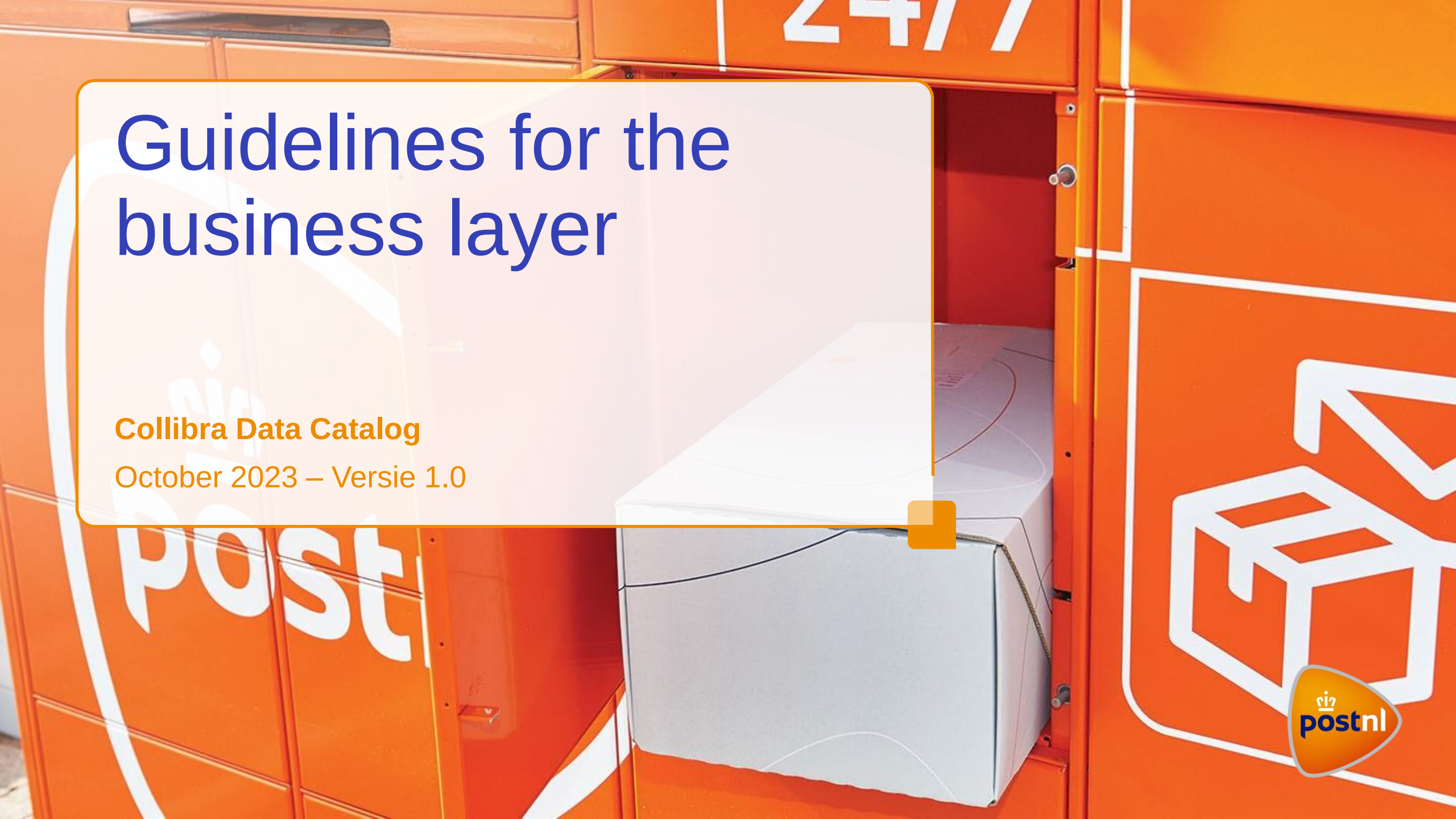
Guidelines for the business layer

Collibra Data Catalog

October 2023 – Versie 1.0



Index

Introduction

The PostNL Data Catalog

Introduction

PostNL uses the Collibra Data Governance Center (DGC) platform to support its data organization in the areas of data quality, data governance, and data analysis. It assists in locating data, provides metadata inventory, and offers information necessary to determine if the data is suitable for its intended use.

