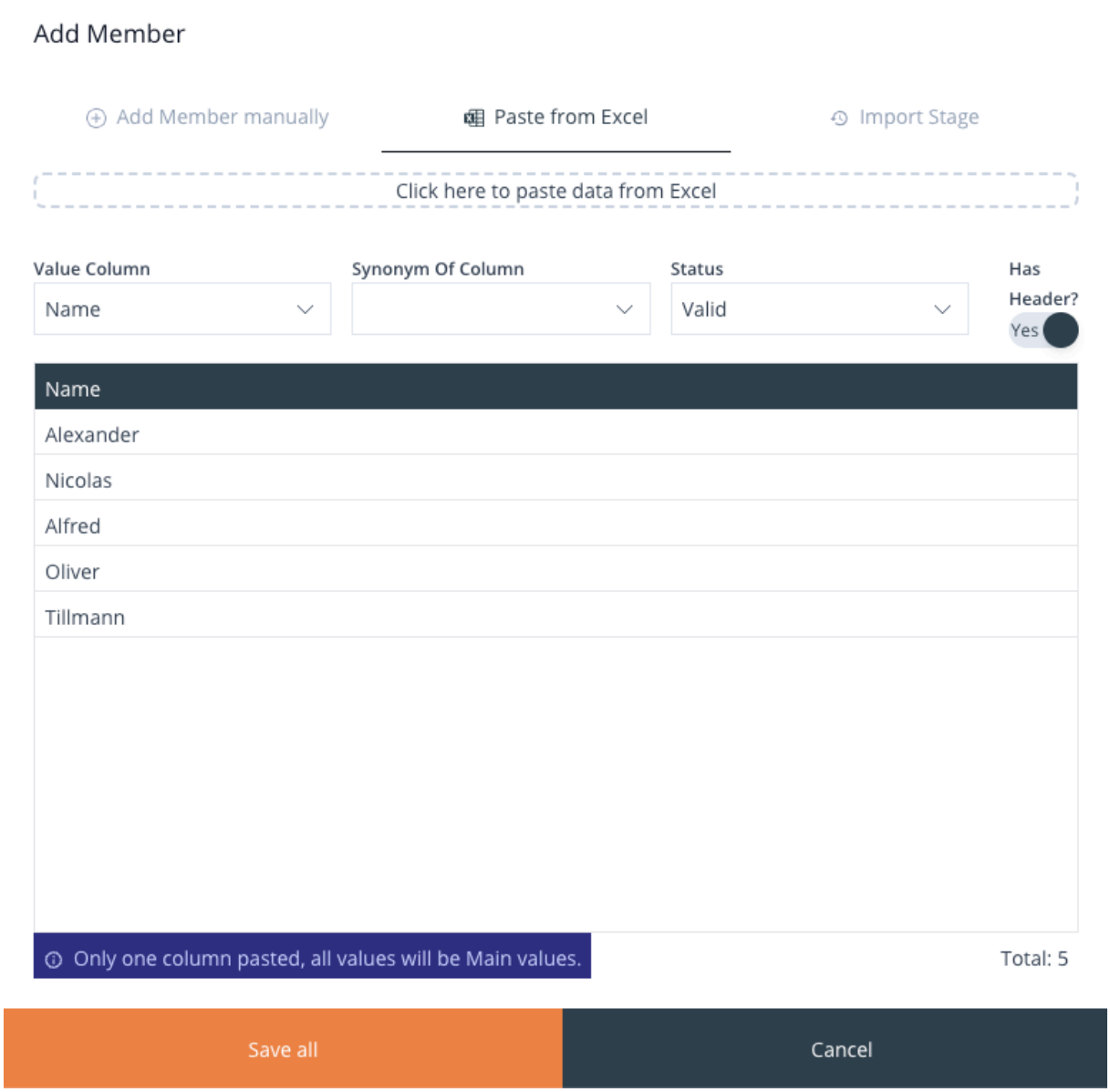


Figure 53: Add Member by pasting from Excel

The `Synonym of Column` dropdown empowers the user to designate the column containing the primary Members. This feature becomes relevant when the column comprising the primary values has been added beforehand.

> **Important Note!** When adding the Synonym columns, it is mandatory that you select the column with the main values as well, before pasting the values into the Drag & Drop component.

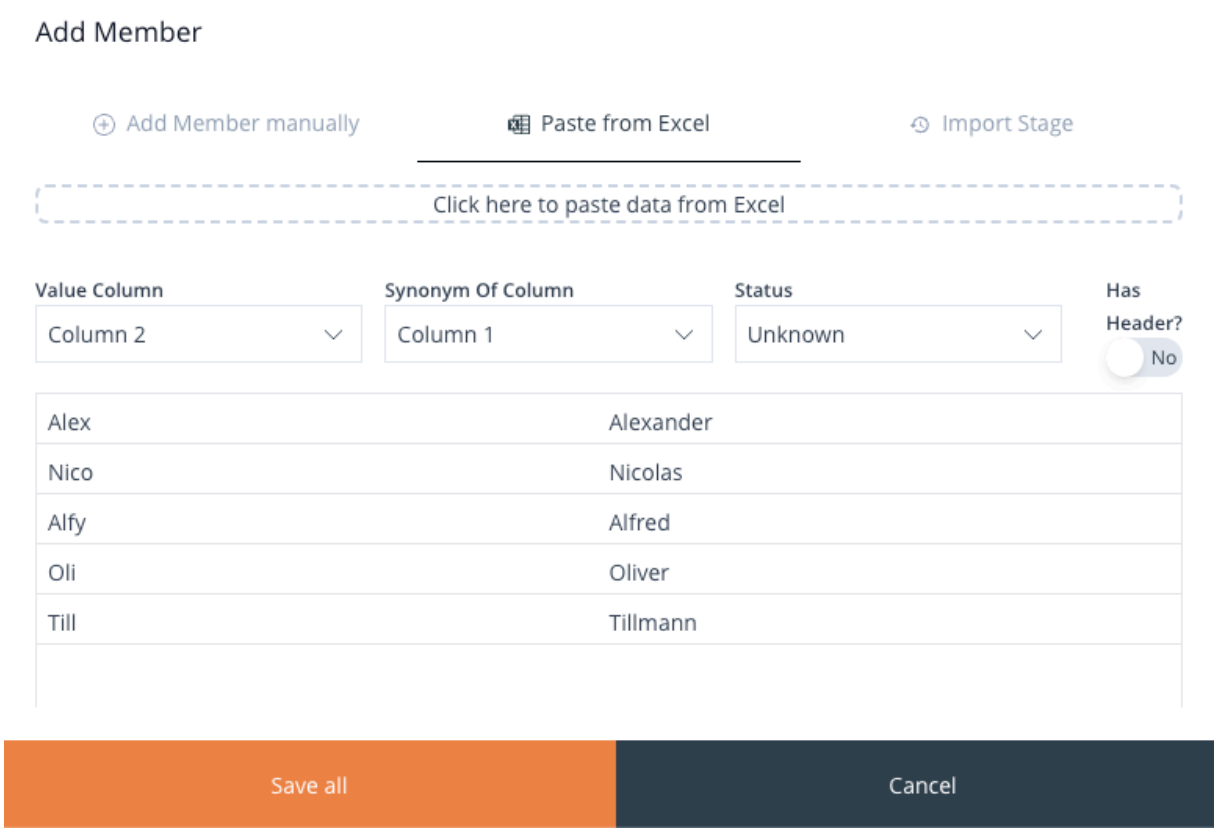


Figure 54: Add Member by pasting from Excel including Synonyms

The `Status` dropdown provides the option to choose between Valid, Invalid, or the default selection of Unknown for the Members.

Lastly, located on the far right side, we find the `Has Header` slider. When activated, this slider designates the top row of the column as a header, automatically assigning the header names to the copied columns. All this information is also available on the panel itself, with detailed instructions.

Upon completing all necessary configurations, proceed to click the `Save All` button at the bottom of the dialog.

3.14.3.1.2 Import Stage The `Import Stage` feature provides you with the ability to import values that have been recently identified in prior Executions. This means that if during previous operations new values were discovered, you can utilize the `Import Stage` to add these new found values as Members.

3.14.4 Domain Info Panel

To the right side of the page we have the Info Panel which displays information on the selected Domain.

The panel is split into 3 sections:

- Details
- Used in Business Rules
- Last Modified (in Edit Version mode)

3.14.4.1 Details* The details section includes the description and the following info, depending on the data type associated with the Domain:

- String
  - Data Type
  - Length
  - Algorithm
  - Closed
  - Read Only
  - Key Domain
- Number
  - Data Type
  - Precision
  - Scale
  - Closed
  - Read Only
  - Key Domain
- Date/Boolean
  - Data Type
  - Closed
  - Read Only
  - Key Domain

In the top-right corner, the ‘Star’ icon allows the user to add the respective Domain to Favorites.

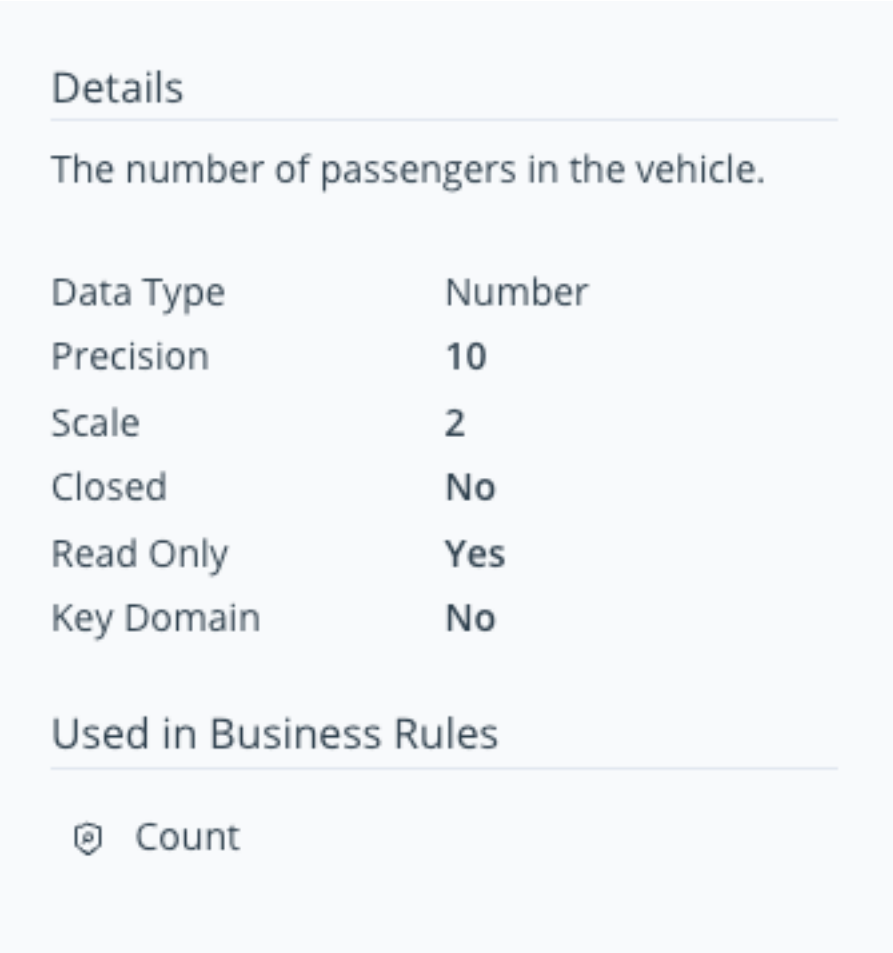


Figure 55: Domain details

3.14.4.2 Last Modified This section will display the name of the user that was the last to make modifications to the Domain, as well as the date on which this occurred.

3.14.4.3 Used in Business Rules This section is only visible in the Edit Version mode, and will list all the Business Rules which are currently using the Domain for Conditions or Actions.

3.14.4.4 Edit/Delete Domains In order to edit and/or delete Domains in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once there, find and click the Domains tab. Select the Domain you wish to edit or delete. In the top-right corner of the Info Panel, notice the `Edit` button. Click to open the `Edit Domain` drawer and form, which allows you to edit the selected Domain in a form similar to the `Add Domain` form, or delete the Domain by using the `Delete` button at the bottom of the form.

3.15 Rulebooks Page

In the Rulebooks section, you'll encounter a roster of Rulebooks. Each Rulebook is accompanied by its own dedicated Business Rule canvas. This canvas showcases all the Business Rules contained within the corresponding Rulebook, along with illustrating the interconnections among these rules.

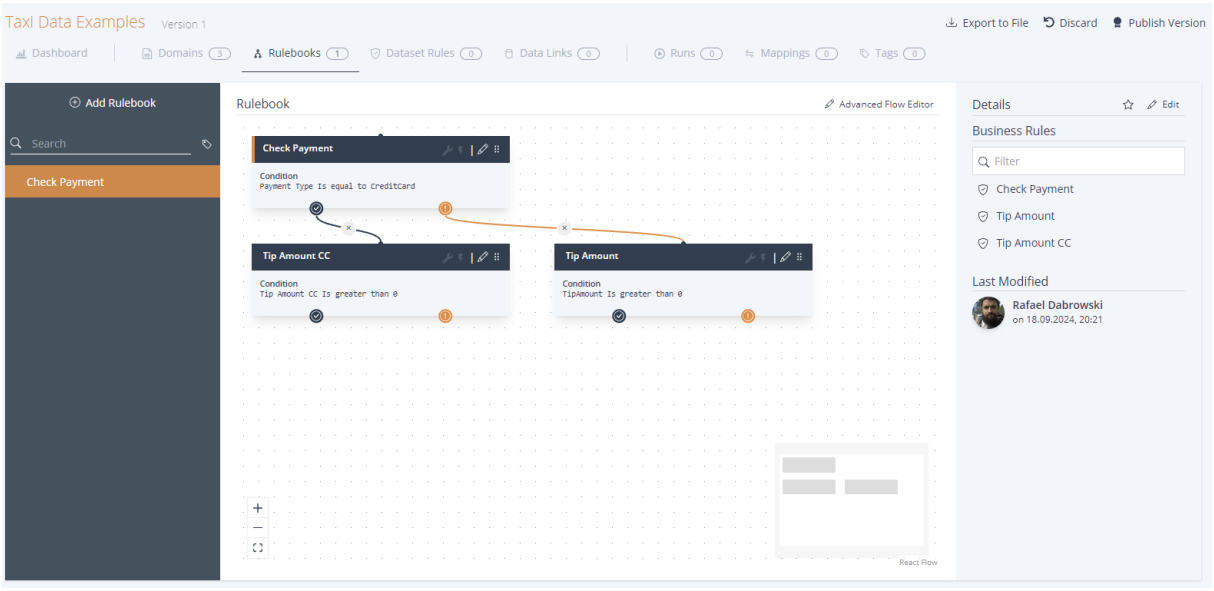


Figure 56: Business Rule Overview

This page is divided into 3 separate panels:

- Rulebooks Browsing Panel
- Business Rules Canvas
- Rulebook Info Panel

3.15.1 Rulebooks Browsing Panel

The Rulebooks Browsing Panel is positioned along the left side of the page. It encapsulates a list of existing Rulebooks, featuring a search input at the list's uppermost part. This input enables you to search for Rulebooks with names containing the specific characters you entered.

3.15.1.1 Add Rulebook In order to add Rulebooks in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

After accessing the Edit Version, navigate to the Rulebooks tab. Here, you'll find the `Add Rulebook` button situated at the upper section of the Rulebook Browsing Panel on the left-hand side.

Upon clicking the `Add Rulebook` button, a drawer will slide out with a form. This form provides you with the option to input a Name and a Description for the newly created Rulebook.

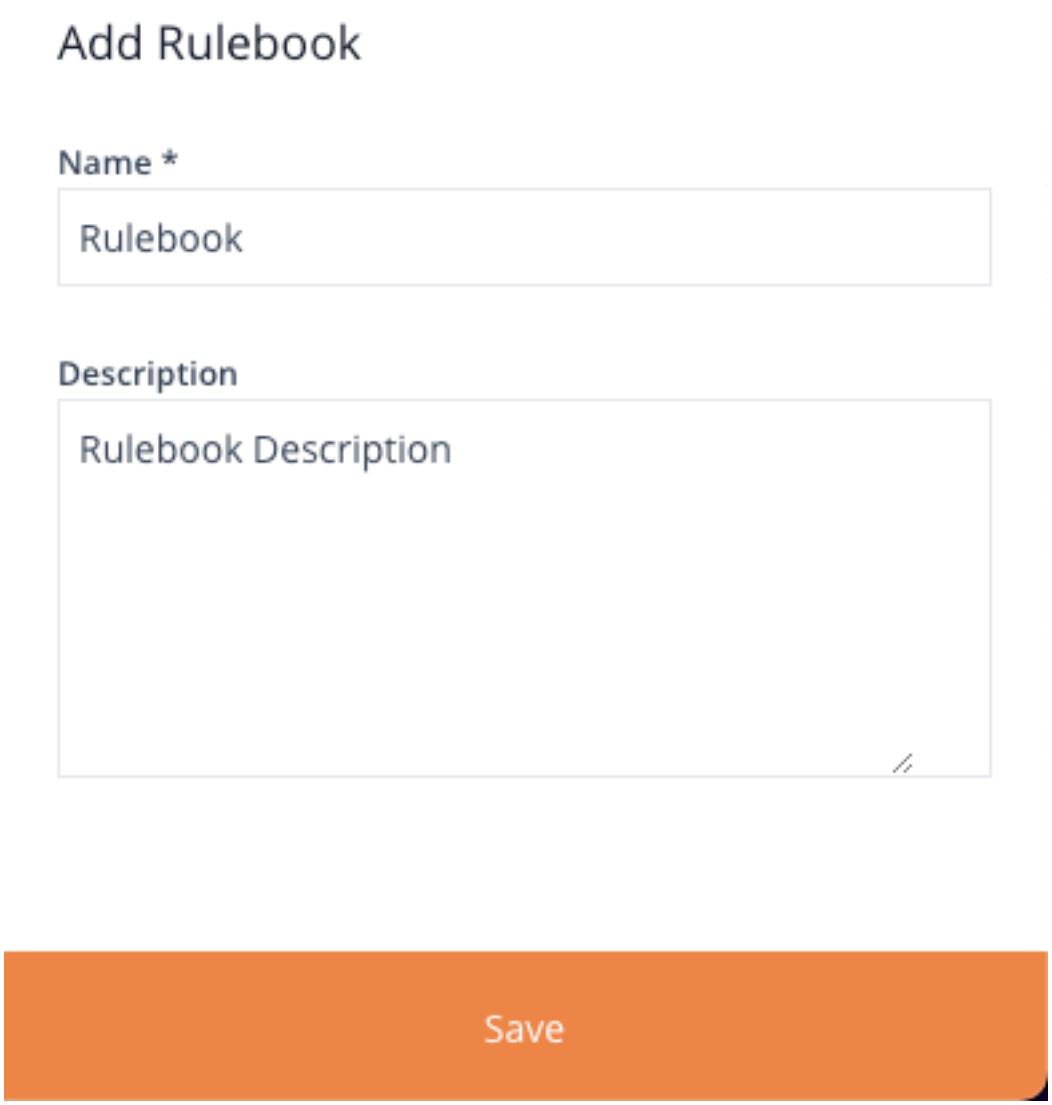


Figure 57: Add Rulebook

Upon selecting the Save button located at the bottom of the form, the following actions will occur:

1. The Rulebook will be appended to the list of existing Rulebooks.
2. A default initial Business Rule will be automatically added to the Rulebook.
3. The Edit Business Rule drawer will be revealed, enabling you to modify the default Business Rule within the Rulebook.

