|  |
| --- |
| Data Governance Atlas  User Manual |
|  |

|  |  |
| --- | --- |
| Data del documento: | 22/11/2023 |

|  |  |
| --- | --- |
| Informazioni Generali | |
| Cliente | Ministero di Giustizia |
| Progetto | Metadata Governance |
| Documento |  |
| Ambito | ATLAS |
| Versione | 0.1 |
| Stato Approvazione | Da approvare |
| Società | RTI |
| Data | 22/11/2023 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Revisioni | | | | |
| Ver. | Data | Descrizione | Autore | Approvazione |
| 0.1 | 22/11/2023 | Prima bozza del manuale utente | Accenture - Nursaule Myrzashova |  |

|  |  |
| --- | --- |
| Lista di Distribuzione | |
| Accenture | Alessia Pacella |

Table of Contents

[Introduction 3](#_Toc151560342)

[General manual for software execution 4](#_Toc151560343)

[Functionality 4](#_Toc151560344)

## Introduction

This manual provides general information and necessary steps about implementation of Data Governance Atlas project which was developed to insert technical and business metadata into the Ministry of Justice platform from provided excel file.

The given excel file contains valuable information about all users of the platform, both business users and developers.

There are four pages with cover, instruction, dataset and column data. The first page describes general information about given file and other three pages. The instruction page provides a list of tracked metadata. The dataset page contains use cases related to metadata management on hive tables when the last column one show also the columns in this hive tables with metadata management.

**Used technologies and software description**

**Apache Atlas** – Data Governance and Metadata framework for Hadoop. It allows users to define a model for the metadata objects they want to manage. The model is composed of definitions called “types”. Instances of ‘types’ called “entities” represent the actual metadata objects that are managed. All functionality of Atlas is exposed to end users via a REST API that allows types and entities to be created, updated and deleted. It is also the primary mechanism to query and discover the types and entities managed by Atlas.

**Apache Ranger** - framework to enable, monitor and manage comprehensive data security across the Hadoop platform. It has advanced security management solution for the Hadoop ecosystem having wide integration with a variety of Hadoop components. By integrating with Atlas, Ranger allows security administrators to define metadata driven security policies for effective governance. Ranger is a consumer to the metadata change events notified by Atlas.

**Python3** – programming language which was used to write all API requests and retrieve information from the server.

The implemented software “API\_codice.py” reads provided excel file called “MdG-MetadataManagementSIAMM.xlsx” and transform collected data in Apache Atlas readable format for sending to the server and adding data in Apache Atlas platform by using REST API. During the execution of the project the general excel file was not modified by passing all parameters to the dictionary form and transformations during the process.

The described software adds different types of tables with columns and create relationships between them on the platform. All connected metadata are stored in different Atlas platform fields: technical properties, user-defined properties and business metadata. For today, script adds tables with columns, updates technical and business metadata of the table and delete columns. The retrieving Atlas types to define the type of tables and columns for the system configured automatically in “atlas\_type.py” script by sending request to hive and impala stores on remote server.

## General manual for software execution

1. Save project locally with necessary excel file in the same folder.

2. Make sure that the excel file has the name “MetadataManagementSIAMM.xlsx”.

3. Open Visual Studio Code and choose data-governance-atlas project folder.

4. Install modules if it is not already installed: >pip Install PyApacheAtlas

5. Launch the “API\_codice.py” script by clicking on play button on the upper right corner of Visual Studio Code.

6. Be patient and wait till program will finish the execution. Warning: if you are running the excel file the first time it can take up to 10-15 minutes to upload all data to Atlas.

7. During the execution the state of work will be printed in the terminal.

8. Go to **https://atlas.dai-training.com/** and login in with **username**: admin, **password**: admin.

9. On the left menu bar click search button to see all entities in the system. If you need to specify and find the appropriate type of the entity as hive\_table, hive\_column on the left menu in **search by type** field specify necessary type. If you need find the specific entity (table, column, dataset) write the name of the entity in the top **search entity** field.

10. Click on the name of the entity in the provided list to see all upload metadata information with relation to the table or columns.

## Functionality

The user can update tables with new fields of metadata, add, delete columns by changing original excel file. However during the implementation different excel files were created to test the performance of software. The next variables with excel files were created:

1. Example of adding table “DS\_SIAMM\_DC\_BENEFICIARIO\_TAB\_FAT\_test” with 10 columns and updating them into the Atlas system.

path\_example= 'develop\MdG-MetadataManagementSIAMM\_example.xlsx'

1. Adding “test” column into existing “DS\_SIAMM\_DC\_BENEFICIARIO\_TAB\_FAT” table.

path\_test\_add\_column = "develop\col\_add\_test.xlsx"

1. Delete COD\_CODICE\_PERSONAFISICASA, COD\_CODICE\_PERSONAGIURIDICASA, COD\_CODICE\_RICHIESTA, COD\_ID\_SESSIONE columns from “DS\_SIAMM\_DC\_BENEFICIARIO\_TAB\_FAT” table.

path\_delete\_column = "develop\col\_delete\_test.xlsx"

1. Delete NUM\_IVA\_MODELLO, ASL\_DENOMINAZIONE, RECORD\_HASH, SCHEMA\_PROVENIENZA, TST\_ELABORATION, SK\_BENEFICIARIO, COD\_CODICE\_BENEFICIARIO from “DS\_SIAMM\_DC\_BENEFICIARIO\_TAB\_FAT” table.

Add “test1\_SK\_TIPOPERSONASA”, “test2\_SK\_TIPOPERSONASA” columns into DS\_SIAMM\_DC\_TIPOPERSONASA\_TAB\_DIM table.

path\_add\_del\_col = "develop\col\_del\_add\_test.xlsx"

1. Edit next attributes only in “DS\_SIAMM\_DC\_BENEFICIARIO\_TAB\_FAT\_test” table:

**Technical properties**

*Descrizione Dataset:* Beneficiario della liquidazione 🡪 Beneficiario della liquidazione\_test\_update

**Business Metadata**

*Dominio/Area di riferimento:* SIAMM 🡪 SIAMM\_test

*Nome entità di Business:* CG\_PROVV\_GIUDIZIARI 🡪 CG\_PROVV\_GIUDIZIARI\_test

**User-defined properties**

*Layer applicativo:* Curated Layer 🡪 Curated Layer\_test

path\_edit\_attr = 'develop\diff\_tech\_attr\_test.xlsx'

1. Edit the one column in DS\_SIAMM\_DC\_BENEFICIARIO\_TAB\_FAT” table with full dataset of tables.

**Technical properties**

*Descrizione Dataset:* Beneficiario della liquidazione 🡪 Beneficiario della liquidazione\_test\_full

**Business Metadata**

*Dominio/Area di riferimento:* SIAMM 🡪 SIAMM\_test\_full

*Categoria di Business:* Casellario Giudiziario 🡪Casellario Giudiziario\_test\_full