The PostNL Data Catalog is a tool that enables PostNL's data organization to enhance the accessibility, accuracy, and relevance of data across the entire company. This provides crucial support for:

- **Data Usage:** Metadata enhances knowledge about data. The more users know about data, the better they can determine its usability and limitations.
- **Data Management:** The data catalog provides insights and a better understanding of the data that PostNL possesses. This makes the data known and manageable and is a prerequisite for the professionalization of data management capabilities such as data governance and data quality management.

The PostNL Data Catalog is one of the solution building blocks within the data capability 'Meta Data Management.'

*"**Metadata Management** is the discipline that involves the collection, maintenance, and standardization of metadata (including access and distribution)."*

Structure of the PostNL Data Catalog

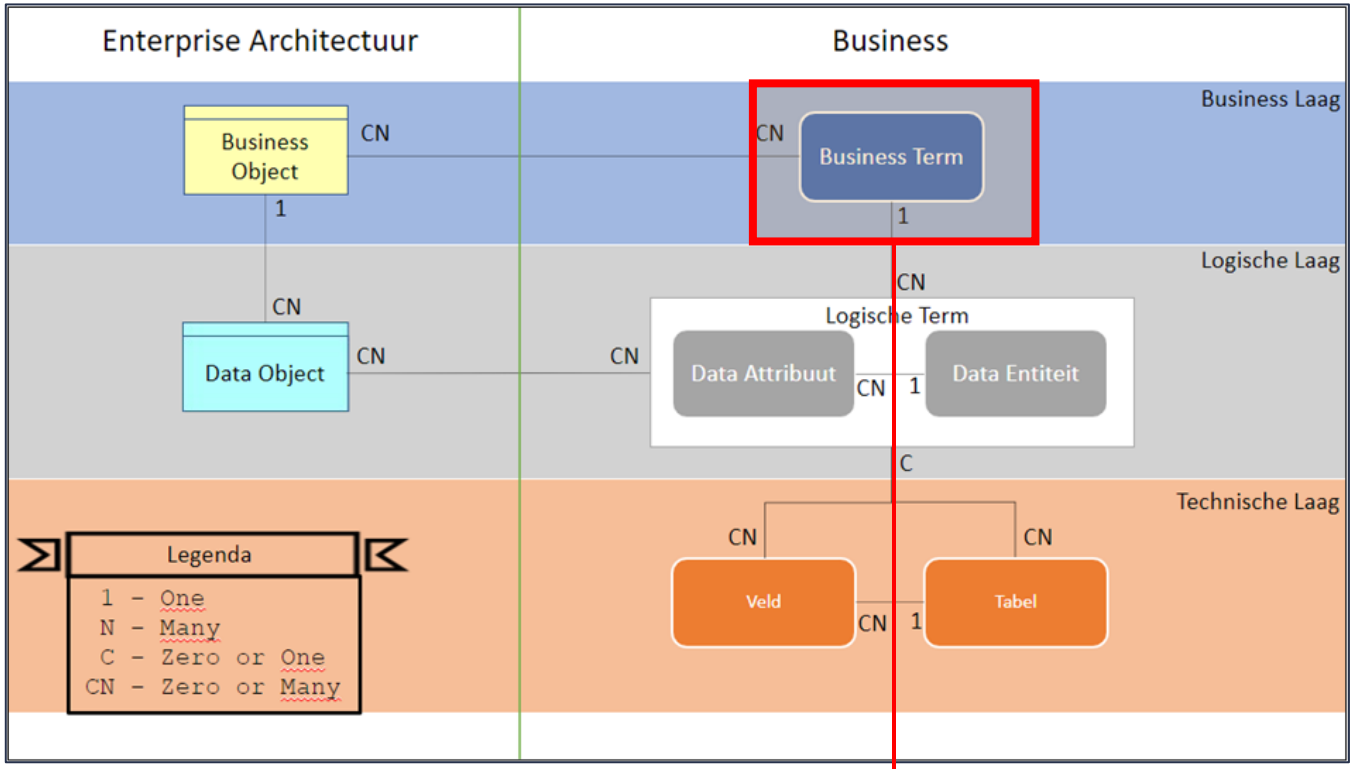
Three layers: business, logical and technical

The PostNL Data Catalog is composed of three different layers:

- **Business layer**
  The terms commonly used by the business and in everyday usage.

- **Logical Layer:**
  The necessary terms (entities and attributes) to link the business terms to the technical layer.

- **Technical layer**
  The fields and tables as they actually appear in the systems and applications.