3.15.2 Add Business Rule to Rulebook

The Add and Edit Business Rule drawer share the same form.

The form is split into 3 distinct sections:

- Business Rule metadata
- Preparation
- Conditions
- Actions

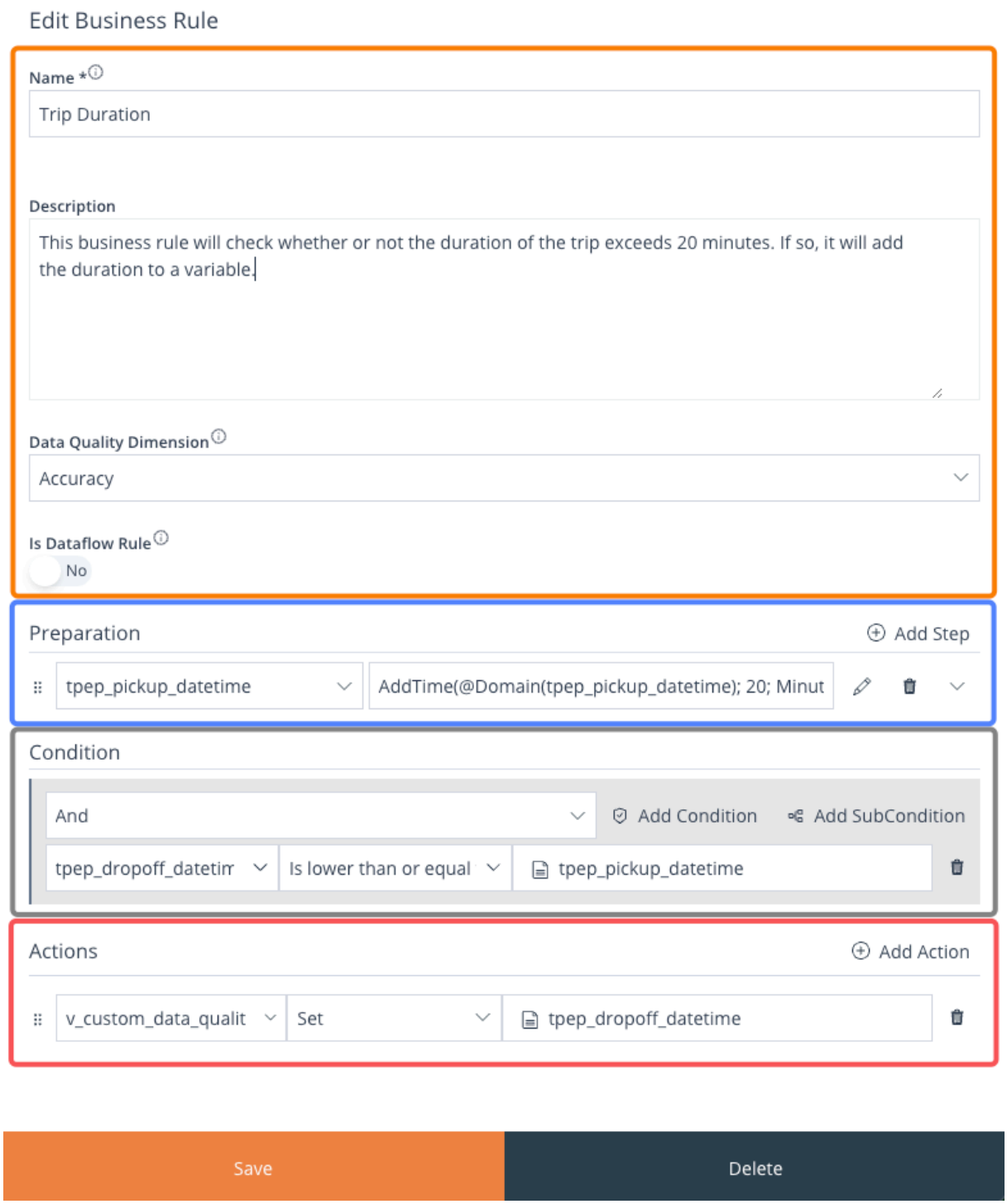


Figure 58: Add Business Rule to Rulebook

3.15.2.1 Business Rule metadata In the Metadata sections, we have 4 inputs:

- **Name**: the name of the Business rule.

- **Description**: the description of the Business rule.
- **Data Quality Dimension** (dropdown) - will offer the option to select one of the 6 Data Quality Dimensions.
- **Is Dataflow Business Rule** (toggle slider) - If enabled, this Business Rule will not determine if a row is valid or not.

  The following Data Quality Dimensions are available:

  - **Completeness**: Measures whether all required data is present and accounted for.
  - **Uniqueness**: Ensures that each record or data point is distinct and not duplicated.
  - **Timeliness**: Assesses whether data is up-to-date and available when needed.
  - **Validity**: Checks if data conforms to the defined formats and Business Rules.
  - **Accuracy**: Verifies that data correctly represents the real-world values it is intended to model.
  - **Consistency**: Ensures that data is uniform and free from contradictions.
  - **None**

3.15.2.2 Preparation

To initiate preparation you need to click the `Add Step` button on the right.

This will add a Preparation step just below.

The item encapsulates:

- **Domain selection dropdown field**: allows you to select the field on which the preparation should be performed.
- **Formula display label**: will display the formula after it is set up.
- **Edit button**: will open the Preparation item editor.
- **Delete button**: will delete the Preparation item.
- **Copy button**: will copy the Preparation item as an additional step.

3.15.2.3 Formula Editor

To open the Formula Editor for a Preparation step, click the `Edit` button on the right.

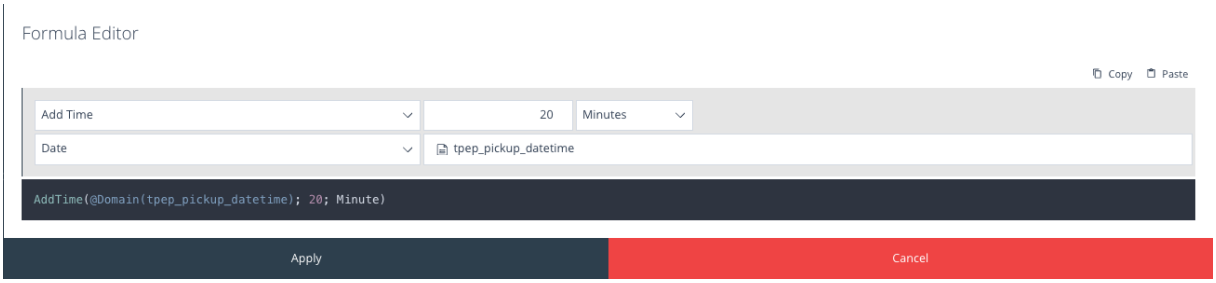


Figure 59: Business Rule Preparation - Formula Editor

Once the Formula Editor pop-up panel opens, you can see it being split into two distinctive parts, the top part will allow you to configure the formula, while the bottom part, with the dark background, will display the configured formula in an Excel like manner. Additional buttons on the panel will allow you to copy and paste formulas.

The first dropdown will allow you to select an operation or function that will be performed, while the input field/s on the right, will allow you to input the values the selected operation or function will perform on.

Depending on the data type of the field we want to perform the preparation on, different preparation operations will be afailable in the Formula Editor.

3.15.2.3.1 For Strings

- **String**: Provides a String value either directly or from a reference like Data Links or Domain.
- **Concat**: Concatenates multiple Strings together, an optional Separator can be entered
- **Lower**: Transform to Lower Case
- **Upper**: Transforms to Upper Case
- **Regex Replace**: Performs a regex Transformation, Capture Groups can be accessed with $1, $2, … in Replace Field
- **Replace**: A regular String Replace
- **SubString**: Gets a part of a String from a zero based Start value and an optional Length
- **Trim**: Removes white spaces from String. Either on Both sides or only from front or end
- **String from Date**: Creates a String from a Date with a specified Format
- **String from Number**: Creates a String from a Number with a specified Format
- **String from Boolean**: Creates a String from a Boolean where The representing Bool Values can be entered

3.15.2.3.2 For Numbers

- **Number**: Provides a Number value either directly or from a reference like Data Links or Domain
- **Addition**: Adds multiple Numbers

- **Multiplication**: Multiplies multiple Numbers
- **Subtraction**: Subtracts multiple Numbers start from top to bottom
- **Division**: Divides multiple Numbers starting from top to bottom
- **Round**: Rounds the Number to desired decimal count. Rounding can be either up, down or commercial
- **Number from String**: Transforms a Number from a String with possibility to provide thousand and decimal separators
- **Number from Date**: Takes the Seconds from Unix Epoch

3.15.2.3.3 For Dates

- **Date**: Provides a Date value either directly or from a reference like Data Links or Domain
- **Add Time**: Adds Time to a Date
- **Date from Number**: Creates a date from a Number based on Unix Epoch seconds
- **Date from String**: Parses a date from a String based on provided Format
- **Current Date**: Provides dates based on the Current Date. Supported:
  - Now
  - Start of today
  - End of today
  - Start of current month
  - End Of current month
  - Start of current year
  - End of current year

3.15.2.3.4 For Boolean

- **Boolean**: Provides a Boolean value either directly or from a reference like3 Data Links or Domain
- **Boolean from String**: Creates a Boolean from a string where the True and False Value can be Provided
- **Boolean from Number**: Creates a Boolean from a Number (0 or 1), otherwise null

For each of these operations you can insert a value which can be static or can reference a value in a different Domain or even in a column in a Data Link.

3.15.2.3.5 Condition This section of the form is where the Business Rule Conditions are defined:

1. First, select the operator `And` or `Or`.
2. Click on `Add Condition`
3. Select a Domain
4. Select a Condition

5. Enter a Value (the value can be another column)

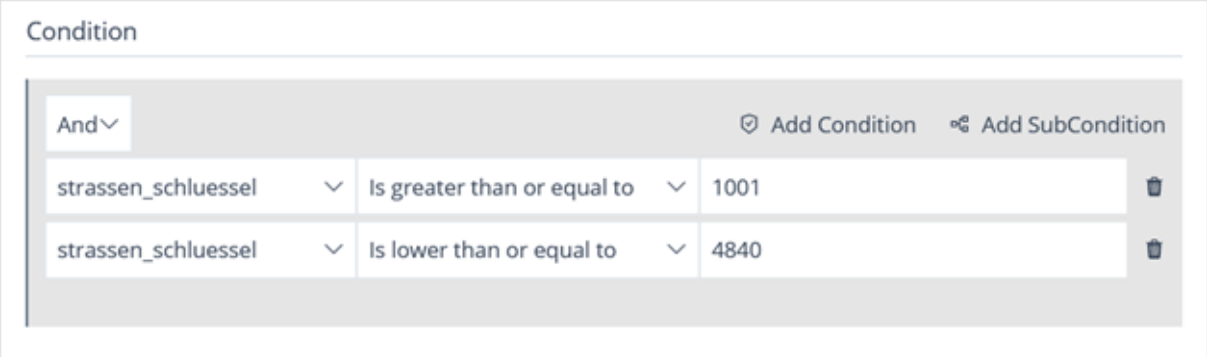


Figure 60: Business Rule Conditions

Repeat the process for every additional Business Rule. You can also add Sub-Conditions which will nest inside the existing Conditions.

Here is a list of Conditions that are available for selection, depending on the data type of Domain you select when configuring a Condition.

Condition	Description	String	Number	Date	Boolean
And	All Conditions must be met.	✓	✓	✓	✓
Or	At least one Condition must be met.	✓	✓	✓	✓
Character Blacklist	String does not contain any of these characters.	✓			
Empty	String is Empty.	✓			
Contains	Text must contain the value.	✓			
Has Date format	Can be Parsed as Date.	✓			
Starts with	Text starts with the value.	✓			
Ends with	Text ends with the value.	✓			
Has exact length of	Text must have the exact number of characters.	✓			
Matches pattern	Text must match the pattern.	✓			
Matches Regex	Text must match the Regex pattern.	✓			
Not Matches pattern	Text must not match the pattern.	✓			
Not Matches Regex	Text must not match the Regex pattern.	✓			
Whitespace	String is only whitespace.	✓			

Condition	Description	String	Number	Date	Boolean
Has maximum length of	Text must not exceed the maximum number of characters.	✓	✓		
Has minimum length of	Text needs the minimum number of characters.	✓	✓		
Is one of	Value must be one of the comma separated sequence.	✓	✓	✓	
Is none of	Value must not be one of the comma separated sequence.	✓	✓	✓	
Is equal to	Both values must be equal (True/False, true/false or 1/0 for bool).	✓	✓	✓	✓
Is not equal to	Both values must not be equal.	✓	✓	✓	✓
Is null	Value is NULL.	✓	✓	✓	✓
Is not null	Value is not NULL.	✓	✓	✓	✓
Is greater than	Left value must be greater.		✓	✓	
Is greater than or equal to	Left value must be greater or equal.		✓	✓	
Is lower than	Left value must be lower.		✓	✓	
Is lower than or equal to	Left value must be lower or equal.		✓	✓	
Is day of week	Day has this day of week.				✓
Is month	Date has this month .				✓
Is not day of week	Day does not have this day of				✓
Is not month	Date does not have this month				✓
Not older than	Date is not older than.				✓
Older than	Date is older than.				✓
Is True	Value is true, if it is: y, yes, true, 1, wahr (case-insensitive).				✓
Is False	Value is false, if it is: n, no, false, 0, falsch (case insensitive).				✓

These can be glued together by Logical Operators:

Logical Operator	Description
And	All Conditions must be met.
Or	At least one Condition must be met.
All True	Condition Check will be skipped

> **Important Note!** The Value for each of the conditions, can also be equal to the value from another Domain in the same Knowledge Base or a value within a Data Link. To do so, simply write @, a dropdown will then present you with a list of the available Domains and Data Links to choose from. See the screenshot below.

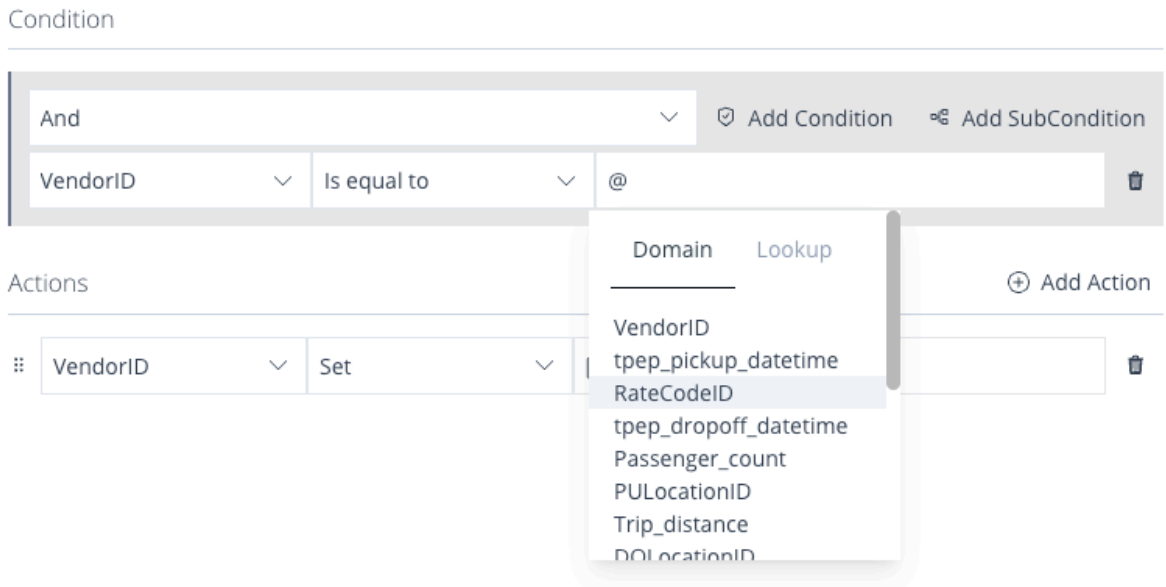


Figure 61: Add Conditions to Rulebook

3.15.2.4 Pattern Matching

In HEDDA.IO Pattern Matching is used within Business Rules as condition operators.

Pattern matching describes a class of problems in which a predefined pattern or rule is provided, and the task is to determine whether a given string (a sequence of letters from a known alphabet) adheres to the specified pattern or rules.

In HEDDA.IO, the Pattern Matching is designed to work only with the * (asterix) and the . (dot).

The * (asterix) serves to match zero or more occurrences of a character. In HEDDA.IO, the * is intended to substitute the .* operator. The . (dot) serves to match any single character except new line.

Example

Let's examine various pattern usages and determine the validity of provided samples.

`A.C` - This pattern will match any string composed of three characters, starting with the letter A, followed by one character in the middle, and ending with the letter C.