*Nome entità di Business:* CG\_PROVV\_GIUDIZIARI 🡪 CG\_PROVV\_GIUDIZIARI\_test\_full

**User-defined properties**

*Sistema Sorgente:* Raw Layer 🡪 Raw Layer\_test\_full

*Layer applicativo:* Curated Layer 🡪 Curated Layer\_test\_full

*Chiave di partizionamento:* DATA\_RIFERIMENTO\_SISTEMA\_ORIG 🡪 \_test\_full

path\_full\_dataset\_modify\_table\_attr = "develop\modify\_full\_test\_dataset.xlsx"

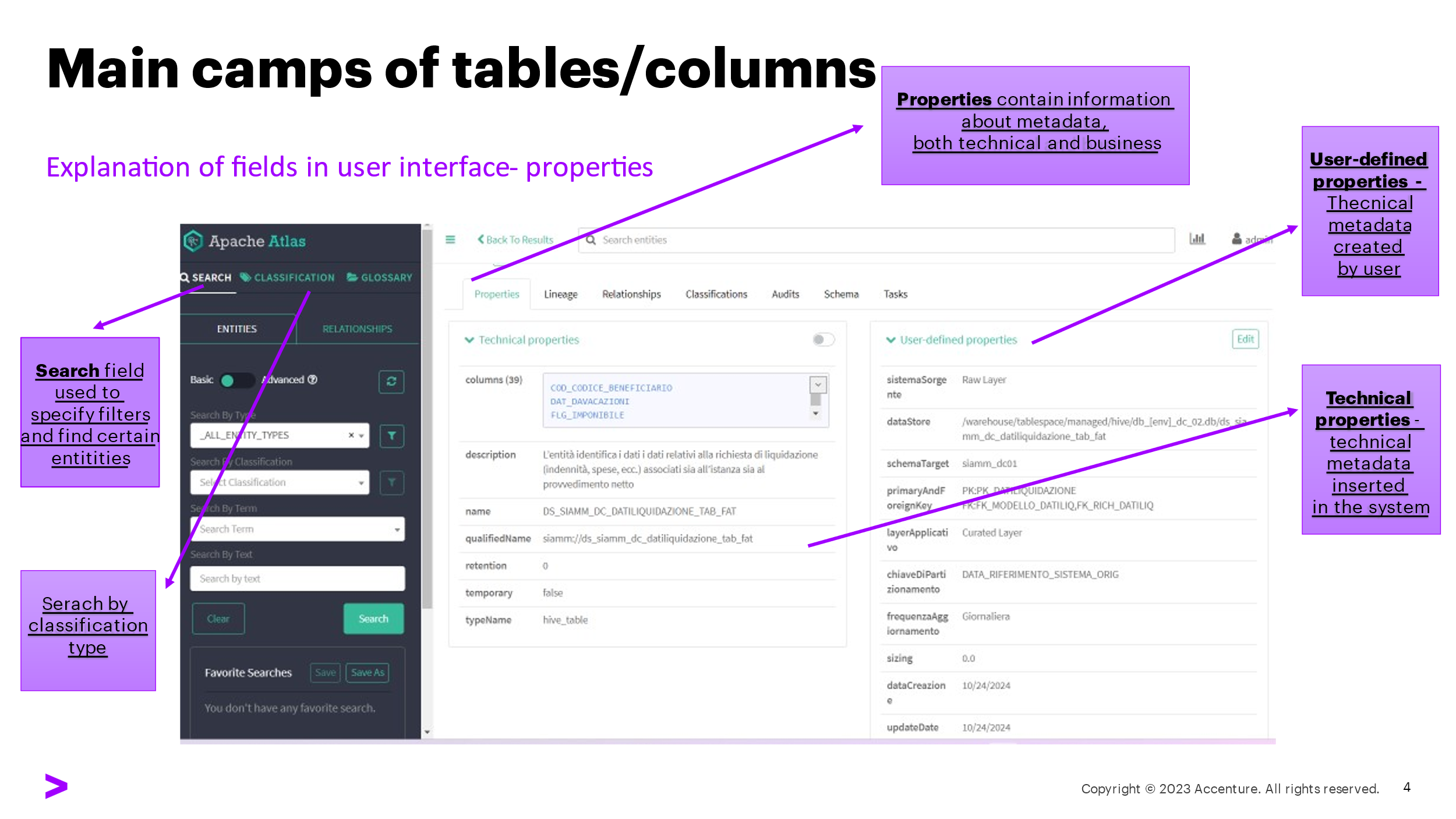
**Note:** To check above functionality is mandatory to change the input “path” variable in the next function in the bottom of the script:

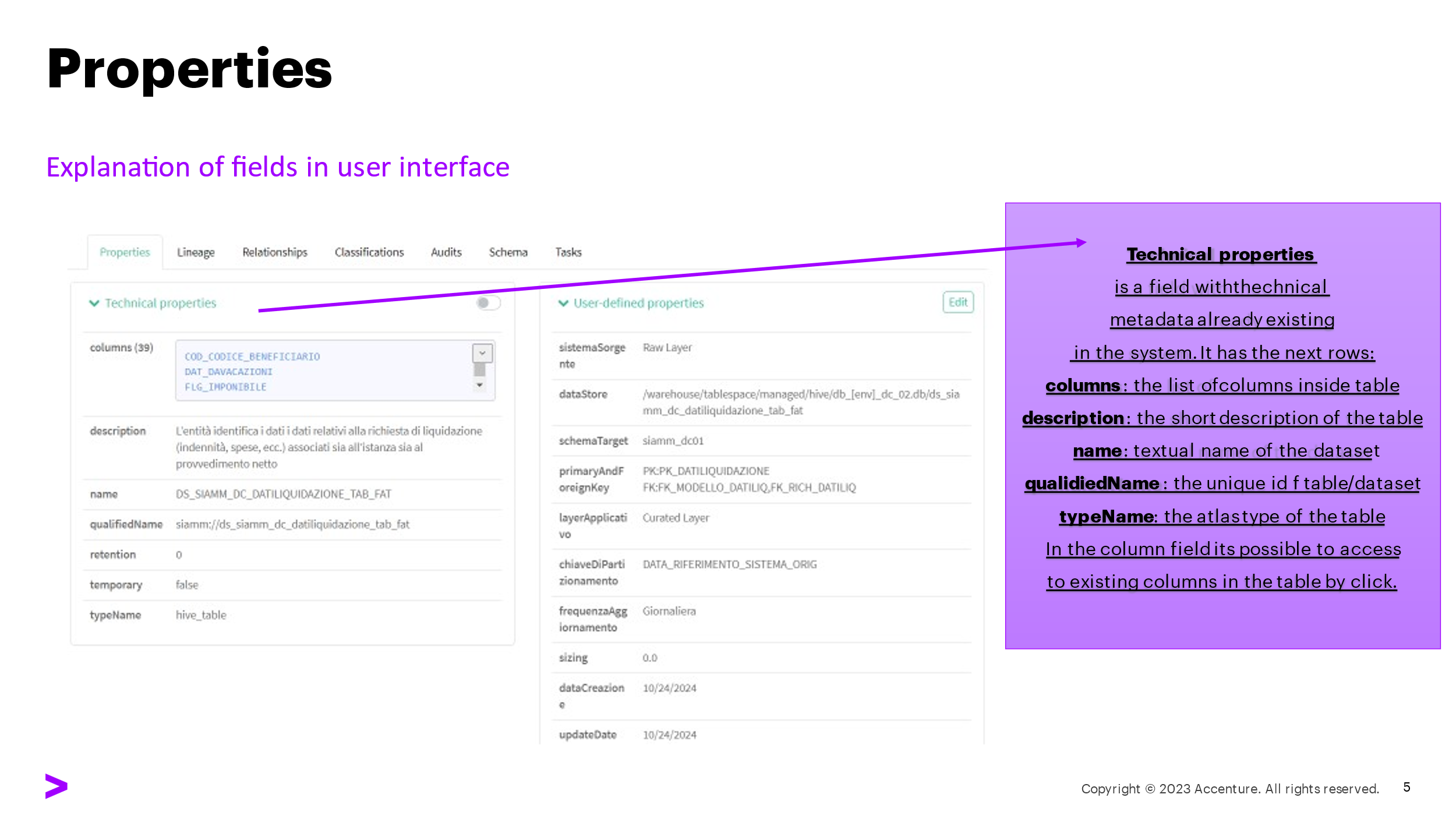
list\_of\_metadata = read\_exel\_tech\_business\_metadata(path)