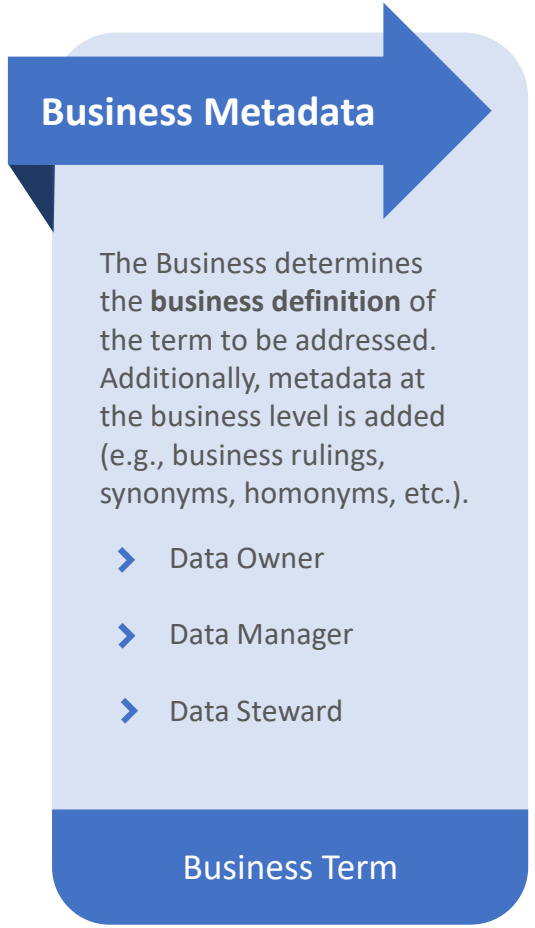


The focus in this document is specifically on the 'Business Layer.' In other words, what is a business term and what is the associated metadata?

Business terms

Basic principles

- "*A business term is a word used to express something or a concept within a specific context.*"
- In the Data Catalog, a business term is created based on business needs. It involves the words and terms used by the business. Therefore, it is the business that has control over what is included and what is not included in the Data Catalog. The business is represented in this by the Data Owner, Data Manager, and Data Steward.
- A business term can be the same as a business object or data object. This means that definitions and other metadata can be one-to-one identical. However, the Architecture owner is responsible for business objects and data objects, including their associated metadata.



Business terms

Creation

There is a lot of freedom to determine what is considered a business term - and what is not. The intention is not to impose so many restrictions that filling out the Data Catalog is seen as a chore. Therefore, trust is placed in the expertise of the business to determine what goes on the business layer and what does not.

- "Handheld," "Branchenummer" (Branch number), or "Postcode" (Postal code) are all clear business terms and, therefore, belong on the business layer.
- "HH__Code," "Branch_number__c," or "PC6_NL" are technical fields and, therefore, do not belong on the business layer.

Business terms should be written as they are used by the business. The spelling follows how you would find the word in the dictionary:

- The first letter is always capitalized, followed by lowercase letters.
- In the case of multiple words, spaces are used.

Business terms

Changing

To add or modify a business term, you need to be a part of the Business Layer Curator group, also known as a 'Filler.' Within this group, there are no restrictions regarding the modification of business terms. This applies to business terms created by yourself as well as those created by others.

The following process rules apply:

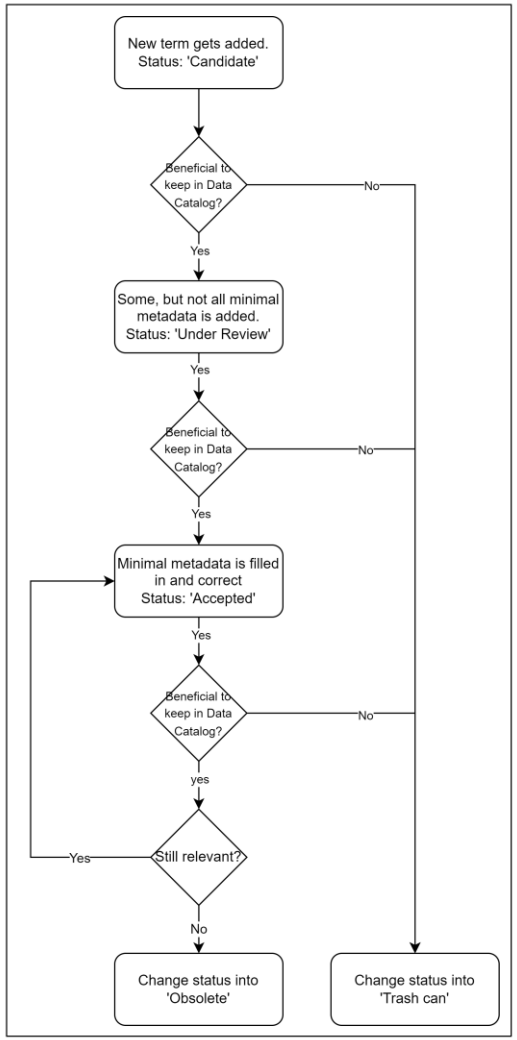
- You are free to modify your own business term (i.e., those you've created yourself).
- You are **not free** to just change a business term created by **someone else**. If the metadata for another business term is incorrect or incomplete, you should contact the owner of that business term to reach a mutual agreement. If the context in which the business term is used deviates significantly from the existing business term, then add the business term in your own, specific context.