Valid Samples: `ABC`,`ADC`, `AVC`, `A^C`, `A  C` Invalid Samples: `ABDC`,`ABD`, `C&C`, `A&$D`

`A*C` - This pattern will match any string composed of any number of characters, as long as the first letter is A, and the last letter is C.

Valid Samples: `ABBC`,`ADBSFC`, `A&%DC`, `ABC`, `A%C` Invalid Samples: `ABDFB`,`AAAA`, `CCBBC`

`INV.....D..-..-....` - This pattern will match any string starting with 'INV,' representingInvoice, followed by any sequence of characters. It then requires the presence of the letter 'D,' symbolizing Date, and concludes with the date itself in the MM-DD-YYYY format.

Valid Samples: `INV12345D12-11-2023`,`INV#1234D03-01-2023`, `INV13BFDD03-01-TEST` Invalid Samples: `INV123456D12-11-2023`,`INV12345D-12-11-2023`, `INV12345DATE-11-2023`

3.15.2.5 RegEx Matching In HEDDA.IO Regular Expression (RegEx) Matching is used within Business Rules as condition operators.

RegEx matching enables you to assess whether a string conforms to a particular syntactic structure. Additionally, it empowers you to search within a string for a substring that aligns with a specified pattern.

Example For instance, let's say we need a Regular Expression to validate an Domain that contains email addresses.

A typical pattern for a valid email address may appear as follows: `[mailbox]@[domain-name].[domain-extension]` or `[local_part@domain_part.top_level_domain]`

In the regex language, this pattern is expressed as: `^[a-zA-Z0-9._-]+@[a-zA-Z0-9-]+\.[a-zA-Z.]{2,3}$`

Let's delve deeper into the aforementioned regular expression.

`[^]` asserts position at the start of the string.

`[a-zA-Z0-9._-]+` Local Part (username/mailbox).

- `[a-z]` matches a single character in the range between a and z (case sensitive)
- `[A-Z]` matches a single character in the range between A and Z (case sensitive)
- `[0-9]` matches a single character in the range between 0 and 9 (case sensitive)
- `[._-]` matches a single character in the list `._-`

- `[+]` matches the previous token between one and unlimited times, as many times as possible, giving back as needed (greedy)

`[@]` matches the character `@` literally.

`[a-zA-Z0-9-]+` Domain Part (Domain name)

Similar to the Local Part, the list of permissible characters is limited to dashes − only.

`[\.]` matches the character `.` literally.

The backslash \ is used as an escape character. Its primary function is to indicate that the character following it should be treated as a literal character rather than as a special character with a special meaning in the regular expression pattern.

`[a-zA-Z.]{2,3}` Top-Level Domain (TLD/Domain extension)

- `[a-zA-Z.]` similar to the Local and Domain parts, excluding digits, and `[.]` matches the character `.` literally (case sensitive).
- `{2,3}` matches the previous token between 2 and 3 times, giving back as needed (greedy)

`[$]` asserts position at the end of the string

In this case, an email such as `john.doe@mail.com` will be valid.

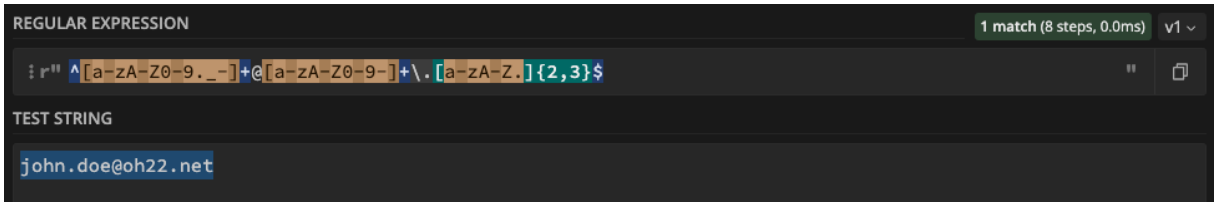


Figure 62: Alt text

Whereas an email like `john.doe#mail.com` will be invalid.

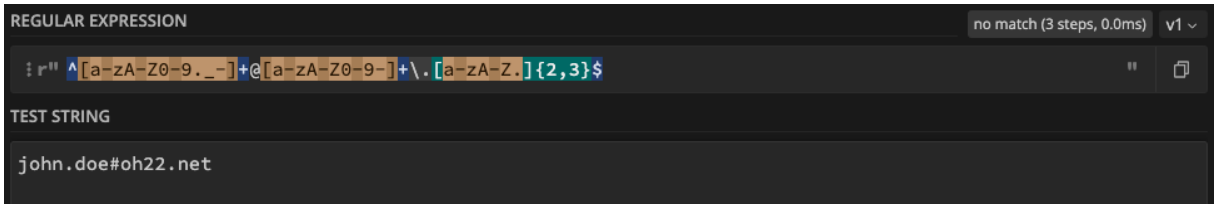


Figure 63: Alt text

3.15.2.6 Actions In this section of the form, you can add Actions, which will be triggered whenever the value in the respective Domain will meet all Conditions of the Business Rule.

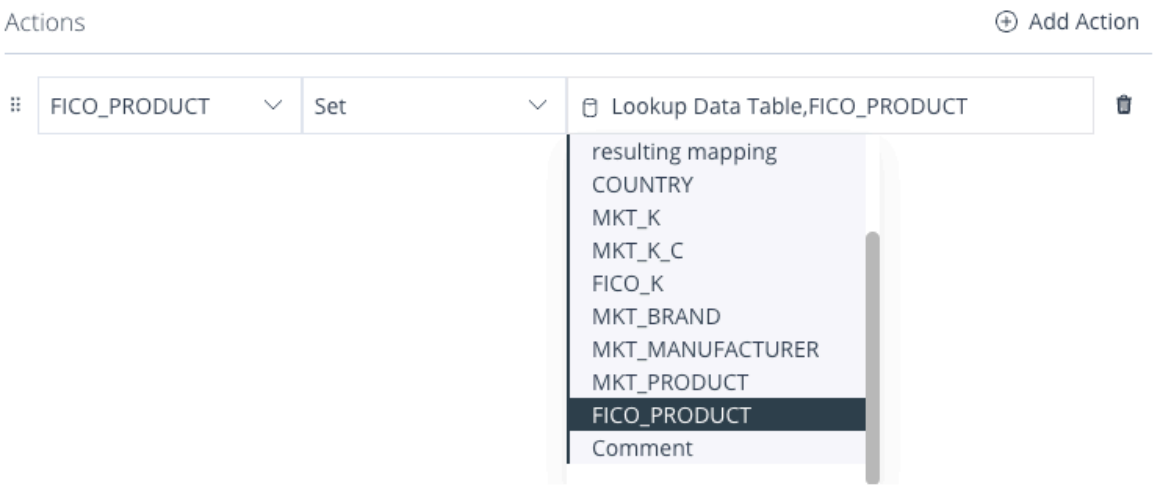


Figure 64: Add Action to Rulebook

To add an Action, simply click the `Add Action` button located on the right side.

A new Action will be appended below, providing you with the ability to choose the Domain on which you wish to execute an Action.

The second input presents itself as a dropdown menu and provides you with a range of Action-Types to select from:

- Set
- Append
- Prepend
- Trim Left
- Trim Right
- Trim
- To Lower
- To Upper
- Replace
- Substring
- Regex Replace

The third input field will only become visible when Actions such as `Set`, `Append`, `Prepend`, `Substring`, `Replace`, or `Regex Replace` are chosen. This field allows you to input a value for the chosen Action.

Similar to Conditions, the **@** symbol can always be utilized to designate a Domain or a column from a Data Link.

3.15.2.6.1 Actions Example - Concatenation The order of execution follows the sequence in which actions are listed. The output of one action becomes immediately available for subsequent actions, either as a Domain to be modified or as a value to alter another Domain. This characteristic can be harnessed for performing more intricate operations.

Example: Concatenation

By utilising `Append` and `Prepend`, it's possible to extend a Domain of the String type, similar to performing concatenation. To achieve this, an initially empty Domain can be assigned the initial value, and successive Append actions can then append additional literals or Domain values. DataFlow Business Rules offer the capability to append specific values or even omit them.

Alternatively, values can be inserted at the beginning of an existing String Domain using Prepend. However, it's important to note that Prepend places the value directly in front of the designated Domain, requiring the values to be prepended in reverse order.

The following example would result in a concatenation of the following parameter list: Value 1, Value 2, Other Text Domain, Target Domain, Value 5, Value 6.

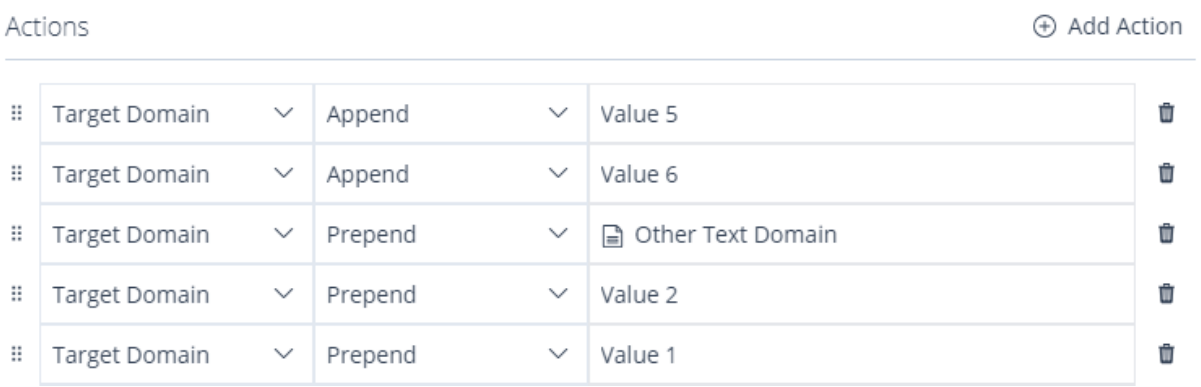


Figure 65: Alt text

3.15.3 Add Subsequent Business Rule

Adding a subsequent Business Rule to a Rulebook will require you to link it with the default or an already existing Business Rule. This means that in order to add a Business Rule, you must click on one of the bottom icons on an existing Business Rule, depending on when you want the rule to be checked and your flow designed. Adding a subsequent Business Rule to a Rulebook will require you to link it with the default or an already existing Business Rule. This means that in order to add a Business Rule, you must click on one of the bottom icons on an existing Business Rule, depending on when you want the rule to be checked and you flow designed.

Figure 66: Add Sub-Business Rule

Clicking on the Checkmark icon will have the new Business Rule checked if the previous Business Rule is Valid. Clicking on the Warning icon will have the new Business Rule checked when the previous Business Rule is invalid.

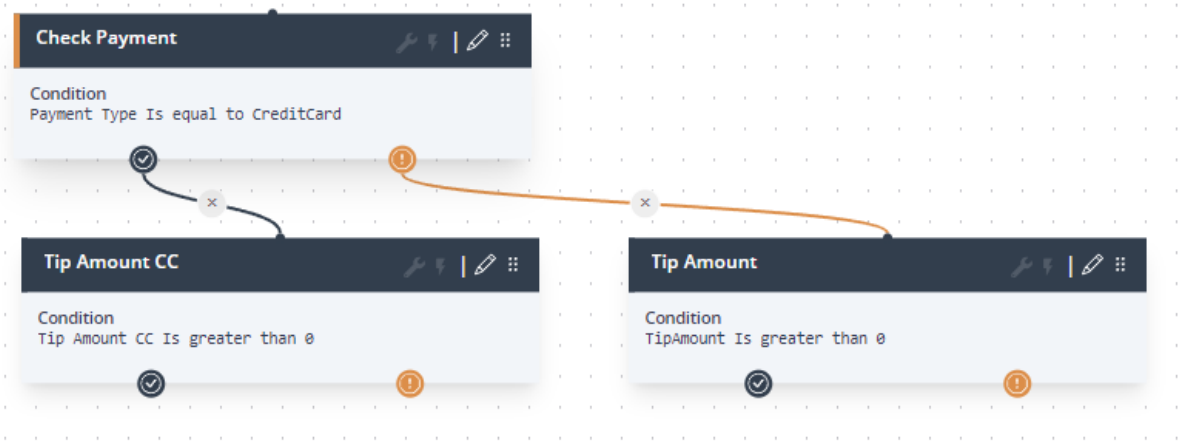


Figure 67: Add Sub-Business Rule

This approach allows you to construct distinct Business Rule flows tailored to align with your validation procedures.

3.15.4 Rulebook Canvas

The Business Rule canvas presents a diagram-like grid where you can add Business Rules and design the sequence of Business Rules as well as the relationship between them. These rules can be used for both validation purposes and managing data/logic flow. The canvas permits zooming in and out and offers the option to center on the chosen object.

Each Business Rule can be dragged around the canvas and positioned at a desired location.

3.15.4.1 Advanced Flow Editor In the Top Right corner you can switch to the Advanced Flow Editor. There it is possible to alter the connections between the Business Rules freely even if it results in Cycles

or Orphans. The changes will not be saved directly but collected and only persisted after clicking on the Save button.

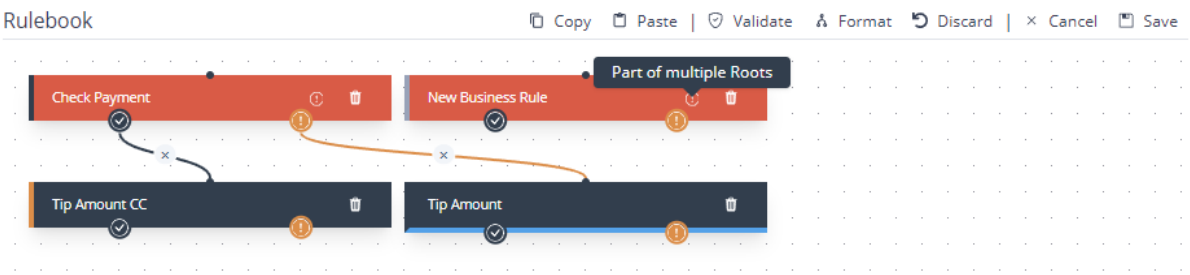


Figure 68: Advanced Editor

Following Options are available:

- Validate: Validates the current flow and displays error hints on Nodes.
- Format: Arranges the Nodes with the Network Simplex Algorithm
- Discard: Resets the Flow to the currently persisted State
- Cancel: Returns to the Regular Flow Display without persisting changes
- Save: Verifies the validity and persists the changes.
- Remove Business Rule: The Trash icon on the Nodes will remove the Business Rule
- Remove Connection: The X icon on the connection will remove This specific Connection
- Add Connection: Drag and Drop a Line from the Valid/Invalid Handle of one Node to the Target Handle of another Node
- Add Business Rule: Drag and Drop a Connection from the Valid/ Invalid Handle onto an Empty space will create a new Node with a default name. Its behavior can be updated once the Flow was saved.Alternatively CTRL+Click on an Empty space in the Canvas.
- Rename a Business Rule: Double click on the Label of a Node and enter the new Name. Pressing Enter or leaving the Node will end the renaming.
- Mark Multiple Business Rules: Hold down SHIFT and draw an rectangle over the nodes to be market. Alternatively, hold CTRL while clicking on a Business Rule.
- Copy & Paste: Marked Business Rules can be copy and pasted between Rulebooks of the same Knowledge Base.

3.15.4.1.1 Indicators

- Node is red: There is an issue with this Node either it is a part of a Cycle, it is part of multiple Root Nodes or has a duplicate Name
  - Solve Cycles: Check for Business Rule Connections which would result in loops and remove those

  - Solve multiple Roots: There should be only one Root Business Rule. Please have Connections to all Business Rules Nodes except for the starting Node.
  - Solve duplicate Name: Rename at least one of those Business Rules to have a unique name within the Knowledge Base.

- Orange left Border: This Business Rule has changes which need to be Saved.
- Gray left Border: This is a newly added Business Rule.
- Blue bottom Border: This Business Rule is currently part of a selection.

3.15.5 Rulebook Info Panel

On the far right-hand side, you'll discover the Info Panel, which displays information related to the selected Rulebook in the Rulebook Browsing Panel.

It is split into two sections, Details and Business Rules.

The Details section contains the description of the Rulebook.

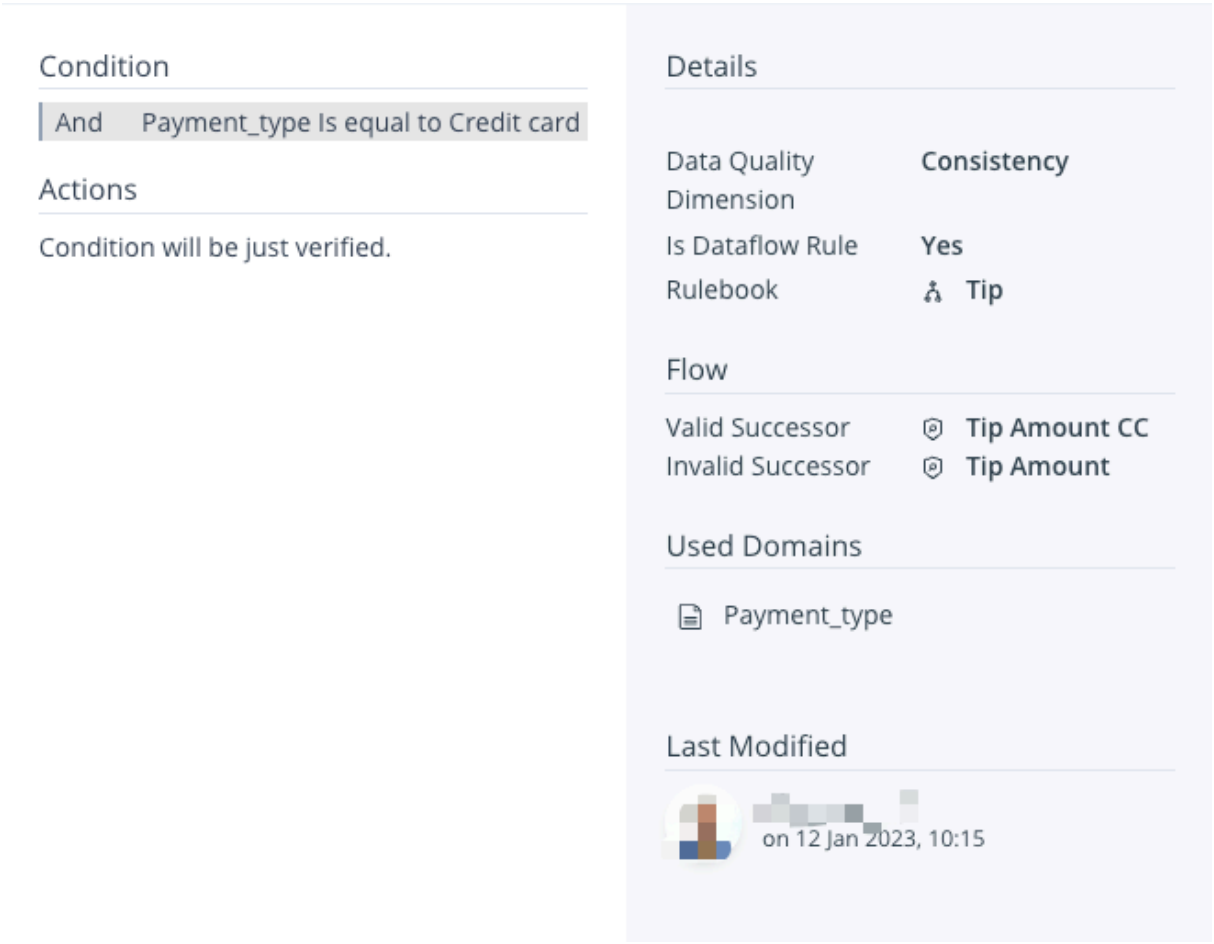
In the top-right corner, the 'Star' icon allows the user to add the respective Rulebook to Favorites. If the Rulebook has a Tag assigned, the user will be able to see the assigned Tag, just below.