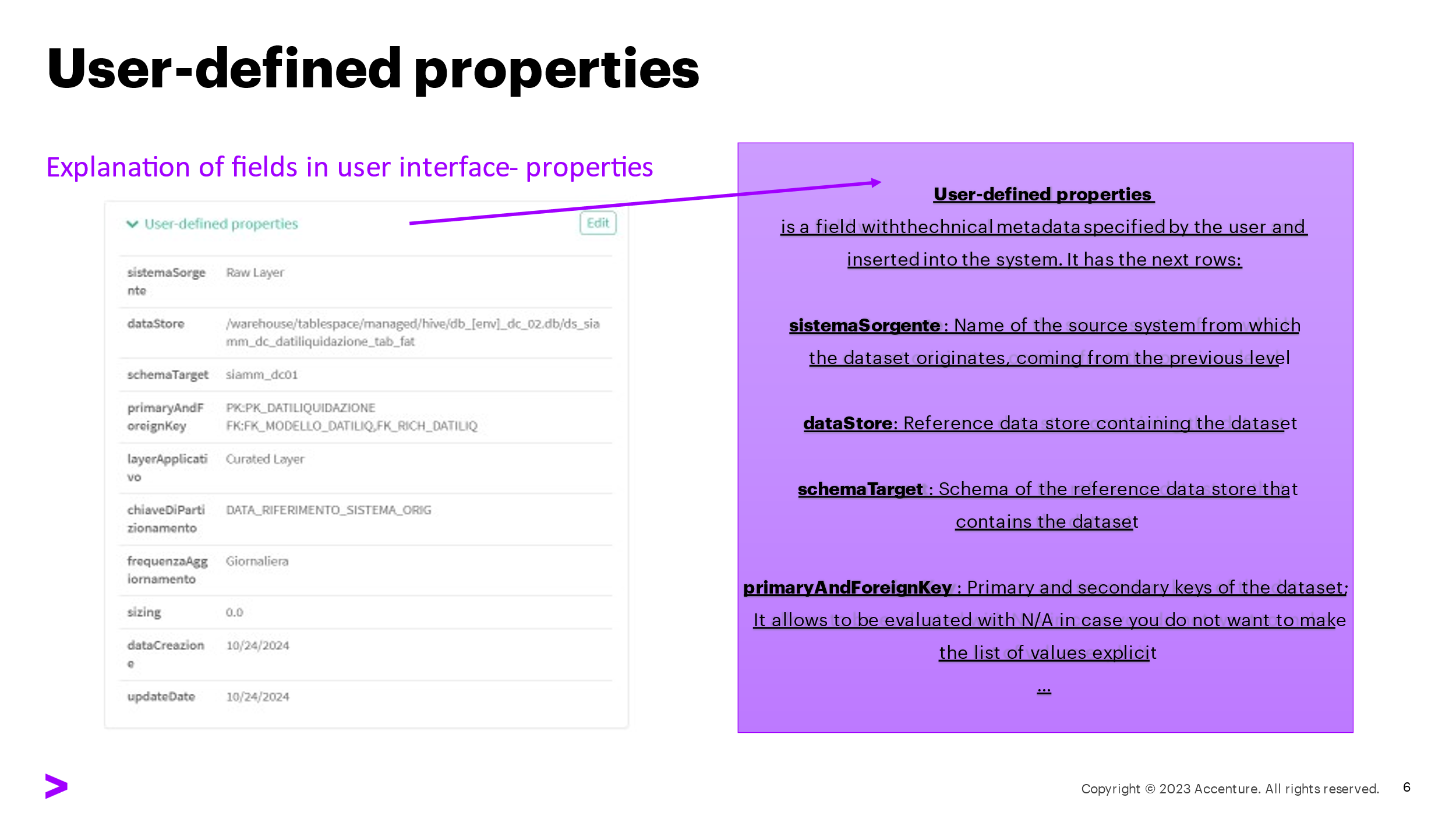
dataset\_dict\_records = convert\_column\_to\_dict\_all(path, sheet\_name="Dataset" )