Business terms

Status

- A term can be in various phases. The status indicates whether it is, for example, an accepted term with metadata or whether it is a term that is *'Under Review*.'

Status	Betekenis
Candidate	Initial status of a term. This means that the term has been created or imported. At this stage, there is no examination of any metadata.
Under Review	Stakeholders review the Asset. This means that metadata has been added, but not all required fields have been filled in or are accurate.
Accepted	The term and the mandatory metadata are fully and correctly filled in. The meaning is endorsed by the key stakeholders.
Obsolete	The term is outdated. The metadata for this term remains available for reference. Periodically, a review will determine if cleanup is necessary.
Trash can	When a term is entered incorrectly, it is given the status of "Trash can." Terms with this status are periodically, with the owner's approval, deleted.
Invalid	Out of scope.
In Progress	Out of scope.
Approval Pending	Out of scope.



Captured metadata

Aligned with the Data Management Organizations

1	**Definition**
2	Definition (English)
3	Explanation
4	Example
5	**Data Manager**
6	**Data Domain**
7	**Language**
8	System Of Record
9	System of Entry
10	Data Labelling
11	**Personally Identifiable Information**
12	PII Type
13	CIA Classification
14	Confidentiality
15	Integrity
16	Availability
17	Business Ruling
18	Standard Value Format
19	Abbreviation
20	Date Created
21	Date Changed
22	Made/Changed by
23	Status
24	Source (origin metadata)
25	***Synonym***
26	***Translation***
27	***Rel Architecture - Objects***
28	***Rel Logical Layer – Data Entity***
29	***Rel Logical Layer – Data Attribute***

In consultation with the Data Management Organizations within PostNL, it has been determined that the so-called metadata fields mentioned here on the left should be recorded in the PostNL Data Catalog.

- The red-colored fields are mandatory.
- The black-colored fields are optional.
- The blue-colored fields indicate the relationship between the three layers and/or business terms.
- Fields 14 through 16 are currently out of scope.
- Fields 20 through 22 are automatically generated by Collibra itself (standard functionality).

The following slides provide further details on each of these fields.

Captured metadata

Explanation of the fields (1/6)

Metadata	Description	Guideline	Mandatory
Definition	A clear description of the business term in one to two sentences.	• Text field • Definition does NOT contain the term itself • Follows the most common spelling and punctuation • Unique to the context in which the term is described	Mandatory
Definition (English)	The English translation of the originally entered definition.	• Text field • Definition does NOT contain the term itself • Follows the most common spelling and punctuation • Should provide an accurate translation of the originally entered definition	Optional
Explanation	An addition to the definition that clearly describes the term.	• Text field • Explanation does NOT contain the term itself • Follows the most common spelling and punctuation • Unique to the context in which the term is described	Optional
Example	Examples of the term.	• Text field • Follows the most common spelling and punctuation • Unique to the context in which the term is described	Optional

Captured metadata

Explanation of the fields (2/6)

Metadata	Description	Guideline	Mandatory
Data Manager	Responsible Data Manager for this specific term.	• Selection list containing the specific Data Manager responsible for this specific term. • Show only the responsible role; not the name. • Actual overview	Mandatory
Data Domain	Responsible domain for this specific term.	• Selection list containing the specific Data Domain responsible for this specific term. • Actual overview	Mandatory
Language	The language in which a term has meaning.	• Selection list with values: NL (Dutch), EN (English), DE (German), and FR (French).	Mandatory
System of Record	Application where the truth is recorded and from which distribution to receiving systems and applications takes place.	• Selection list containing all systems and applications used within PostNL defined as Golden Records. • May have the same value as the SoE (System of Entry). • If the term has lineage up to the technical layer, this field is mandatory.	Optional
System of Entry	Application where the data is initially entered or generated.	• Selection list containing all systems and applications used within PostNL for data entry. • May have the same value as the SoR (System of Record). • If the term has lineage up to the technical layer, this field is mandatory.	Optional