The Business Rules section features a list of all the Business Rules on the selected Rulebook canvas.

Clicking any of the Business Rules, will navigate you to a menu containing information about the selected Business Rule, including:

- **Preparation**: presents the formula for the configured preparation steps.
- **Conditions**: presents the Business Rule Conditions and operators in a more condensed format.
- **Actions**: presents the Actions that have been configured to execute whenever the value within the Domain satisfies the Condition of the Business Rule.
- **Details**: presents the description of the Business Rule, the Data Quality Dimension it pertains to, and the Group to which the Business Rule is linked.
- **Flow**: displays the parent rules on the Rulebook canvas.
- **Used Domains**: enumerates all the Domains that were utilized to create Conditions within the Business Rule.
- **Last Modified**: shows information about the user who last modified the Business Rule, along with the date on which the modification took place.

In the top-right corner, the 'Star' icon allows the user to add the respective Business Rule to Favorites.



You can always click on the Rulebook in the list on the left, to go back to the Rulebook overview.

3.15.5.1 Edit/Delete Rulebook In order to edit and/or delete a Rulebook in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once there, find an click the Rulebook tab. Select the Rulebook you wish to edit or delete. In the top-right corner of the Info Panel, notice the `Edit` button. Click to open the `Edit Rulebook` drawer, which allows you to edit the selected Rulebook in a form similar to the `Add Rulebook` form or delete the Rulebook by using the `Delete` button at the bottom of the form.

3.16 Dataset Rules Page

This page can be accessed from the Knowledge Base navigation bar, by clicking the `Dataset Rules` tab.

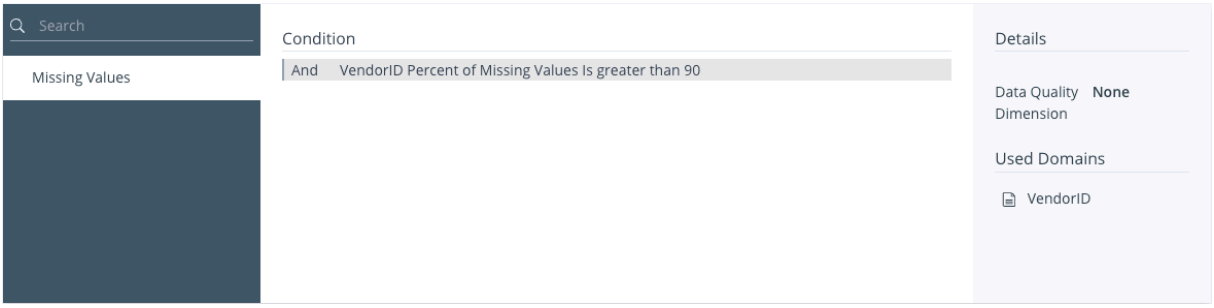


Figure 69: Dataset Rules

The page is divided into 3 separate panels:

- Dataset Rules Browsing Panel
- Dataset Rule Overview
- Dataset Rule Info Panel

We will take a closer look at each of these sections.

3.16.1 Dataset Rules Browsing Panel

Upon landing on the page, you can see the panel to the left that contains the list of already created Dataset Rules. Positioned above this list is a search input facilitating the quick retrieval of Dataset Rules with names that match the input character/s.

In the Edit Version mode, above of the search input, the `Add Dataset Rule` button is located.

3.16.1.1 Add Dataset Rule In order to add a Dataset Rule in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once there, find the `Dataset Rules` tab and click on it to be navigated to the Dataset Rules page. Locate and click the `Add Dataset Rule` at the top of the Dataset Rules Browsing Panel.

This will open the `Add Dataset Rule` drawer allowing you to configure and create a new Dataset Rule.

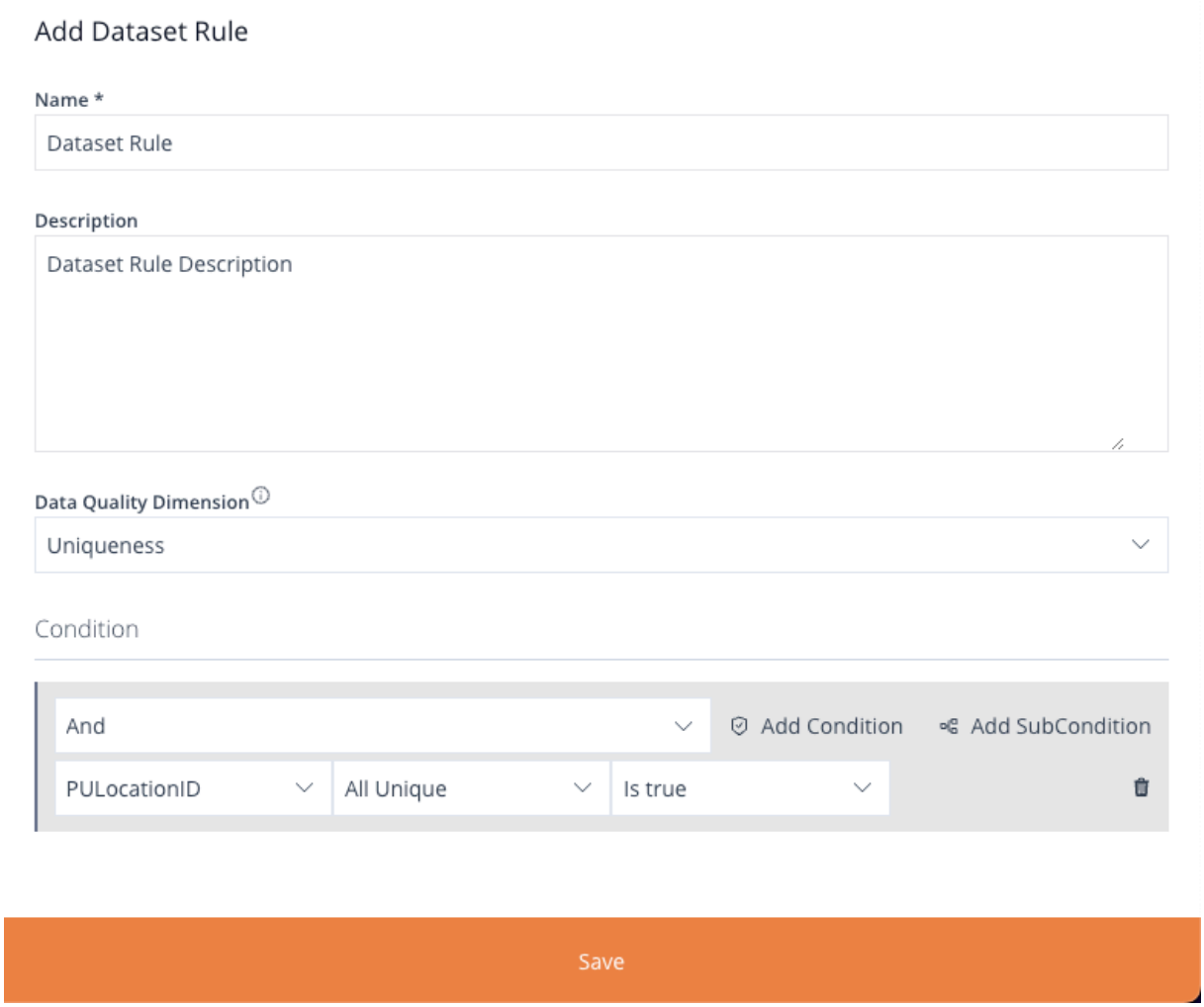


Figure 70: Add Dataset Rule

The form inside the drawer is split into 2 sections:

- Dataset Rule metadata
- Conditions

3.16.1.1.1 Metadata

This section contains the basic information about the Dataset Rule:

- **Name**: the name of the Dataset Rule.
- **Description**: the description of the Dataset Rule.
- **Data Quality Dimension** (dropdown) - will offer the option to select one of the 6 Data Quality Dimensions.

  The following Data Quality Dimensions are available:

- **Completeness**: Measures whether all required data is present and accounted for.
- **Uniqueness**: Ensures that each record or data point is distinct and not duplicated.
- **Timeliness**: Assesses whether data is up-to-date and available when needed.
- **Validity**: Checks if data conforms to the defined formats and Business Rules.
- **Accuracy**: Verifies that data correctly represents the real-world values it is intended to model.
- **Consistency**: Ensures that data is uniform and free from contradictions.
- **None**

3.16.1.1.2 Conditions The Condition section is quite similar to the one we have in the Business Rules form, with an additional field being the difference.

The additional field allows the values within the Domain to be queried according to specific Conditions. These being:

- Number of Distinct Values
- All Unique
- Maximum String Length
- Minimum String Length
- Median String Length
- Mean String Length
- Number of Missing Values
- Percent of Missing Values
- Number of Valid Values

These are then checked against default Business Rule Conditions.

3.16.2 Dataset Rule Overview

In the middle section, you will find an overview of the Conditions formulated within the chosen Dataset Rule. Each entry within this list features the corresponding operator and Condition.

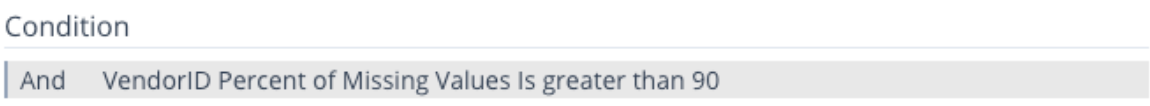


Figure 71: Dataset Rule Overview

3.17 Dataset Rule Info Panel

On the right-hand side, the Info Panel hosts the Details and Used Domains sections.

The Details section encompasses the description of the Dataset Rule and the associated Data Quality Dimension. In the top-right corner, the 'Star' icon allows the user to add the respective Dataset Rule to Favorites.

The Used Domains section highlights the Domain name on which the Condition was established.

Clicking any of the Domains in the list, will navigate you to the respective Domain on the Domains page.

3.17.1 Edit/Delete Dataset Rule

In order to edit or delete a Dataset Rule in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once there, find an click the Dataset Rule tab. Select the Dataset Rule you wish to edit or delete. In the top-right corner of the Info Panel, notice the `Edit` button. Click to open the `Edit Dataset Rule` drawer and form, which allows you to edit the selected Dataset Rule in a form similar to the `Add Dataset Rule` form, or delete the Dataset Rule by using the `Delete` button at the bottom of the form.

3.18 Data Links Page

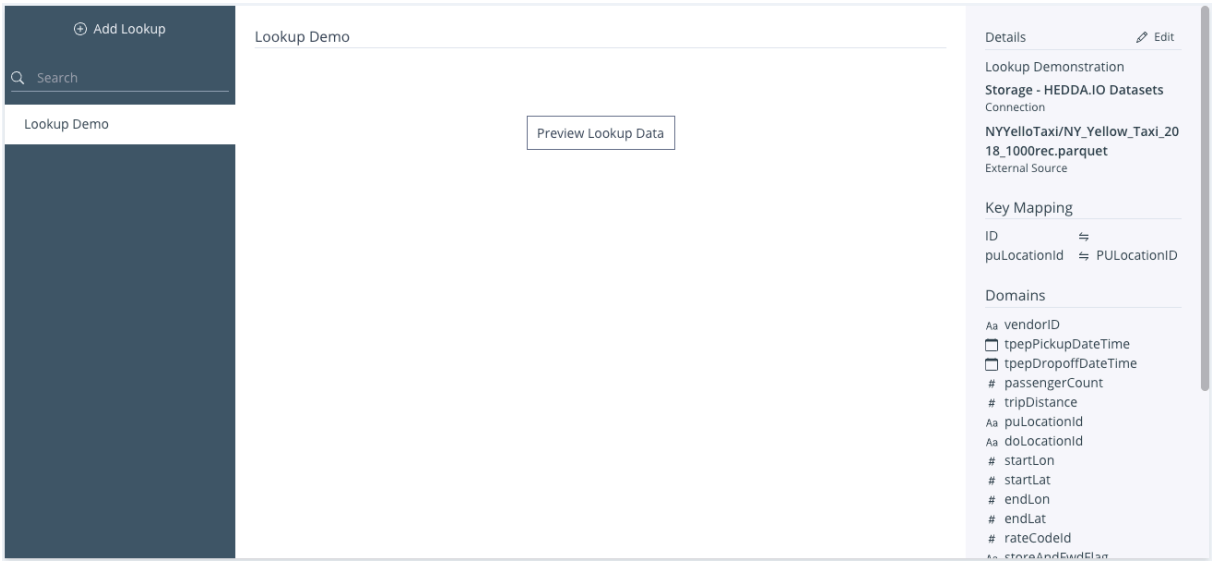


Figure 72: Data Links

The page is divided into three separate panels:

- Data Links Browsing Panel
- Data Links Overview
- Data Links Info Panel

Let's take a closer look at each of these sections.

3.18.1 Data Links Browsing Panel

Similar to previous pages, upon navigation to the Data Links page, the left-hand panel presents a list of Data Links with a search input just above it.

In the Edit Version mode, above of the search input, the `Add Data Link` button is located.

3.18.1.1 Add Data Link In order to add a Data Link in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once there, click the `Data Links` tab in the navigation bar to navigate to the Data Links page.

In the Browsing Panel on the left, click the `Add Data Link` in the top section. This will open the `Add Data Link` drawer containing a form that will allow you to configure a Data Link using the a predefined External Connection.

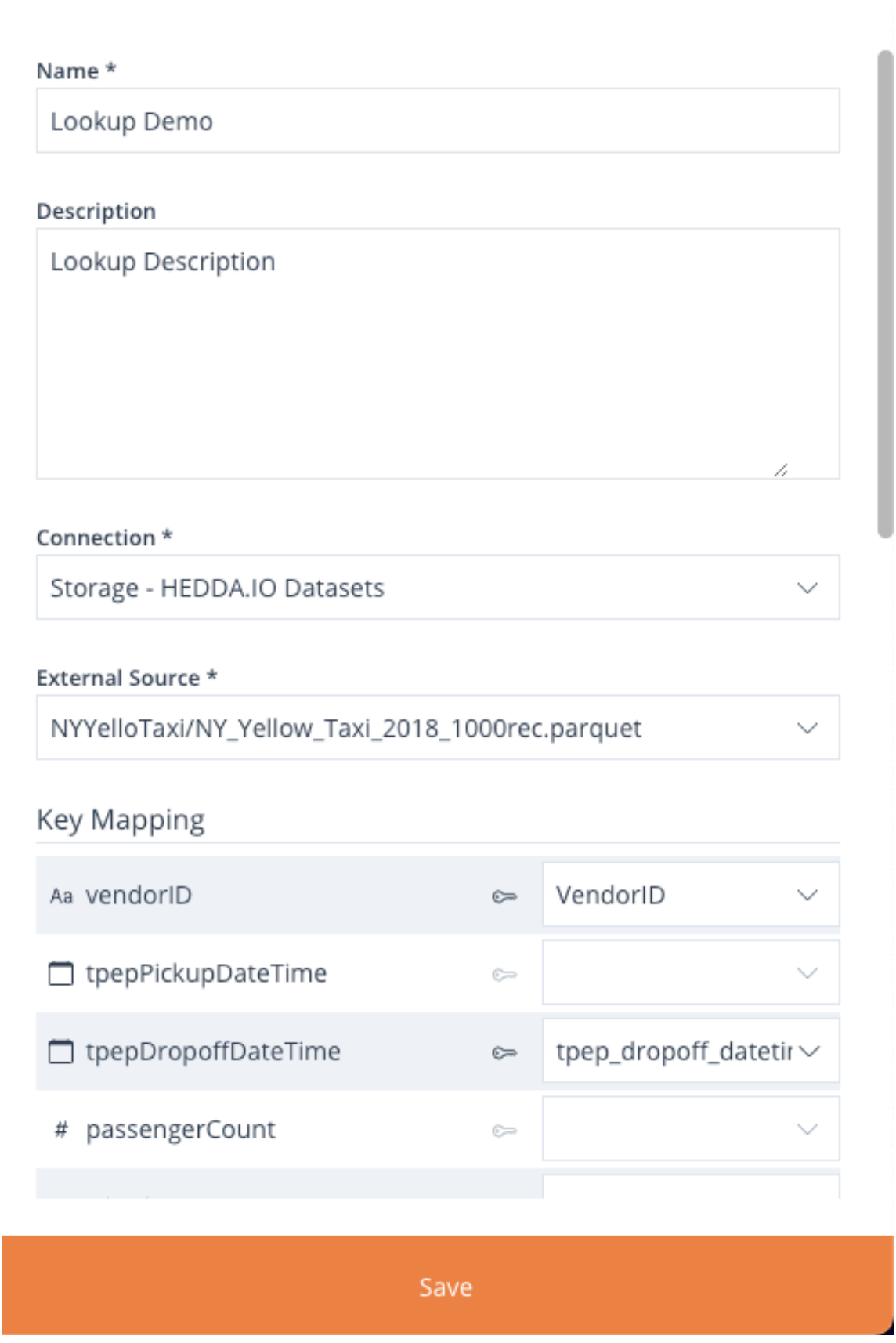


Figure 73: Data Links Browsing Panel

The form is split into two sections, Data Link metadata and Key Mapping.