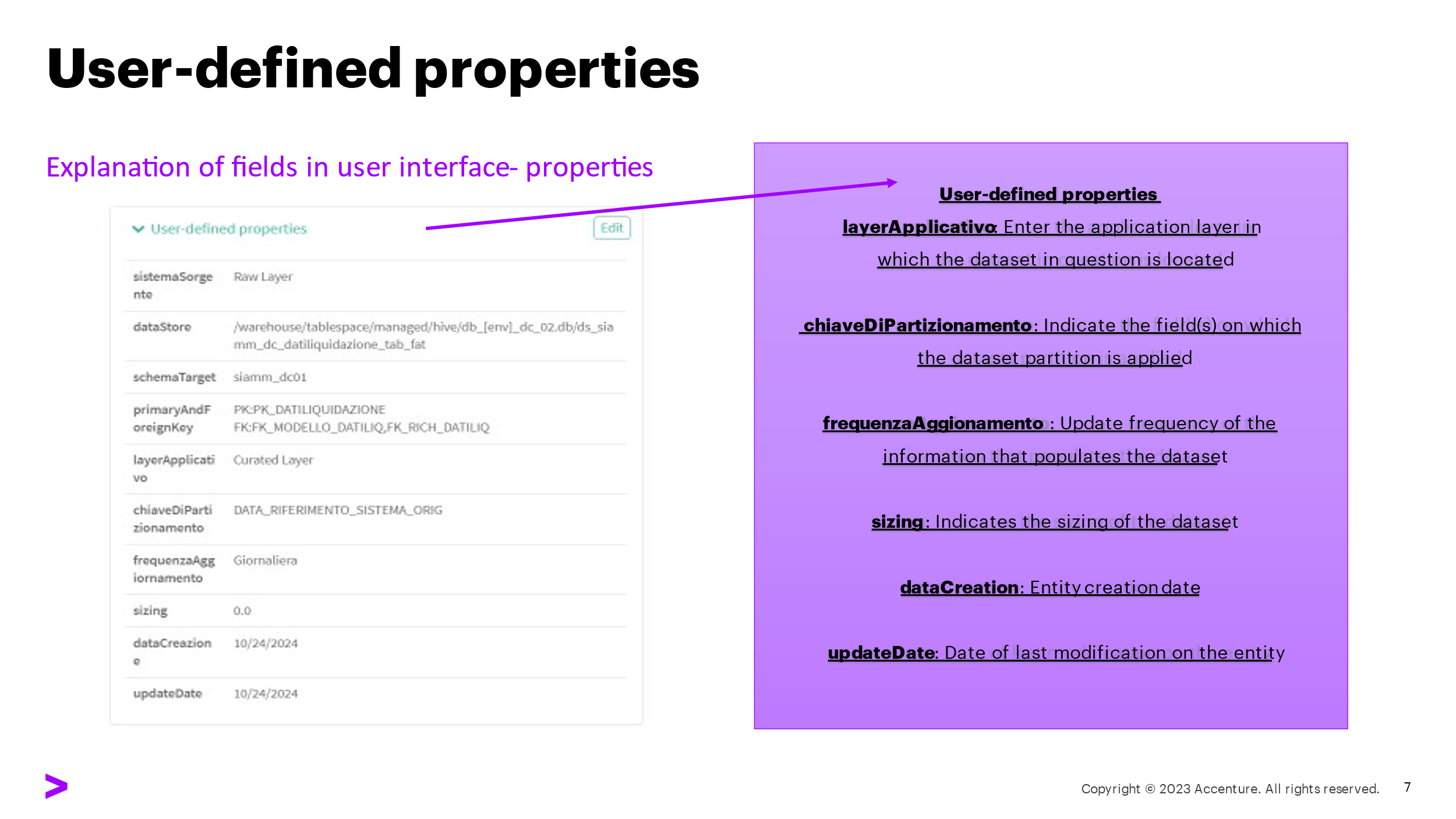
columns\_dict\_records = convert\_column\_to\_dict\_all(path, sheet\_name="Colonne" )