Metadata

The metadata section contains the following input fields:

- **Name**: the name of the Data Link.
- **Description**: the description of the Data Link.
- **Connection**: select the External Connection from the dropdown.
- **External Source**: from the dropdown, select the path to the dataset that you want to use as a Data Link.

Key Mapping

Once the dataset is selected in the External Source dropdown, it will populate the Key Mapping section according to the name of all the columns in the dataset.

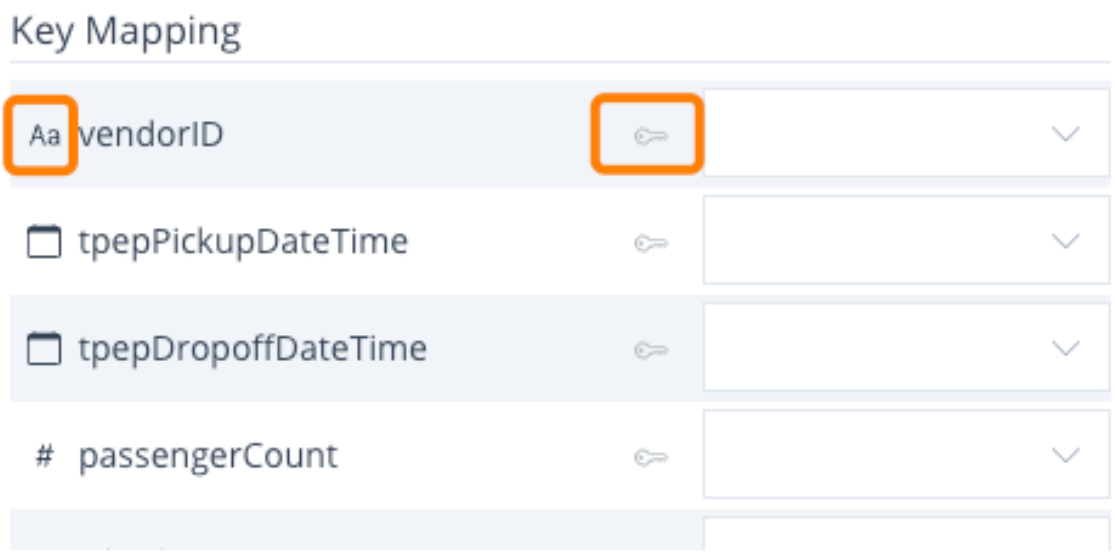


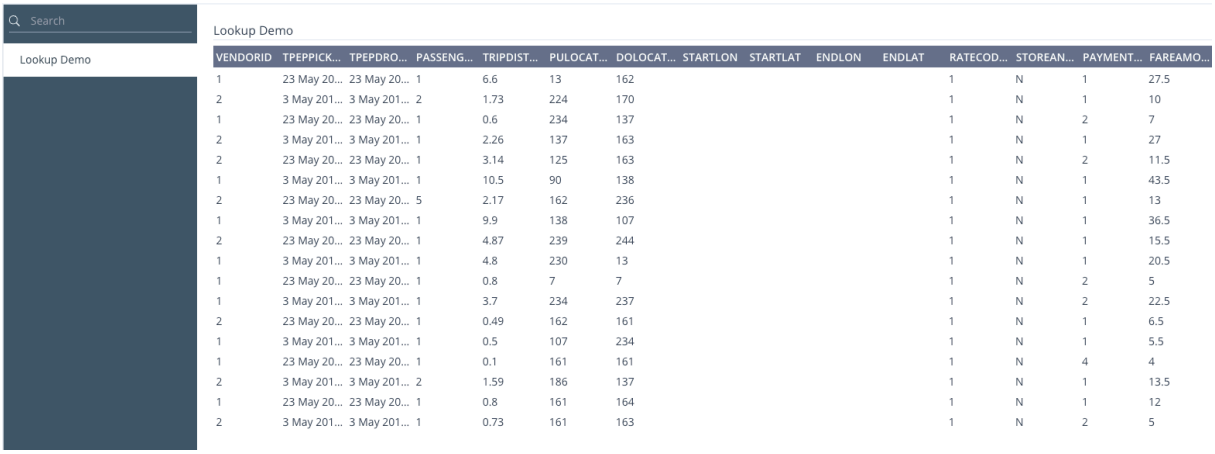
Figure 74: Key Mapping

Each item in the Key Mapping section has an icon representing the datatype on the left, a Key icon in the middle, which upon being clicked, will mark the column as being an identifier or a key, at which point, the dropdown field on the right will become available, allowing you to map it to an existing Domain.

Once all is configured, click the Save button to save the Data Link and close the drawer.

3.18.2 Data Links Overview

In the center, the Data Link data preview includes the `Preview Data Link Data` button, which fetches data from the dataset that was selected in the configuration of the Data Link.



VENDORID	TPEPPICK...	TPEPDRO...	PASSENG...	TRIPDIST...	PULOCAT...	DOLOCAT...	STARTLON	STARTLAT	ENDLON	ENDLAT	RATECOD...	STOREAN...	PAYMENT...	FAREAMO...
1	23 May 20...	23 May 20...	1	6.6	13	162					1	N	1	27.5
2	3 May 201...	3 May 201...	2	1.73	224	170					1	N	1	10
1	23 May 20...	23 May 20...	1	0.6	234	137					1	N	2	7
2	3 May 201...	3 May 201...	1	2.26	137	163					1	N	1	27
2	23 May 20...	23 May 20...	1	3.14	125	163					1	N	2	11.5
1	3 May 201...	3 May 201...	1	10.5	90	138					1	N	1	43.5
2	23 May 20...	23 May 20...	5	2.17	162	236					1	N	1	13
1	3 May 201...	3 May 201...	1	9.9	138	107					1	N	1	36.5
2	23 May 20...	23 May 20...	1	4.87	239	244					1	N	1	15.5
1	3 May 201...	3 May 201...	1	4.8	230	13					1	N	1	20.5
1	23 May 20...	23 May 20...	1	0.8	7	7					1	N	2	5
1	3 May 201...	3 May 201...	1	3.7	234	237					1	N	2	22.5
2	23 May 20...	23 May 20...	1	0.49	162	161					1	N	1	6.5
1	3 May 201...	3 May 201...	1	0.5	107	234					1	N	1	5.5
1	23 May 20...	23 May 20...	1	0.1	161	161					1	N	4	4
2	3 May 201...	3 May 201...	2	1.59	186	137					1	N	1	13.5
1	23 May 20...	23 May 20...	1	0.8	161	164					1	N	1	12
2	3 May 201...	3 May 201...	1	0.73	161	163					1	N	2	5

Figure 75: Data Links Overview

3.18.3 Data Links Info Panel

On the right side, we have the Info Panel, which is split into 3 sections:

- Details
- Key Mapping
- Domains

The details section encompasses the description, External Connection, and Data Link dataset name. In the top-right corner, the 'Star' icon allows the user to add the respective Data Link to Favorites.

The Key Mapping section comprises the columns used as keys as well as the Domain associated with it.

The Domains section lists all columns within the Data Link dataset.

3.18.3.1 Edit/Delete Data Link In order to edit or delete a Data Link in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once there, find an click the Data Links tab. Select the Data Link you wish to edit or delete. In the top-right corner of the Info Panel, notice the `Edit` button. Click to open the `Edit Data Link` drawer containing a form that allows you to edit the selected Data Link in a form similar to the `Add Data Link` form, or delete the Data Link by using the `Delete` button at the bottom of the form.

3.19 Runs Page

This is the section where you can set up a Run, which is necessary when attempting to execute a dataset against a Knowledge Base.

The page is divided into 3 separate panels:

- Runs Browsing Panel
- Runs Overview
- Runs Info Panel

Figure 76: Run Overview

3.19.1 Runs Browser Panel

On the left-hand side column, we have the Browsing Panel containing the list of previously created Runs with a search input that allows you to find the Runs based on the input character/s.

In the Edit Version mode, on top of the search input, the `Add Run` button is located.

3.19.1.1 Add Run In order to add a Run in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once there, click the `Runs` tab in the navigation bar to navigate to the Runs page.

At the top of the Browsing Panel is the `Add Run` button. Clicking on the button will open the `Add Run` drawer and the form.

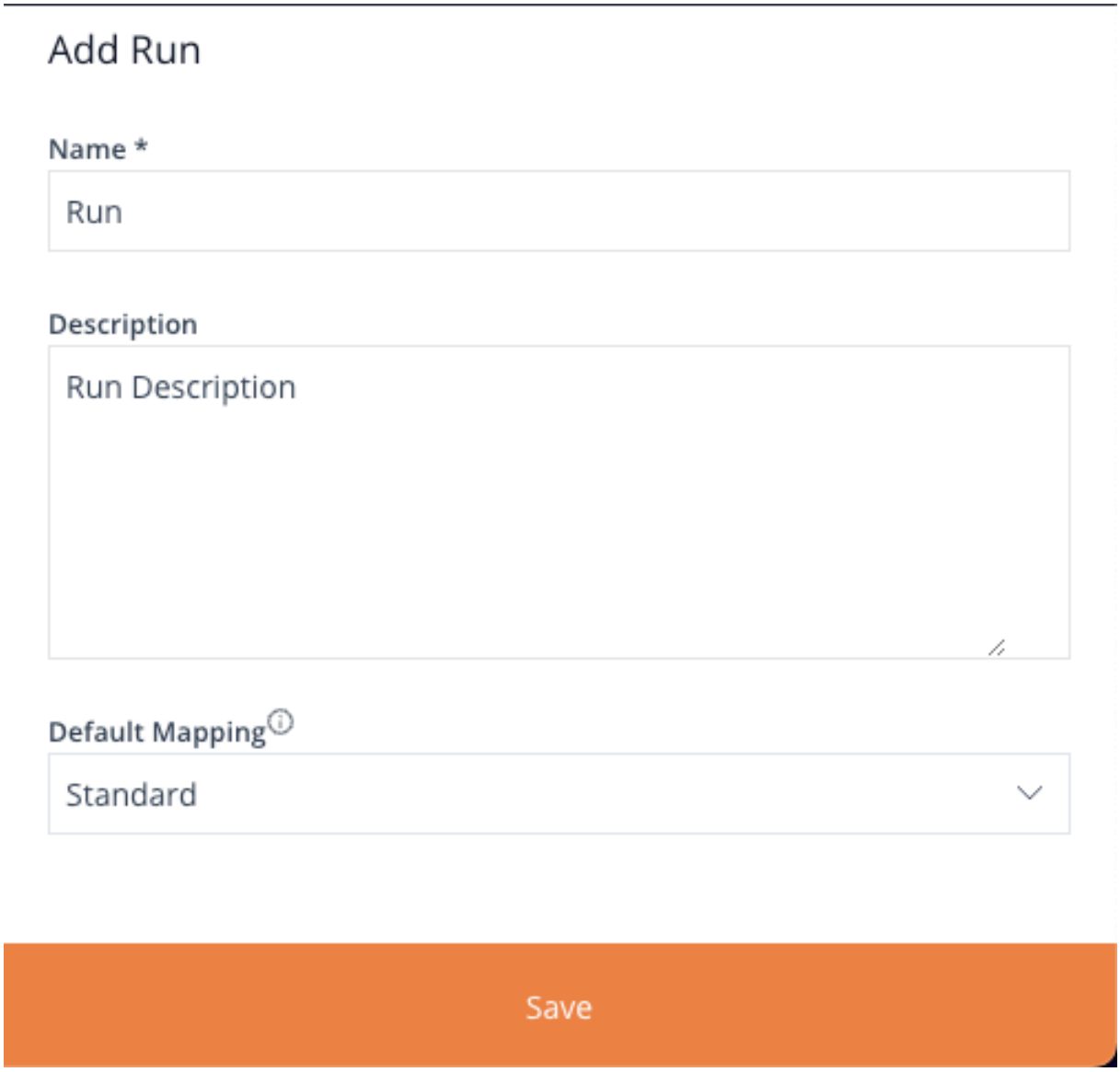


Figure 77: Add Run

This form has the following inputs:

1. The name of the Run.
2. The description of the Run.
3. Default Mapping

Once done, click “Save” on the bottom of the form.

3.19.2 Runs Overview

The Overview panel includes the statistics that we can find on the Knowledge Overview. However, while the Knowledge Base overview displays statistics for all Runs, here we have the same statistics but for each Run separately.

The Overview will present pertinent details concerning all Executions up until the current date.

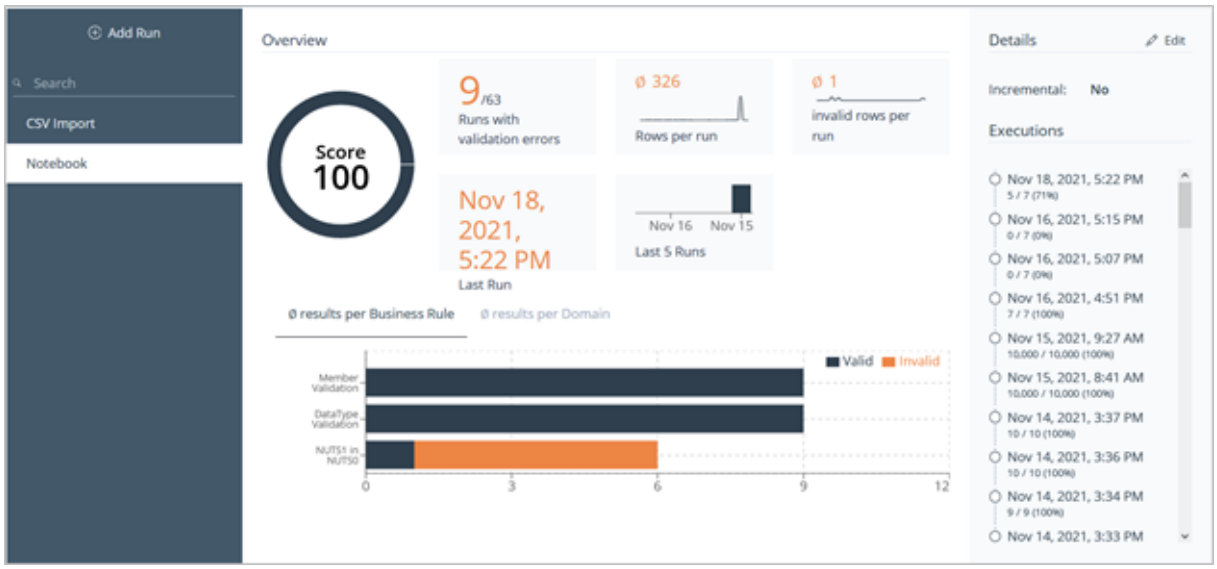
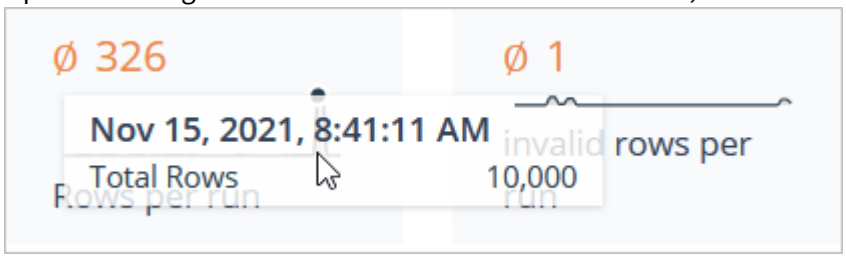


Figure 78: Run Overview

Here is an enumeration of the diverse statistics available in the overview:

- Score
- Runs with validation errors
- Rows per Run
- Invalid rows per Run
- Date of the last Run
- Last 5 Runs

Upon hovering over each of these visual elements, detailed information will be unveiled.



Towards the bottom of the overview, you'll encounter three tabs, each housing distinct sets of statistics:

- result per Business Rule
- result per Domain
- result per Dimension

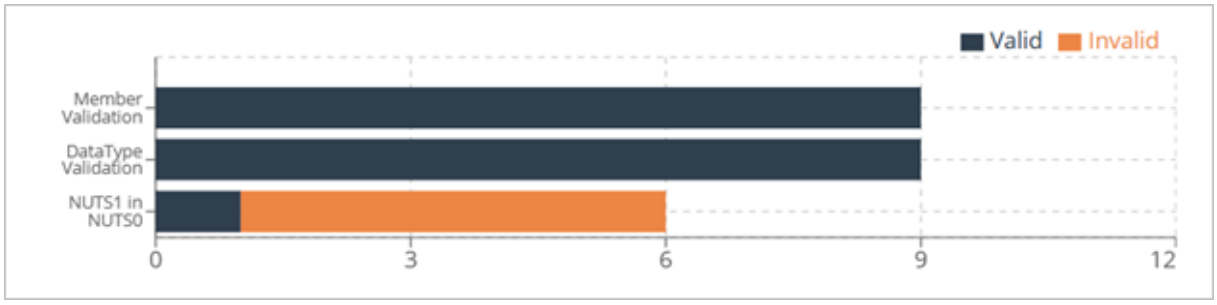


Figure 79: Result per Business Rule

3.19.2.1 Result per Business Rule The Legend clarifies by colors used to denote valid and invalid rows. Subsequently, this section showcases each of the Business Rules utilized in each Execution, presented in an elegantly designed horizontal column diagram.

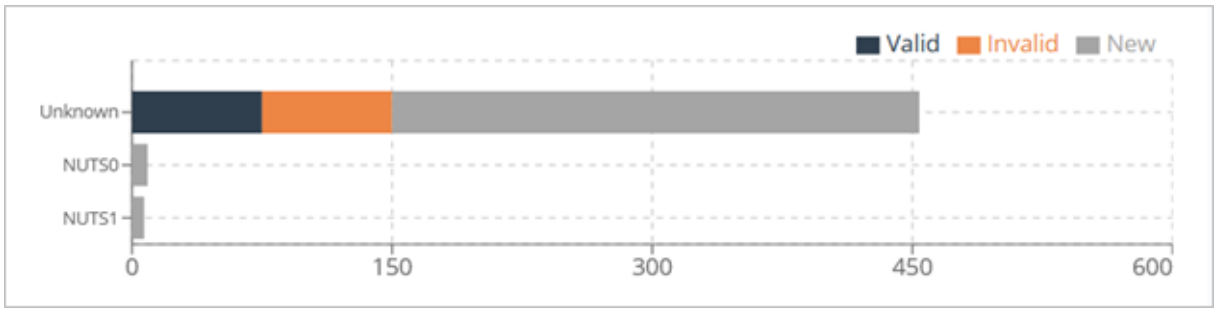


Figure 80: Results per Domain

3.19.2.2 Result per Domain The Legend clarifies by colors used to denote valid and invalid entries.

The diagram will sequentially present statistics for columns from the dataset that align with Domains in the Knowledge Base.

3.19.2.3 Result per Dimension This tab will exhibit statistics derived from Data Quality Dimensions, provided any were chosen during the configuration of Business Rules.

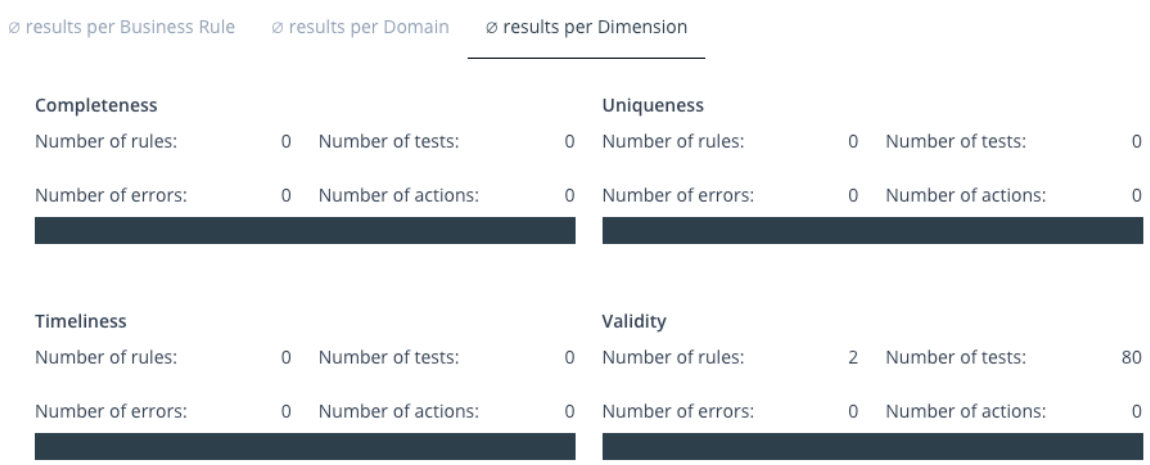


Figure 81: Result per Dimension

In addition to the count of Business Rules, errors and Actions associated with each Dimension, it will also present the tally of valid and invalid rows per Business Rules. This representation will be depicted by a bar graph, wherein the orange color denotes valid rules and the blue color signifies invalid ones.

3.19.3 Runs Info Panel

Both the Details and the Executions sections are hosted by the Info Panel located on the right-hand side of the page.

The panel is split into three distinct sections:

- Details
- Tags
- Executions

Details

The Details section showcases the metadata for the selected Run. In the top-right corner, the 'Star' icon allows the user to add the respective Run to Favorites. Here, the user can also see the description and the name of the Mapping configured for the selected Run.

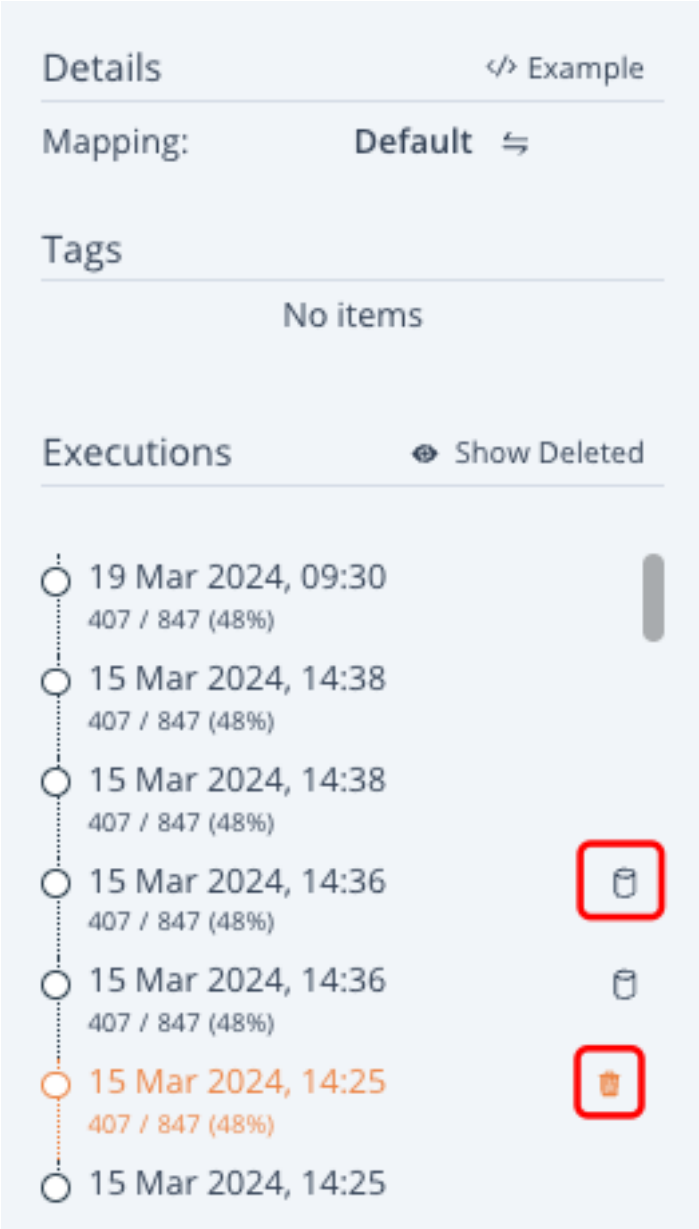


Figure 82: Runs

Furthermore, the `Example` button positioned in the top-right corner furnishes a comprehensive example of how to execute a dataset against the current HEDDA.IO Knowledge Base using the selected Run. The example contains the code for both Python and .NET HEDDA.IO libraries.

Python C#

```python
from pyheddaio.workflow import Hedda

ApiUrl = "<Api Url>"
ApiKey = "<Your API Key>"
df = <dataFrame>

# Create a HEDDA.IO client
hedda = Hedda.create(base_url=ApiUrl, api_key=ApiKey)

# Configure project, knowledge base, domain mappings, etc.
hedda.use_project("New York Taxi Testing")
hedda.use_knowledge_base("Yellow Taxi Trip Testing")
hedda.use_profiling()
hedda.enable_data_upload()

# Pass data to the client
hedda.use_data(df)

# Use this run to store the results in
hedda.use_run("Standard")
result = hedda.start()
```

Figure 83: Python Run Example

Clicking on the Mapping will navigate you to the respective Mapping.

Tags

The Tags section will display the Tags configured for the selected Run.

3.19.3.1 Executions

The Executions section is located below the Tags section in the Info Panel.

Each item in the list of Executions will have a Date and Time, which represents the moment in time it was executed, as well as the number of valid records / number of total records.

On the right side of each Execution you will find the `delete` button, represented by a trashcan icon.

If the execution was performed using the Preview feature, it will have a `has data` icon, represented by a databank icon. This will open the Execution in Preview mode.

Clicking on any of these Executions in the list, will open a drawer that will show advanced info about the Execution.

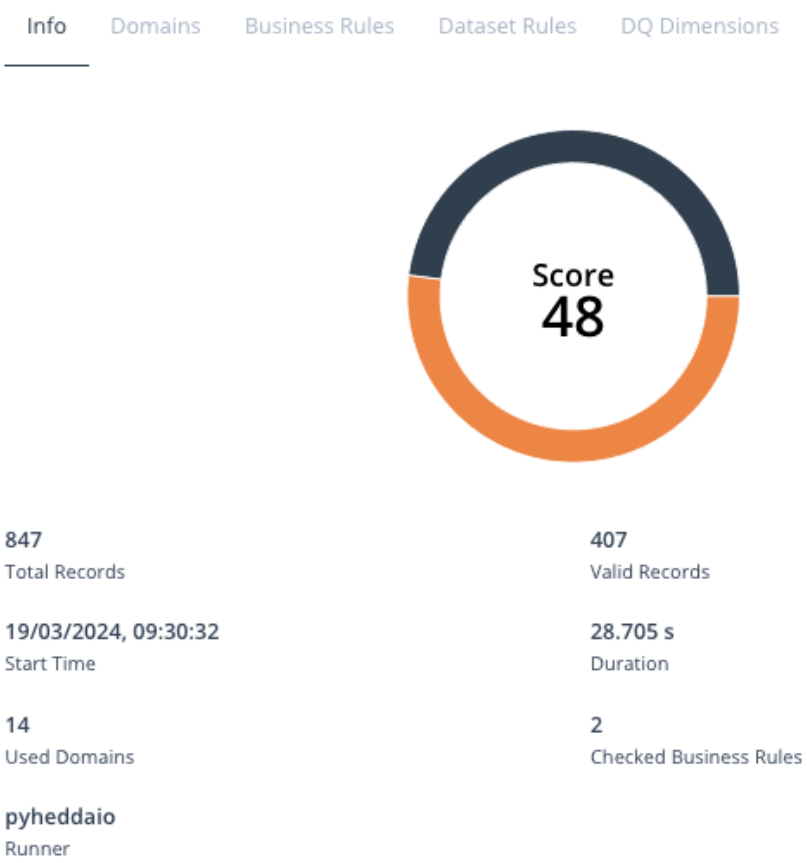


Figure 84: Executions

The drawer presents you with 6 tabs:

- Info
- Domain
- Business Rules
- Dataset Rules
- DQ Dimensions
- Member
- Profiling

Each containing comprehensive statistics about the Execution.

3.19.3.1.1 Info

This tab functions as an Execution overview.

Here is the type of information available on the info tab:

- Score
- Total Records
- Valid Records
- Start Time
- Duration
- Used Domains
- Checked Business Rules
- Runner

3.19.3.1.2 Domain Statistics

Here you can find statistics about the Domains used in the Run.

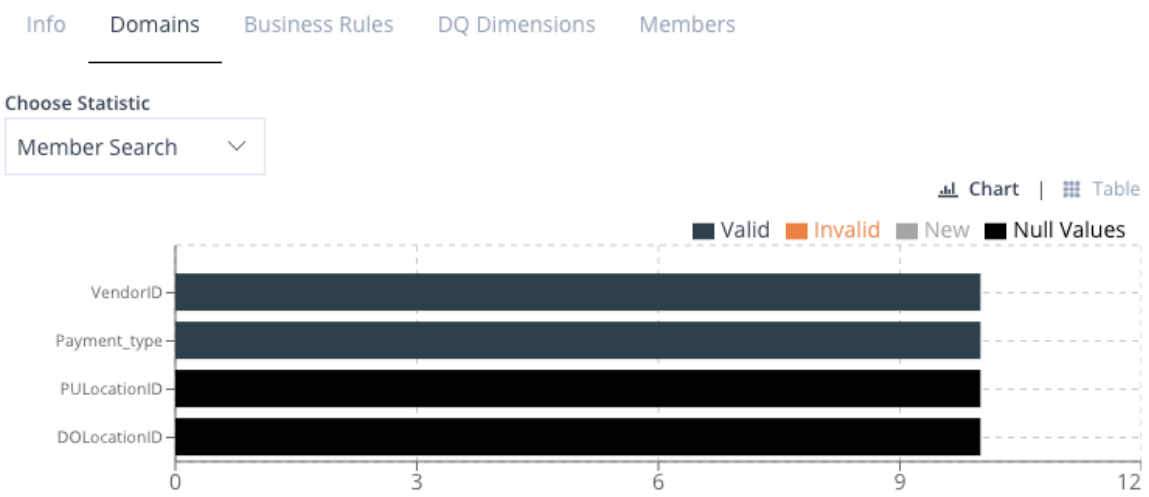


Figure 85: Domain Statistics

At the top, you have the option to view statistics for Members, Data Type Validation, or Value Insights.

You can pick between seeing the statistics as a chat or in a table format on this tab. Just click the buttons on the right side, above the legend, to switch between them.

Figure 86: Domain Statistics

The Legend is there to assist in understanding the colors used to indicate Valid, Invalid, New, or Null Values when in Chart mode.

3.19.3.1.3 Business Rule Statistics Here are the statistics for every Rulebook and subsequently every Business Rule used in the respective Run.

On the top you will find two dropdowns that will filter the statistics based on different criteria:

- Rulebook
- Exclude Dataflow Rules

The `Rulebook` dropdown will allow you to see statistics per Rulebook, thus filtering the data based on the selected Rulebook.

Selecting `None` will only leave the Member and Data Type validation.

The `Exclude Dataflow Rules` dropdown will leave or exclude the Business Rules that were tagged as dataflow rules.

The Legend will help understand which color was used for the Valid, or Invalid values.

You can pick between seeing the statistics as a chart or in a table format on this tab. Just click the buttons on the right side, above the legend, to switch between them.

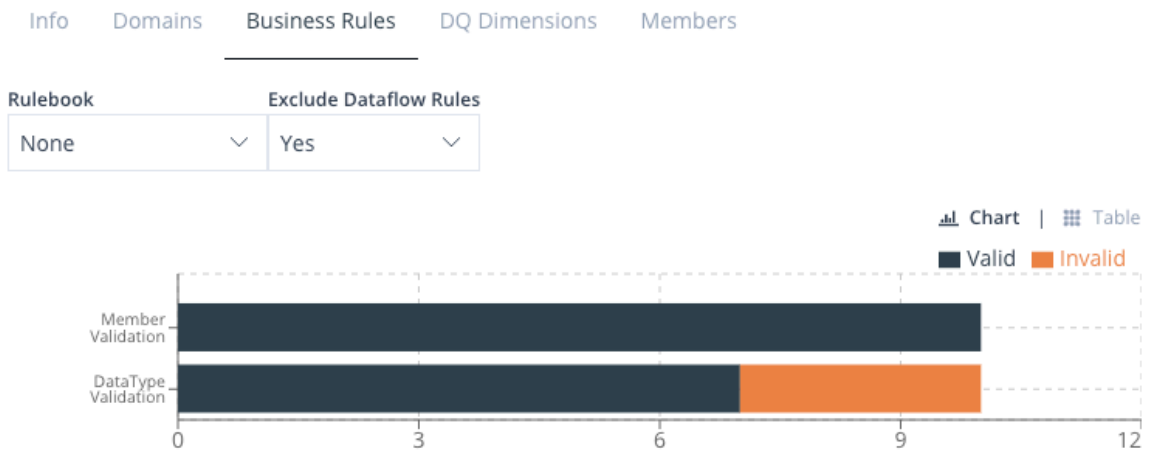


Figure 87: Business Rule Statistics

3.19.3.1.4 Dataset Rules Statistics This tab presets statistics centered around the Dataset Rules. For each Dataset Rule configured in the Knowledge Base, it will display the DQ Dimension it was assigned, and the Validity of the Dataset Rule.

Figure 88: Dataset Rules Statistics

3.19.3.1.5 Data Quality Dimensions Statistics This tab presents statistics centered around Data Quality Dimensions (DQ Dimentions). This entails that for each Dimension type, it showcases the count of Business Rules with configured dimensions, the number of tests conducted, the portion of those tests encountering errors, and the number of executed Actions.

Figure 89: DQ Dimensions Statistics

3.19.3.1.6 Member Statistics In this tab, statistics are compiled for all Members that are found, matched, and new.

At the top, you can opt to exclusively view Members associated with a particular Domain. Additionally, you have the ability to filter the statistical data according to the following criteria:

- New
- New with match
- New without match
- Matched
- All

The horizontal column diagram will present Members based on the selected filter.

You have the choice to view the statistics either as a chart or in a table format on this tab. Simply click the buttons situated on the right side, above the legend, to switch between these display formats.

The Legend exhibits a single color that signifies the Found items.

3.19.3.1.7 Profiling This statistics tab will only show if the Python HEDDA.IO was used in the Run.

This tab includes by far the most statistical data, however, this data is mostly insights into the structure, relationships, and overall characteristics of the dataset. It allows studying data for easy analysis.

The open-source tool pandas-profiling helps craft comprehensive data reports. Click on pandas-profiling to read more.

Reviewing the Profiling report allows you to identify data anomalies and comprehend the essential actions required for data cleansing and preparation. For instance, the ultimate dataset version will be free of duplicated information. This enables the identification of missing values, identical data entries, incorrect data types, or any other data-related issues. The report facilitates the task of fulfilling all the criteria on the data cleaning and preparation checklist.

Profiling statistics in HEDDA.IO includes statistics such as:

Component	Description
Domains	Which columns contain what
Correlations	Mean, average, etc.
Missing	Tells you about how your data is distributed
Alerts	Again, for visualizing distributions

Component	Description
Samples	Allows you to see what is related

At the very top of the Profiling panel, you can see data related to the Memory Size, Record Size, and Missing Cells.

Figure 90: Profiling

Let's take a closer look at the components.

Profiling - Domains

This part of the report provides an extensive examination of all the columns within the dataset.

The statistics displayed differ based on the data type of the Domain. The Domain can be selected from the dropdown located above the statistical data. This encompasses, for instance:

- Overview
  - Distinct values/percentage
  - Missing cells/percentage
  - Unique values/percentage
  - Memory Size
- Characters and Unicode
  - Total Characters
  - Distinct Characters
  - Distinct Categories
  - Distinct Script
  - Distinct Blocks
- Length
  - Max length
  - Min Length
  - Median Length
  - Mean Length

- Samples

Profiling - Correlations

Correlations refer to the statistical measure of the strength and direction of a relationship between two or more variables.