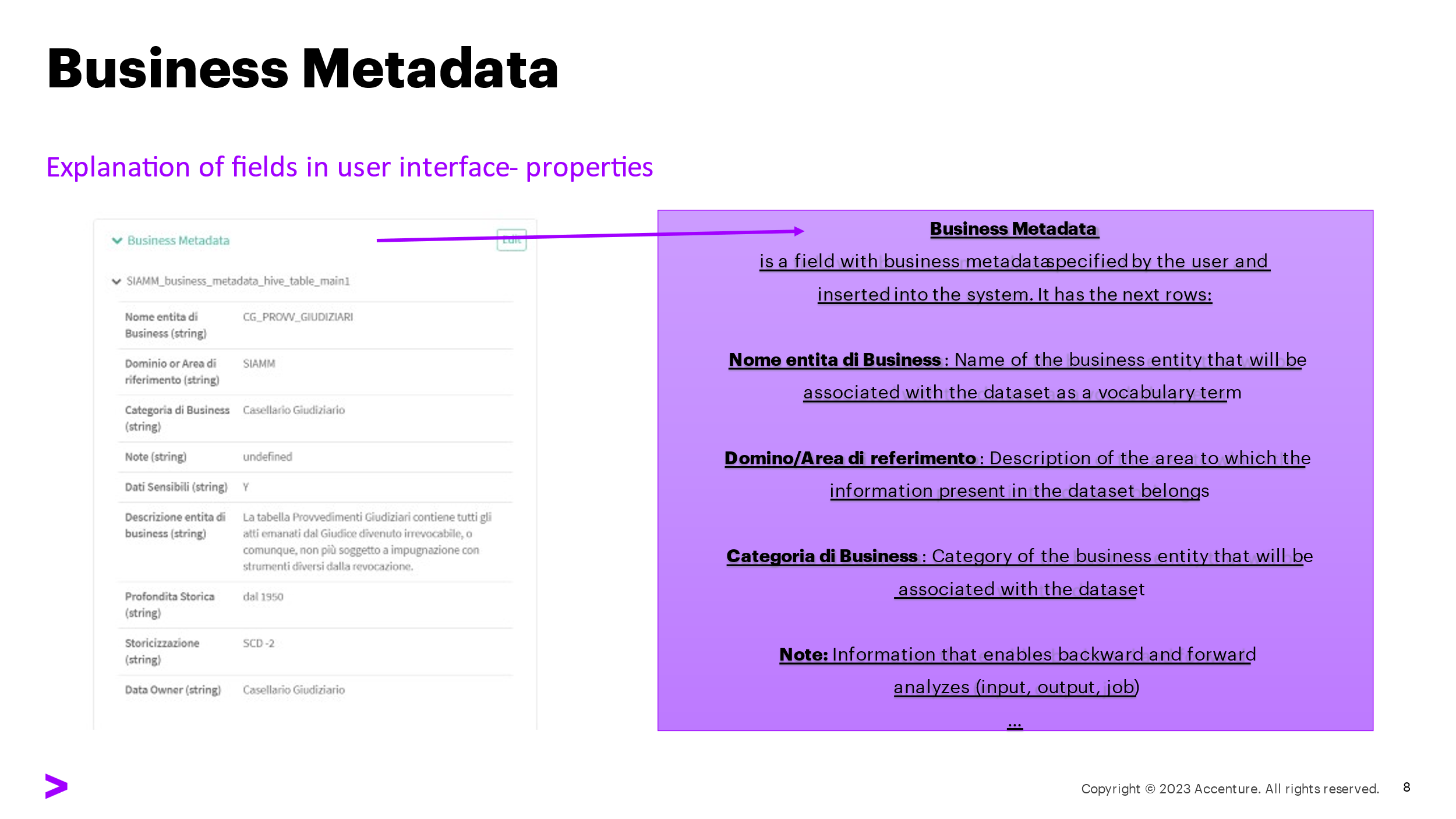
**General User Manual with Interface**

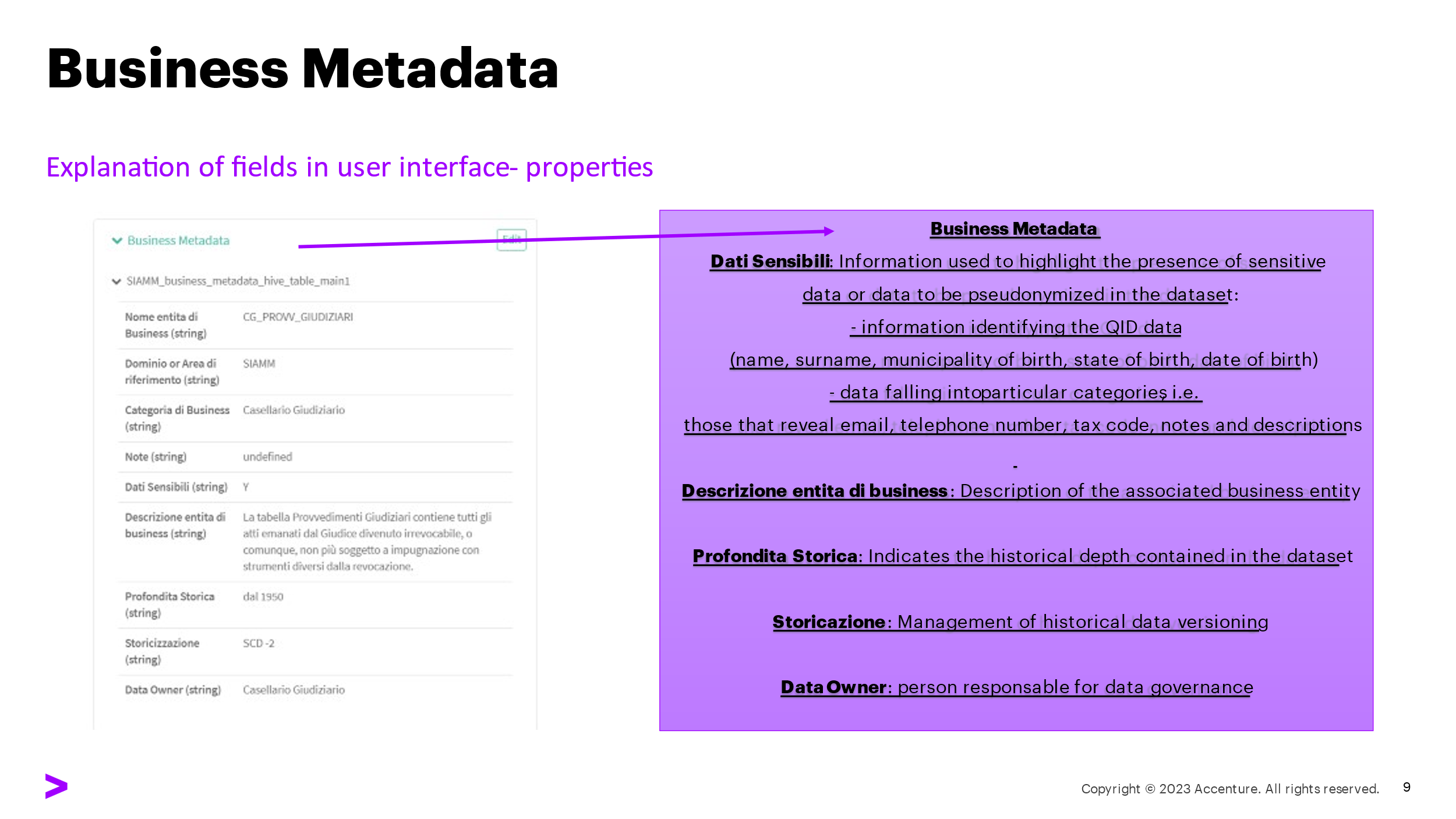