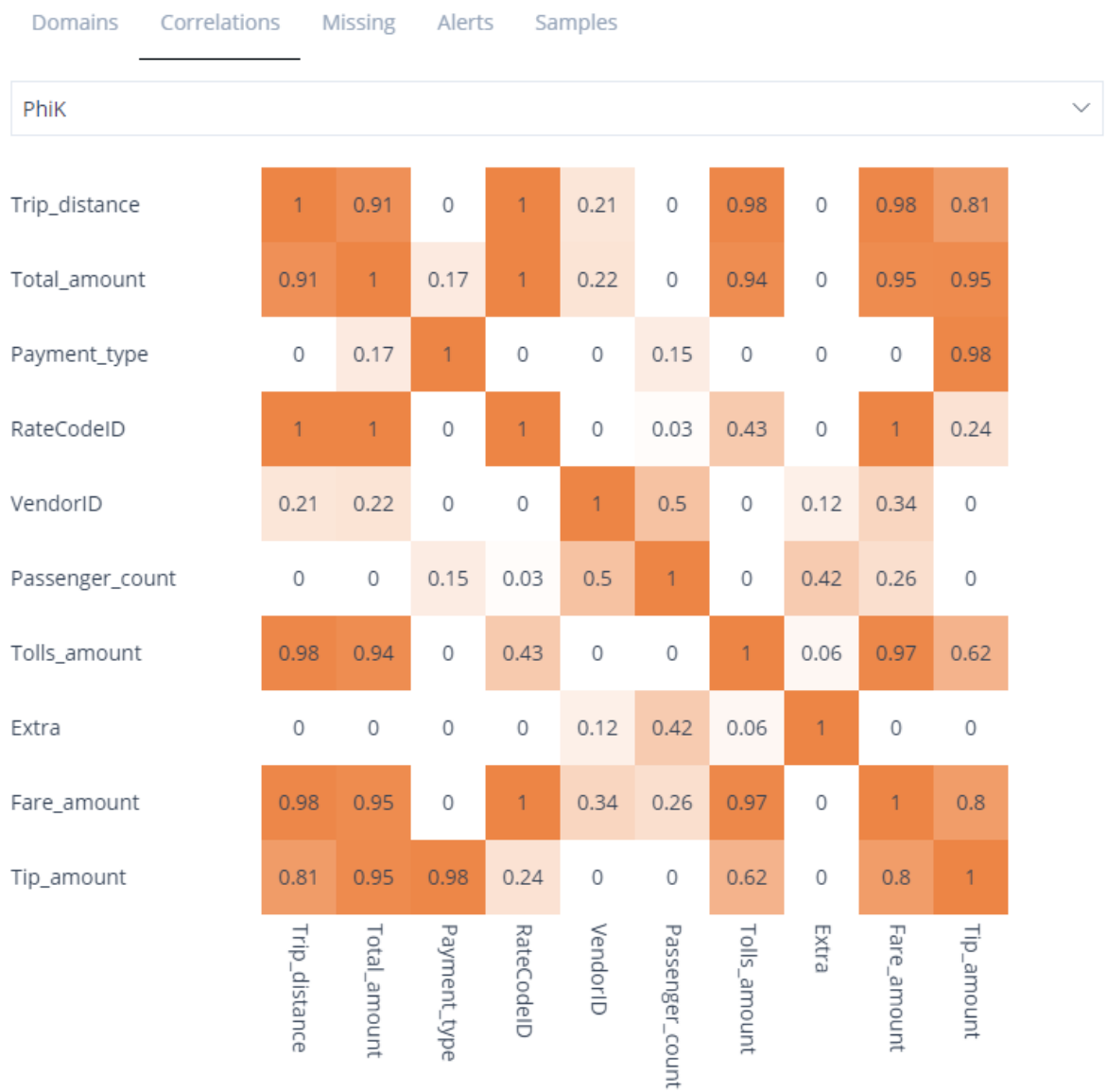


Figure 91: Correlations

A threshold of 1 signifies a strong correlation, while lower values indicate weaker correlations.

You have the option to select from various correlation types:

- PhiK

- Cramers
- Spearman
- Kendall
- Pearson

Profiling - Missing

This section displays visualizations of the missing values within the dataset.

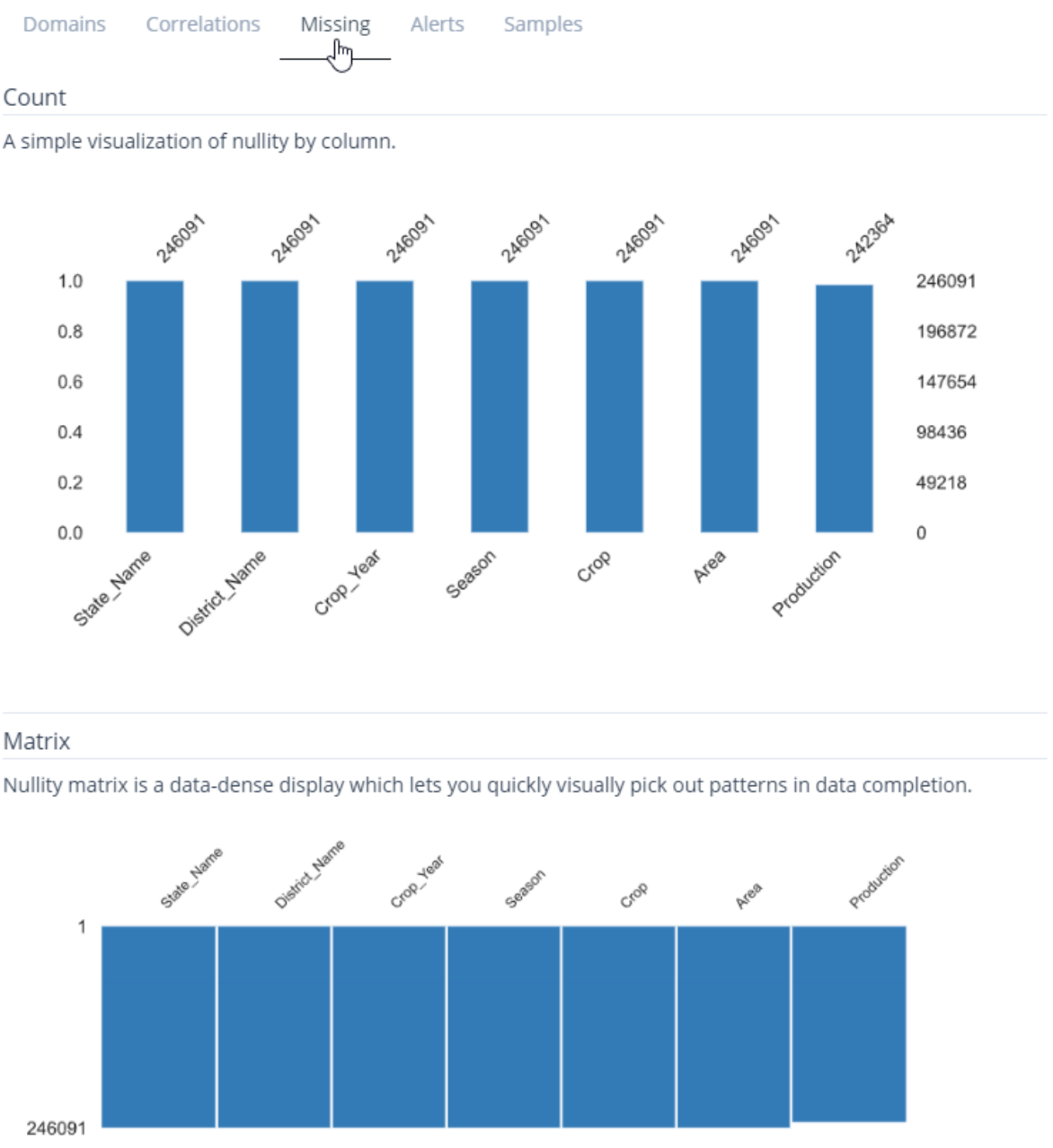


Figure 92: Missing

There are two types of charts: Count and matrix. The count chart is a simple bar chart where the x-axis represents the column names, and the y-axis indicates the number of values present (excluding null values), depicted by the length of the bars.

The same principle applies to the matrix chart.

Profiling - Alerts

The Alerts segment within the report provides an extensive compilation of potential data quality concerns. However, to confirm whether an alert indeed signifies a data quality problem, Domain validation is necessary.

The table provided below outlines all conceivable data quality alerts and their implications. Certain warnings pertain to individual columns, some relate to correlations between columns, and others encompass the entire dataset.

Alert	Description
Constant	Column only contains one value.
Zeros	Column only contains zeros.
High Correlation	Correlations (either Spearman, Cramer, Pearson, Kendall, PhiK) are above the warning threshold (configurable).
High Cardinality	Whether the column has more than 50 distinct values. Threshold is configurable.
Skewness	Column's univariate distribution presents skewness. Threshold value is configurable.
Missing Values	Column has missing values.
Infinite Values	Column has infinite values (either np.inf or -np.inf).
Unique Values	All values of the column are unique (count of unique values equals column's length).
Date	Column contains Date records.
Uniform	Column follows a uniform distribution (Chi-squared test score > 0.999, threshold score is configurable).
Constant length	For strings/date columns whose entries all have the same length.
Rejected	Variable has mixed types or is constant (thus not suitable for meaningful analysis).
Unsupported	Column can't be analyzed (type is not supported, has mixed types, has lists/dicts/tuples, is empty, wrongly formatted).
Duplicates	Dataset-level warning signaling the presence of more than 10 duplicated records.
Empty	Dataset-level warning signaling there is no data to be analyzed.

Profiling - Samples

This section displays the first and last ten rows of the dataset.

3.19.3.2 Edit/Delete Run In order to edit and/or delete a Run in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once there, find and click the Runs tab. Select the Run you wish to edit or delete. In the top-right corner of the Info Panel, notice the `Edit` button. Click to open the `Edit Run` drawer with the form allowing you to edit the selected Run in a form similar to the `Add Run` form, or delete the Run by using the `Delete` button at the bottom of the form.

3.20 Mappings Page

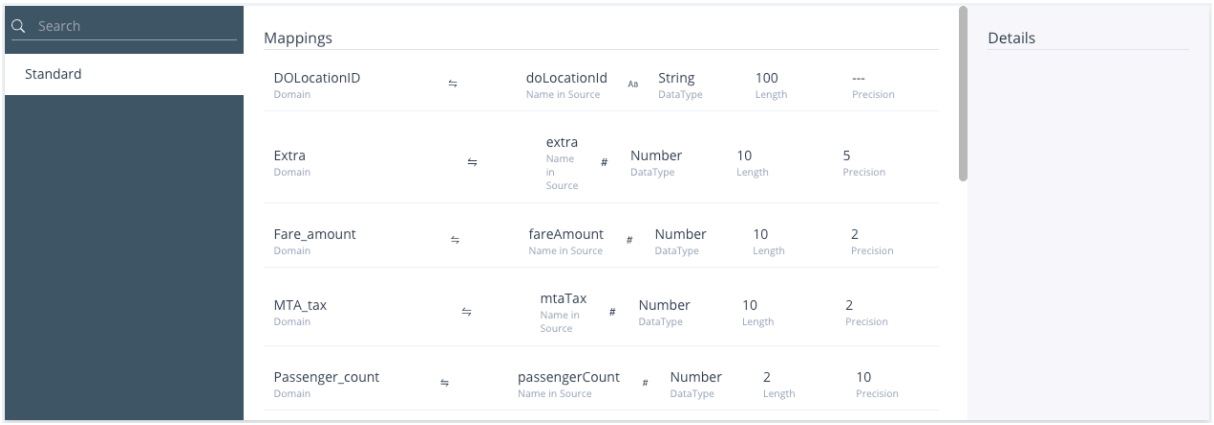


Figure 93: Mappings

The page is divided into 3 separate panels:

- Mapping Browsing Panel
- Mapping Overview
- Mapping Info Panel

3.20.1 Mapping Browsing Panel

Upon landing on the page, you'll observe the customary left-hand column containing the list of pre-established Mappings. Positioned above this list is a search input that will help you find Mappings whose names align with the input character/s.

In the Edit Version mode, on top of the search input, the `Add Mapping` button is located.

3.20.1.1 Add Mapping

In order to add a Mapping in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once there, click the `Mappings` tab in the navigation bar to navigate to the Mappings page.

Clicking the `Add Mapping` button located on top of the Browsing Panel will open the `Add Mapping` drawer. The drawer contains a form that will allow you to configure a Mapping to be used for the Run.

The form is split into two sections:

- Mapping Metadata
- Domain Mappings

Metadata

This section allows you to enter the name and the description of the Mapping.

Domain Mapping On the bottom side of the form is where we can configure our Mapping.

Clicking the `Copy Existing` will allow you to copy the Mapping configuration from an already existing Mapping.

Right under the “Select Domain” label, you have a list of the available Domains listed horizontally.

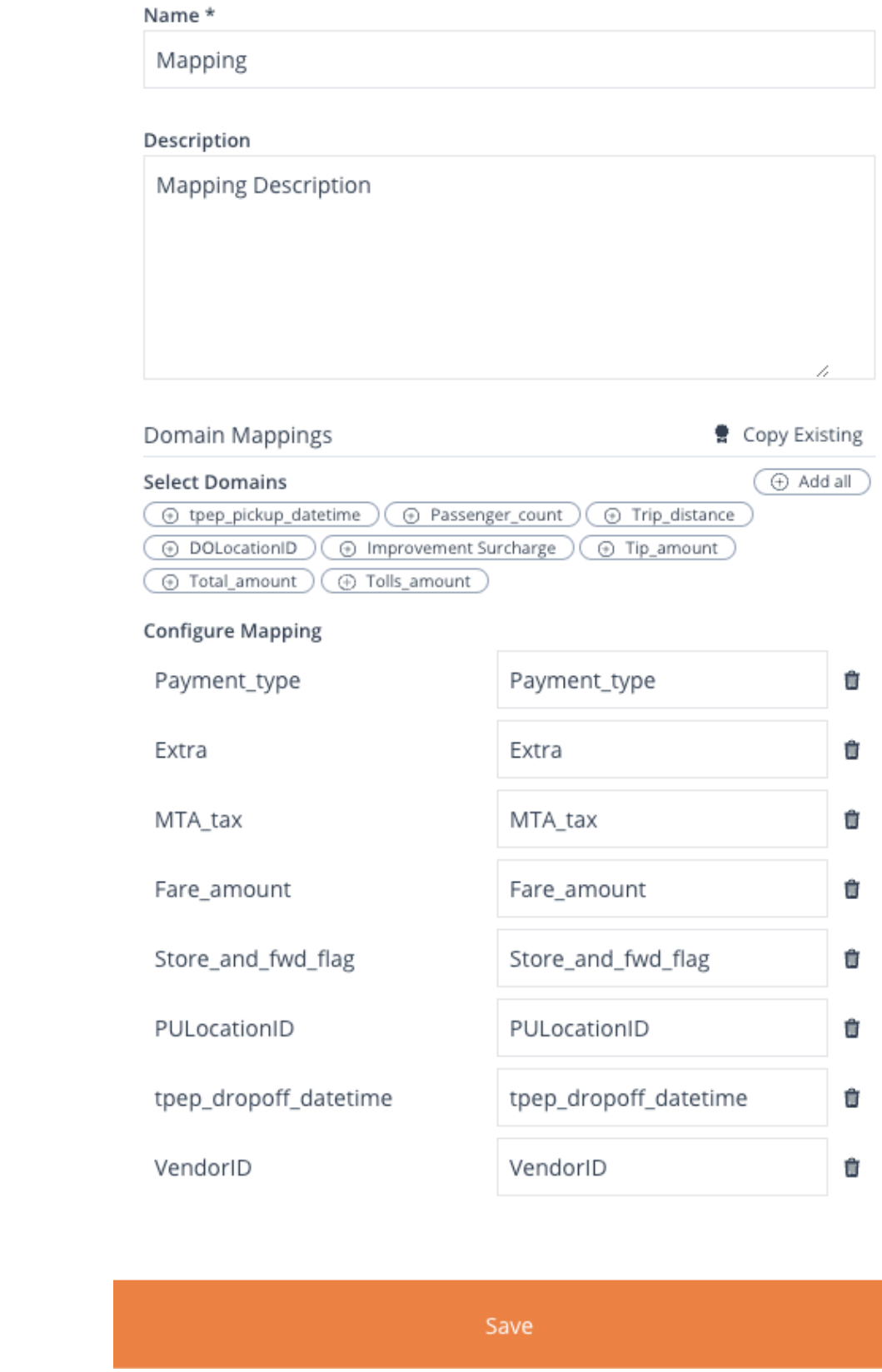


Figure 94: Domain Mapping

Clicking on each of them, will add the respective Domain to the `Configure Mapping` label.

Alternatively, you can click the `Add All` which will add all Domains.

If the column name in the dataset and the Domain in HEDDA.IO coincide, the Domains will require no editing. Alternatively, you will need to manually enter the name of the column associated with the selected Domain.

You can click the `trashcan` icon on the right of each item, in order to remove it from under the `Configure Mapping` label.

Once all is configured, click the `Save` button at the bottom of the form to save and close the drawer.

3.20.2 Mappings Overview

In the middle section you have the list of mapped Domains along with their respective data type, length and precision, if the case.

3.20.3 Mapping Info Panel

On the right-hand column, we have the details section containing the description of the Mapping. In the top-right corner, the 'Star' icon allows the user to add the respective Mapping to Favorites.

3.20.3.1 Edit/Delete Mapping In order to edit and/or delete a Mapping in HEDDA.IO, you must switch to the Edit Version mode, which is done by clicking the `Edit Version` button in the top-right corner of the page.

Once there, find an click the Mappings tab. Select the Mapping you wish to edit or delete. In the top-right corner of the Info Panel, notice the `Edit` button. Click to open the `Edit Mapping` drawer and the form allowing you to edit the selected Mapping in a form similar to the `Add Mapping` form, or delete the Mapping by using the `Delete` button at the bottom of the form.