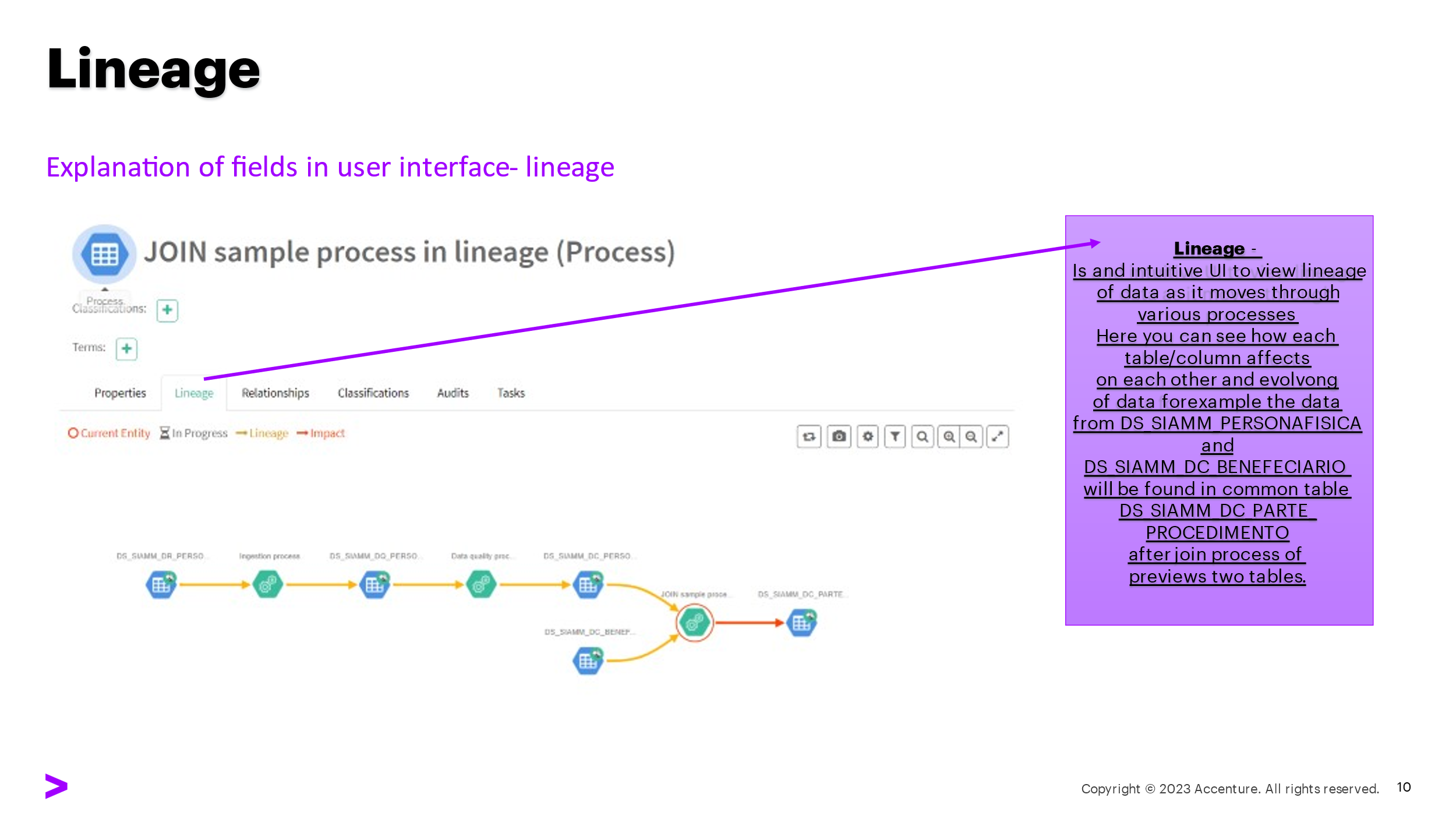


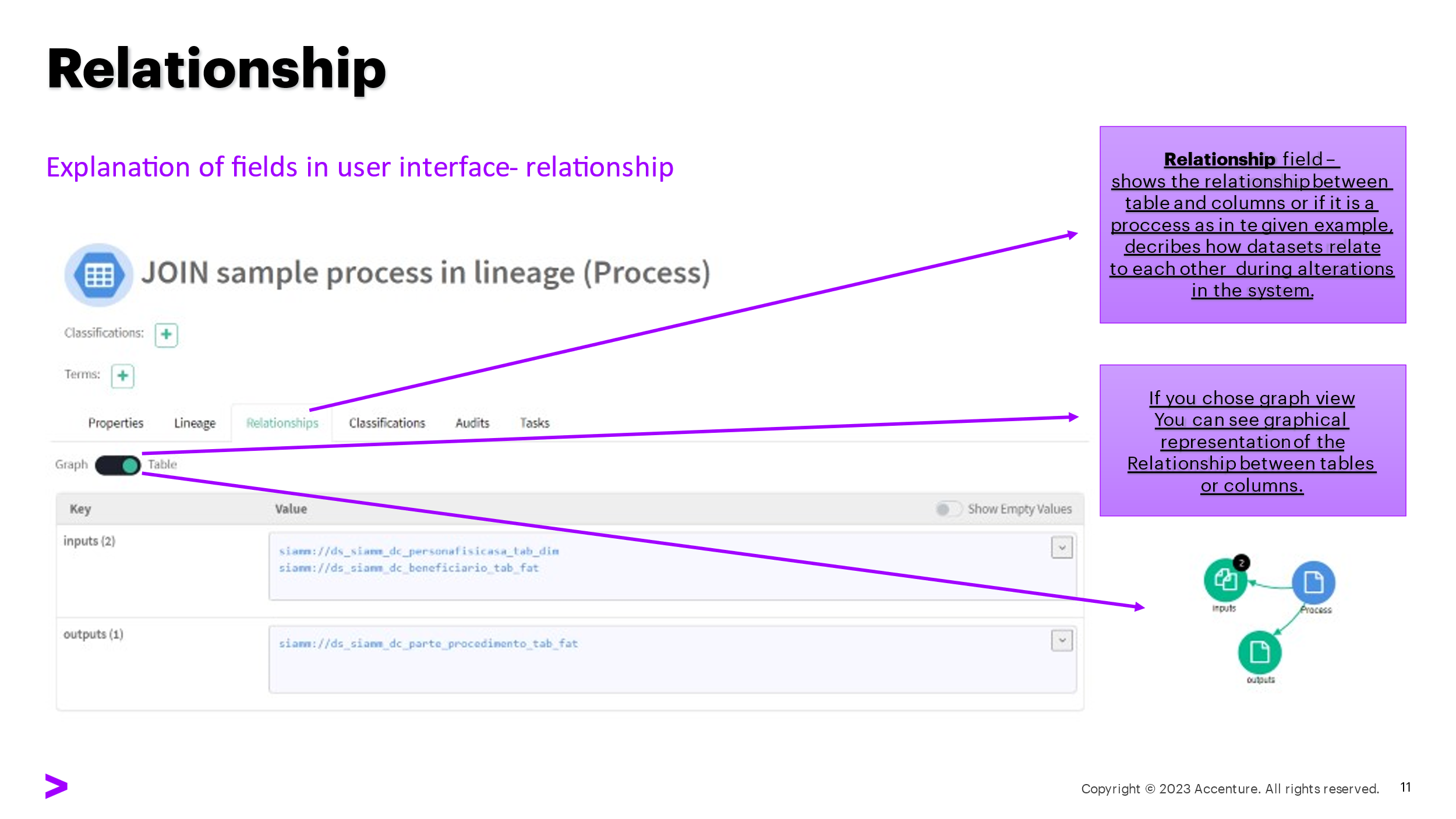


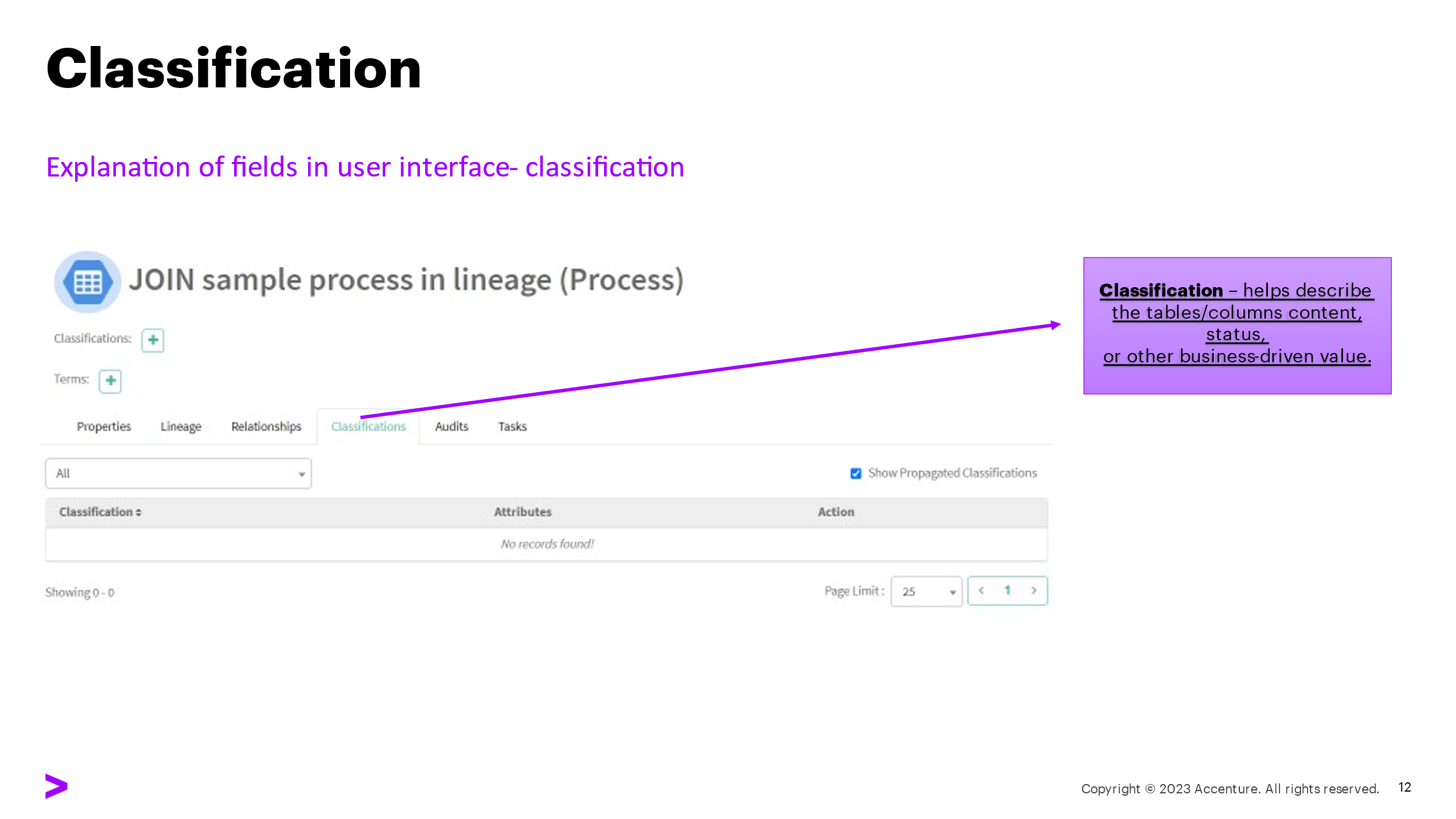


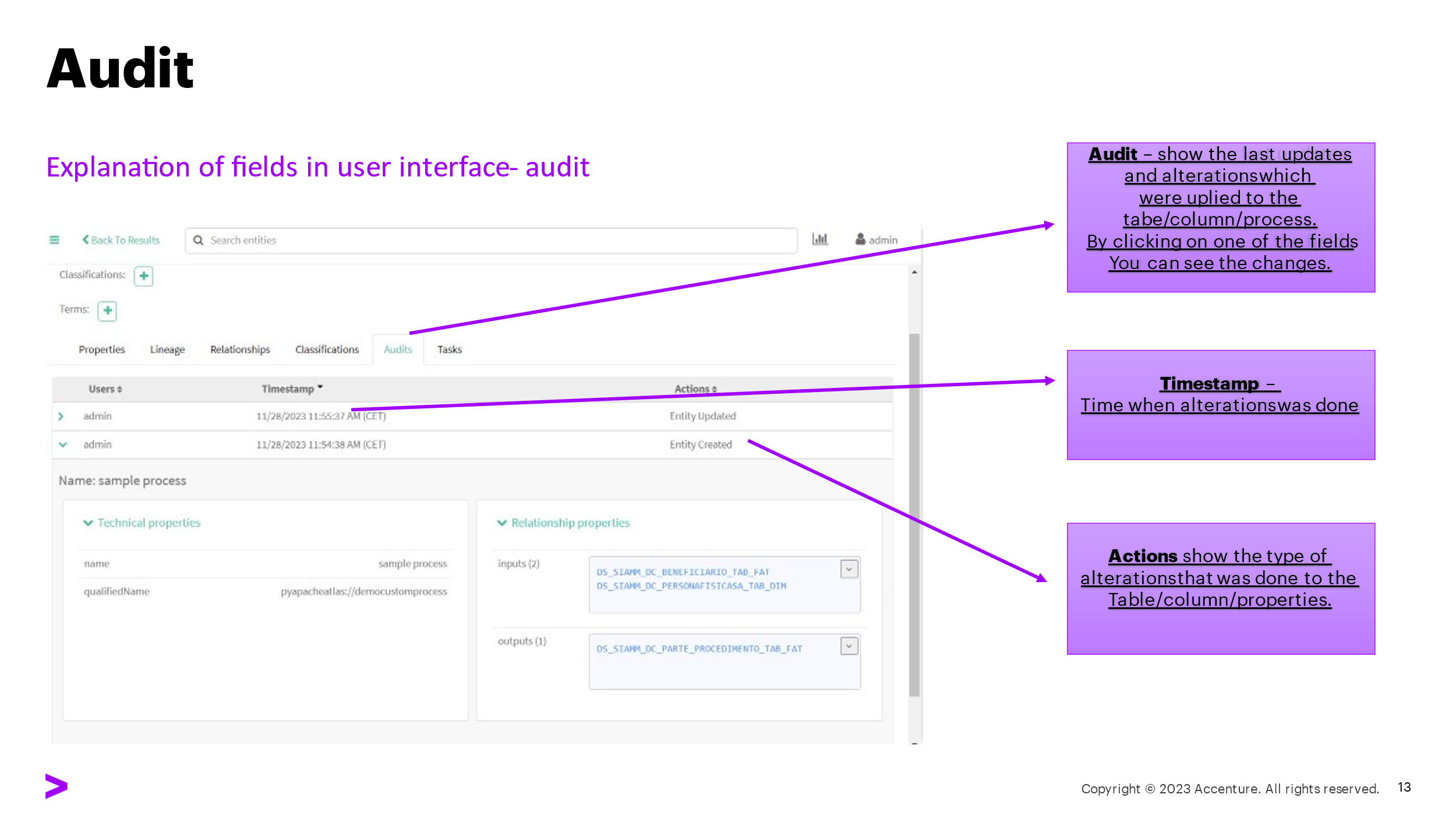


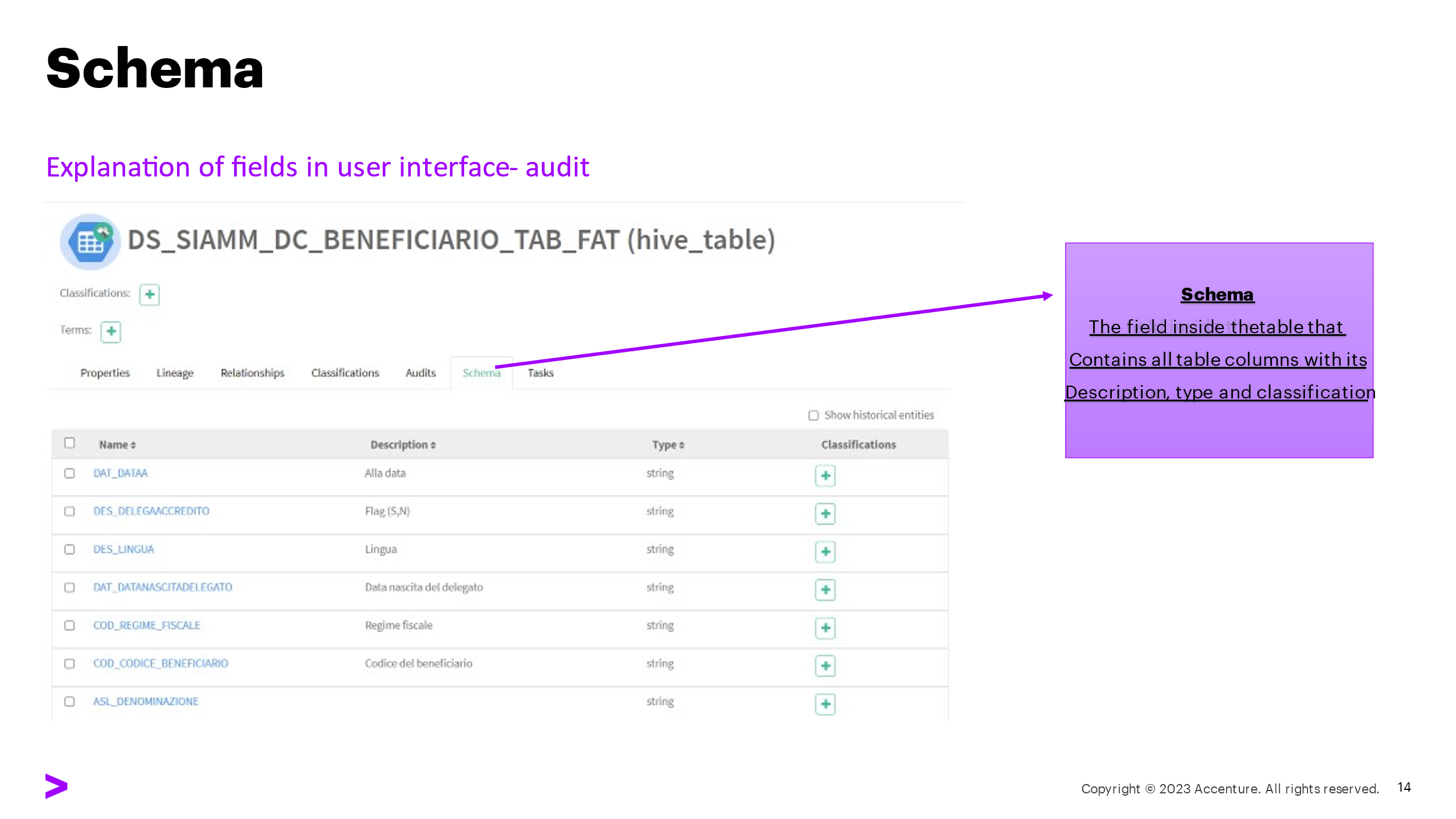


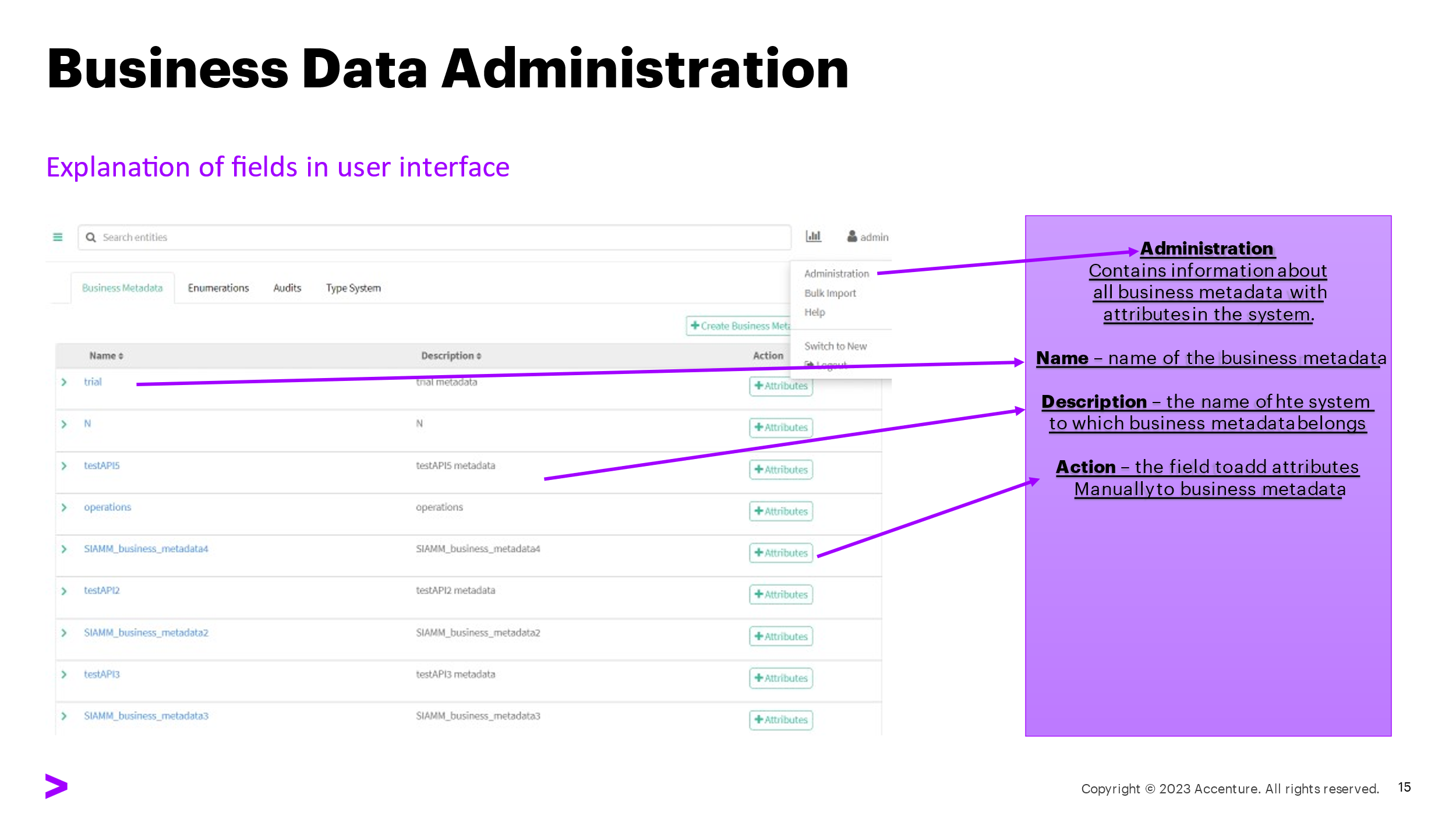












|  |
| --- |
|  |

Copyright © 2023 Accenture

All rights reserved.

Accenture and its logo are trademarks of Accenture.