3.20.4 Tags Page

Tags enable users to categorize Rulebooks and Runs by applying a user-defined Tag to them. It's important to note that a Run with an assigned Tag will validate not only Rulebooks with the same Tag but also Rulebooks without any Tag, as these are treated as general Rulebooks.

The overview panel will display the Rulebook/s and Run/s associated with the selected Tag. Additionally, it offers the possibility to unlink the Tag from a Rulebook or Run.

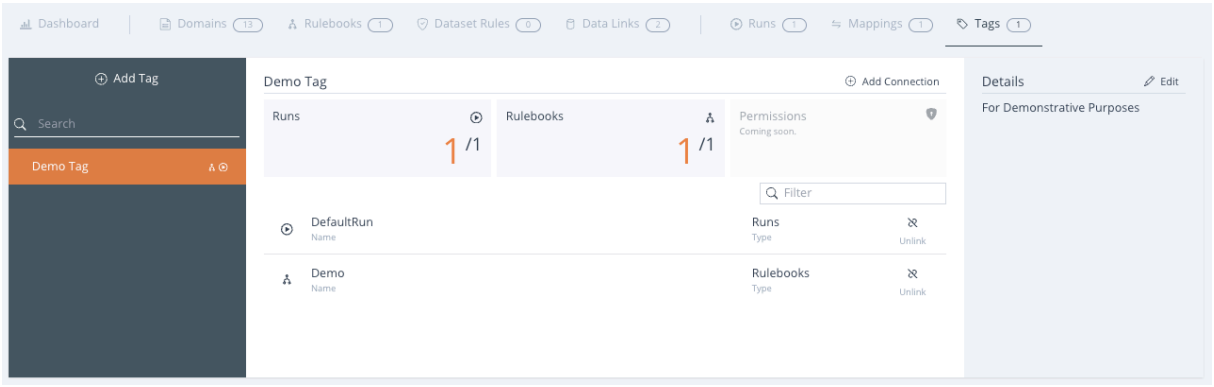


Figure 95: Tags Overview

3.20.4.1 Add Tag To add a Tag, go to the Tags tab, then click the `Add Tag` button in the top-left corner of the overview panel.

This will open the `Add Tag` drawer, with a form that has 2 fields: name and description.

Click `Save` to add the Tag.

3.20.4.2 Tag Info Panel The Details section will display all the Rulebooks and Runs the selected Tag was configured into. In the top-right corner, the 'Star' icon allows the user to add the respective Tag to Favorites.

3.20.4.3 Edit and/or Delete Tag To Edit or Delete a Tag, click the `Edit` button in the top-right corner of the page, in the Details section.

This will open a drawer with a `Edit Tag` form, allowing you to `Save` it, once modifications were made, or `Delete` it, with both buttons available at the bottom of the drawer.

4 Appendix

4.1 Algorithm Comparison

4.1.1 53639

Algorithm	5363,	53609	D-53639	65520
ColognePhonetic	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
Levenshtein	**0.80**	**0.80**	~~0.71~~	~~0.00~~

Algorithm	5363,	53609	D-53639	65520
DoubleMetaphone	**0.99**	**0.99**	~~0.75~~	**0.99**
Jaro	**0.87**	**0.87**	**0.90**	~~0.43~~
LongestCommonSubstring	**0.80**	~~0.60~~	~~0.00~~	~~0.00~~
Caverphone	**0.89**	**0.90**	**0.90**	~~0.70~~
Nysiis	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
Phonem	**0.80**	**0.80**	**0.83**	~~0.00~~
Phonetex	**0.99**	**0.99**	**0.86**	**0.86**
Phonex	**0.99**	**0.99**	**0.86**	**0.86**
Phonix	**0.99**	**0.99**	**0.86**	**0.86**
ExactMatchCaseInsensitive	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
ExactMatch	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
LevenshteinDamerau	**0.80**	**0.80**	~~0.71~~	~~0.00~~
KeyboardDistanceQwertz	**0.95**	**0.90**	~~0.56~~	**0.91**
KeyboardDistanceQwerty	**0.95**	**0.90**	~~0.55~~	**0.91**
KeyboardDistanceDvorak	**0.89**	**0.90**	~~0.56~~	**0.91**

4.1.2 Tillmann

Algorithm	Tillamnn	Tillman	Tilmann	Illmann	Iillmann	Tillmanb	Hilma	tillmann
ColognePhonetic	~~0.67~~	**0.99**	**0.99**	~~0.75~~	~~0.75~~	~~0.73~~	~~0.42~~	**0.99**
Levenshtein	~~0.75~~	**0.88**	**0.88**	~~0.75~~	**0.88**	**0.88**	~~0.50~~	**0.88**
DoubleMetaphone	**0.99**	**0.99**	**0.99**	~~0.75~~	~~0.75~~	**0.99**	~~0.50~~	**0.99**
Jaro	**0.96**	**0.96**	**0.96**	**0.87**	**0.92**	**0.92**	~~0.77~~	**0.92**
LongestCommonSubstring	~~0.50~~	**0.88**	~~0.38~~	~~0.00~~	~~0.00~~	**0.88**	~~0.00~~	~~0.00~~
Caverphone	**0.90**	**0.99**	**0.99**	**0.89**	**0.89**	**0.90**	~~0.78~~	**0.99**
Nysiis	**0.83**	**0.99**	**0.99**	~~0.67~~	~~0.67~~	**0.86**	~~0.50~~	**0.99**
Phonem	~~0.67~~	**0.99**	**0.99**	**0.83**	**0.83**	**0.86**	~~0.67~~	**0.99**
Phonetex	**0.99**	**0.99**	**0.99**	**0.90**	**0.90**	**0.91**	**0.80**	**0.99**

Algorithm	Tillamnn	Tillman	Tilmann	Illmann	Iillmann	Tillmanb	Hilma	tillmann
Phonex	**0.99**	**0.99**	**0.99**	**0.89**	**0.89**	**0.90**	**0.89**	**0.89**
Phonix	**0.99**	**0.99**	**0.99**	**0.89**	**0.89**	**0.90**	**0.89**	**0.99**
ExactMatchCaseInsensitive	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	**0.99**
ExactMatch	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
LevenshteinDamerau	**0.88**	**0.88**	**0.88**	~~0.75~~	**0.88**	**0.88**	~~0.50~~	**0.88**
KeyboardDistanceQwertz	**0.89**	**0.88**	~~0.75~~	~~0.71~~	**0.97**	**0.99**	~~0.53~~	**0.99**
KeyboardDistanceQwerty	**0.89**	**0.88**	~~0.75~~	~~0.71~~	**0.97**	**0.99**	~~0.53~~	**0.99**
KeyboardDistanceDvorak	**0.89**	**0.88**	~~0.71~~	~~0.63~~	**0.97**	**0.97**	~~0.52~~	**0.99**

4.1.3 Eitelberg

Algorithm	Eiteblerg	Eikelberg	Eichelberg	Heidelberg	Eickelbäck
ColognePhonetic	~~0.78~~	**0.83**	**0.83**	**0.99**	~~0.56~~
Levenshtein	~~0.78~~	**0.89**	**0.80**	~~0.70~~	~~0.50~~
DoubleMetaphone	~~0.50~~	~~0.75~~	~~0.75~~	~~0.75~~	~~0.75~~
Jaro	**0.96**	**0.93**	**0.90**	~~0.71~~	~~0.69~~
LongestCommonSubstring	~~0.44~~	~~0.22~~	~~0.20~~	~~0.00~~	~~0.20~~
Caverphone	**0.91**	**0.90**	**0.90**	**0.99**	**0.82**
Nysiis	~~0.75~~	**0.88**	**0.88**	~~0.67~~	~~0.50~~
Phonem	~~0.78~~	**0.89**	~~0.78~~	**0.90**	~~0.78~~
Phonetex	**0.83**	**0.92**	**0.92**	**0.83**	~~0.75~~
Phonex	**0.83**	**0.92**	**0.92**	**0.92**	**0.83**
Phonix	**0.82**	**0.91**	**0.91**	**0.91**	**0.91**
ExactMatchCaseInsensitive	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
ExactMatch	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
LevenshteinDamerau	**0.89**	**0.89**	**0.80**	~~0.70~~	~~0.50~~
KeyboardDistanceQwertz	**0.93**	**0.97**	~~0.74~~	~~0.69~~	~~0.67~~
KeyboardDistanceQwerty	**0.93**	**0.97**	~~0.74~~	~~0.69~~	~~0.63~~

Algorithm	Eiteblerg	Eikelberg	Eichelberg	Heidelberg	Eickelbäck
KeyboardDistanceDvorak	**0.93**	**0.97**	~~0.69~~	~~0.64~~	~~0.66~~

4.1.4 Germany

Algorithm	Gemrany	Germant	German	Gery	France	Virgin Islands
ColognePhonetic	~~0.67~~	~~0.73~~	**0.99**	~~0.33~~	~~0.50~~	~~0.17~~
Levenshtein	~~0.71~~	**0.86**	**0.86**	~~0.57~~	~~0.29~~	~~0.21~~
DoubleMetaphone	~~0.25~~	**0.99**	**0.99**	~~0.50~~	~~0.25~~	~~0.50~~
Jaro	**0.95**	**0.90**	**0.95**	~~0.00~~	~~0.00~~	~~0.48~~
LongestCommonSubstring	~~0.29~~	**0.86**	**0.86**	~~0.43~~	~~0.00~~	~~0.00~~
Caverphone	**0.90**	**0.90**	**0.99**	~~0.78~~	~~0.70~~	~~0.57~~
Nysiis	~~0.71~~	~~0.71~~	**0.86**	~~0.57~~	~~0.43~~	~~0.36~~
Phonem	~~0.71~~	**0.86**	**0.86**	~~0.57~~	~~0.29~~	~~0.23~~
Phonetex	**0.80**	**0.91**	**0.99**	**0.80**	~~0.70~~	~~0.53~~
Phonex	**0.90**	**0.90**	**0.99**	**0.89**	**0.80**	~~0.62~~
Phonix	**0.80**	**0.89**	**0.99**	**0.88**	~~0.70~~	~~0.50~~
ExactMatchCaseInsensitive	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
ExactMatch	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
LevenshteinDamerau	**0.86**	**0.86**	**0.86**	~~0.57~~	~~0.29~~	~~0.21~~
KeyboardDistanceQwertz	**0.92**	**0.95**	**0.86**	~~0.51~~	~~0.73~~	~~0.41~~
KeyboardDistanceQwerty	**0.92**	**0.99**	**0.86**	~~0.55~~	~~0.73~~	~~0.43~~
KeyboardDistanceDvorak	**0.94**	**0.97**	**0.86**	~~0.54~~	~~0.53~~	~~0.47~~

4.1.5 4401 4th Ave. S.

Algorithm	4401 4th Avi. S.	4404 4th Ave S	12305 5th Helena Dr.	Kellerstr. 30
ColognePhonetic	**0.99**	**0.99**	~~0.07~~	~~0.00~~

Algorithm	4401 4th Avi. S.	4404 4th Ave S	12305 5th Helena Dr.	Kellerstr. 30
Levenshtein	**0.94**	**0.81**	~~0.40~~	~~0.12~~
DoubleMetaphone	**0.99**	**0.99**	~~0.25~~	~~0.12~~
Jaro	**0.96**	**0.89**	~~0.60~~	~~0.39~~
LongestCommonSubstring	~~0.69~~	~~0.19~~	~~0.00~~	~~0.00~~
Caverphone	**0.99**	**0.81**	~~0.56~~	~~0.50~~
Nysiis	**0.99**	~~0.60~~	~~0.33~~	~~0.38~~
Phonem	**0.90**	**0.90**	~~0.20~~	~~0.00~~
Phonetex	**0.99**	**0.99**	~~0.58~~	~~0.58~~
Phonex	**0.99**	**0.99**	~~0.58~~	~~0.58~~
Phonix	**0.99**	**0.99**	~~0.58~~	~~0.60~~
ExactMatchCaseInsensitive	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
ExactMatch	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
LevenshteinDamerau	**0.94**	**0.81**	~~0.40~~	~~0.12~~
KeyboardDistanceQwertz	**0.97**	~~0.79~~	~~0.60~~	~~0.63~~
KeyboardDistanceQwerty	**0.97**	~~0.79~~	~~0.60~~	~~0.63~~
KeyboardDistanceDvorak	**0.99**	~~0.79~~	~~0.52~~	~~0.72~~

4.1.6 London

Algorithm	Lnodon	Londonn	Londn	Llondon	Lomdon	Londom
ColognePhonetic	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**
Levenshtein	~~0.67~~	**0.86**	**0.83**	**0.86**	**0.83**	**0.83**
DoubleMetaphone	**0.99**	**0.99**	**0.99**	**0.99**	~~0.75~~	~~0.75~~
Jaro	**0.94**	**0.95**	**0.94**	**0.95**	~~0.79~~	**0.89**
LongestCommonSubstring	~~0.17~~	**0.86**	~~0.67~~	~~0.14~~	~~0.33~~	**0.83**
Caverphone	**0.90**	**0.99**	**0.99**	**0.99**	**0.90**	**0.90**
Nysiis	~~0.67~~	**0.99**	**0.83**	**0.99**	**0.99**	**0.99**
Phonem	~~0.67~~	**0.99**	**0.83**	**0.99**	**0.83**	**0.83**

Algorithm	Lnodon	Londonn	Londn	Llondon	Lomdon	Londom
Phonetex	**0.99**	**0.99**	**0.99**	**0.91**	**0.90**	**0.90**
Phonex	**0.99**	**0.99**	**0.99**	**0.91**	**0.99**	**0.99**
Phonix	**0.99**	**0.99**	**0.99**	**0.91**	**0.99**	**0.99**
ExactMatchCaseInsensitive	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
ExactMatch	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
LevenshteinDamerau	**0.83**	**0.86**	**0.83**	**0.86**	**0.83**	**0.83**
KeyboardDistanceQwertz	**0.91**	**0.86**	~~0.79~~	~~0.67~~	**0.99**	**0.99**
KeyboardDistanceQwerty	**0.91**	**0.86**	~~0.79~~	~~0.67~~	**0.99**	**0.99**
KeyboardDistanceDvorak	**0.83**	**0.86**	~~0.75~~	~~0.55~~	**0.97**	**0.97**

4.1.7 Gawedzki

Algorithm	Gaewdzki	Kawetski	Gavetski	Gawetschi	Kawetschi
ColognePhonetic	**0.99**	**0.99**	**0.99**	~~0.67~~	~~0.67~~
Levenshtein	~~0.75~~	~~0.62~~	~~0.62~~	~~0.56~~	~~0.44~~
DoubleMetaphone	**0.99**	**0.99**	~~0.62~~	~~0.50~~	~~0.50~~
Jaro	**0.96**	~~0.75~~	~~0.75~~	~~0.73~~	~~0.65~~
LongestCommonSubstring	~~0.25~~	~~0.00~~	~~0.25~~	~~0.44~~	~~0.00~~
Caverphone	**0.91**	**0.99**	**0.91**	**0.99**	**0.99**
Nysiis	**0.99**	~~0.60~~	~~0.57~~	**0.80**	~~0.60~~
Phonem	~~0.75~~	**0.88**	**0.88**	~~0.70~~	~~0.70~~
Phonetex	**0.99**	**0.80**	**0.82**	**0.80**	~~0.70~~
Phonex	**0.99**	**0.89**	**0.90**	**0.99**	**0.89**
Phonix	**0.99**	**0.90**	**0.91**	**0.90**	**0.80**
ExactMatchCaseInsensitive	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
ExactMatch	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
LevenshteinDamerau	**0.88**	~~0.62~~	~~0.62~~	~~0.56~~	~~0.44~~
KeyboardDistanceQwertz	**0.98**	**0.91**	**0.91**	~~0.77~~	~~0.75~~

Algorithm	Gaewdzki	Kawetski	Gavetski	Gawetschi	Kawetschi
KeyboardDistanceQwerty	**0.98**	**0.94**	**0.94**	~~0.80~~	~~0.77~~
KeyboardDistanceDvorak	**0.90**	**0.94**	**0.96**	**0.81**	~~0.78~~

4.1.8 Bayer Bepanthen 500ml

Algorithm	Bayer Bepanten 500	Bayer Bebanthen 500l	BAY Bepanthen 500ml	Bepanthen 500ml	Bebanteen 500ml	Bayer Bepanthen 250ml
ColognePhonetic	**0.83**	**0.99**	~~0.62~~	~~0.25~~	~~0.25~~	**0.99**
Levenshtein	**0.86**	**0.90**	**0.81**	~~0.71~~	~~0.62~~	**0.90**
DoubleMetaphone	**0.99**	**0.99**	~~0.50~~	~~0.25~~	~~0.25~~	**0.99**
Jaro	**0.95**	**0.95**	**0.81**	**0.80**	~~0.77~~	**0.97**
LongestCommonSubstring	~~0.57~~	~~0.38~~	~~0.05~~	~~0.05~~	~~0.05~~	~~0.76~~
Caverphone	**0.94**	**0.94**	**0.94**	**0.88**	**0.88**	**0.94**
Nysiis	**0.92**	**0.92**	**0.92**	~~0.75~~	~~0.75~~	**0.99**
Phonem	**0.89**	**0.94**	**0.89**	~~0.72~~	~~0.72~~	**0.89**
Phonetex	**0.87**	**0.87**	**0.93**	**0.87**	**0.87**	**0.99**
Phonex	**0.92**	**0.99**	**0.92**	**0.92**	**0.92**	**0.99**
Phonix	**0.92**	**0.99**	**0.92**	**0.92**	**0.92**	**0.99**
ExactMatchCaseInsensitive	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
ExactMatch	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~	~~0.00~~
LevenshteinDamerau	**0.86**	**0.90**	**0.81**	~~0.71~~	~~0.62~~	**0.90**
KeyboardDistanceQwertz	~~0.78~~	**0.92**	~~0.72~~	~~0.52~~	~~0.56~~	**0.97**
KeyboardDistanceQwerty	~~0.78~~	**0.92**	~~0.72~~	~~0.54~~	~~0.56~~	**0.97**
KeyboardDistanceDvorak	~~0.77~~	**0.92**	~~0.69~~	~~0.59~~	~~0.60~~	**0.97**

Thanks for reading, and we hope this documentation helps you to navigate HEDDA.IO with ease. If you haver any questions, don't hesitate to contact us.