Captured metadata

Explanation of the fields (3/6)

Metadata	Description	Guideline	Mandatory

Data Labelling	Indicates the label applicable to data protection.	• Selection list with the values:
		  • **Public*:** Corresponds to data with 'none' confidentiality value
		  • **Internal**: Corresponds to data with 'low' confidentiality value
		  • **Confidential****: Corresponds to data with 'medium' confidentiality value
		  • **Secret**: Corresponds to data with 'high' confidentiality value
		• Must be aligned with and comply with the rules and principles of the PostNL Cybersecurity Office.
		• ** If the column contains PII data, it should be labelled with minimum "Confidential".
			Optional

Personally Identifiable Information	Indicates whether it contains personal data.	• Selection list with the values:
		  • Yes, does contain Personally Identifiable Information.
		  • No, does not contain Personally Identifiable Information.
		  • Uncertain if it contains Personally Identifiable Information.
		  • Unknown.
			Mandatory

PII Type	Indicates the type of personal data it contains.	• Multiple-choice selection list with the values:
		  • Business Partner
		  • Consumer
		  • Employee
		• Can be expanded further in the future.
		• If 'Personal Identifiable Information' is filled in, this field is mandatory.
			Optional

Captured metadata

Explanation of the fields (4/6)

Metadata	Description	Guideline	Mandatory
CIA Classification	The classification of data and related systems is determined by the impact on PostNL when the requirements for Confidentiality, Integrity, and Availability are not met.	• Selection list with the values Baseline and Above Baseline. • When the impact of any of the three categories (Confidentiality, Integrity, or Availability) is high, the CIA Classification is Above Baseline. In all other cases, Baseline is sufficient. • Must be aligned with and comply with the rules and principles of the PostNL Security Office.	Optional

Confidentiality	What is the impact when there is unauthorized disclosure of information.	• Selection list with the values: low, medium, and high:
		  • **Low:** The loss of Confidentiality, Integrity, or Availability is expected to have a limited adverse effect on business processes, assets, or individuals.
		  • **Medium**: The loss of Confidentiality, Integrity, or Availability is expected to have an adverse effect on business processes, assets, or individuals.
		  • **High**: The loss of Confidentiality, Integrity, or Availability is expected to have a severe or catastrophic adverse effect on business processes, assets, or individuals..
		• Must be aligned with and comply with the rules and principles of the PostNL Security Office.
			Optional

Integrity	What is the impact when there is unauthorized alteration or destruction of information.		Optional
Availability	What is the impact when there is a disruption in access for the use of information or information systems.		Optional

Captured metadata

Explanation of the fields (5/6)

Metadata	Description	Guideline	Mandatory
Business Ruling	The business rules that apply to a term.	• Text field • Example: Dutch Postal Code: Consists of four digits and two letters. The first digit cannot start with 0. The letter combinations 'SS,' 'SD,' and 'SA' are not used.	Optional
Standard Value Format	The standard format that applies to a term.	• Text field • Specifies the format in which the term should be filled out. • Can be filled in as a regular expression. • Example: Dutch Postal Code: /^[1-9][0-9][0-9][0-9][A-Z][A-Z]$/gm	Optional
Abbreviation	Abbreviation of the term.	• Text field • If an abbreviation is used as a term, it should also be entered as a new term.	Optional
Source	Indicates what or who (function/role) is the source of the entered metadata.	• Text field • Can contain one or multiple sources. • Can refer to functions, roles, and/or systems.	Optional

Captured metadata

Explanation of the fields (6/6)

Metadata	Description	Guideline	Mandatory
Synonym	The relationship with terms that have a similar or identical definition within the same Data Domain.	• Has a relationship with another term in Collibra. This requires that the term with which a relationship is to be established exists in Collibra. • Can only exist with other terms on the same layer. • The relationship must be indicated for both term A and term B.	Optional
Translation*	The relationship with terms entered in another language within the same data domain.	• Has a relationship with another term in Collibra. This requires that the term with which a relationship is to be established exists in Collibra. • Can only exist with other terms on the same layer. • Is only applicable to the business layer. • The relationship must be indicated for both term A and term B.	Optional
Relations with other layers	The relationships to other terms will be explained on separate slides.	• Has a relationship with another term in Collibra. This requires that the term with which a relationship is to be established exists in Collibra.	Optional

As a result of the DGB decision, the initial number of fields has been expanded with two optional fields:
- Definition_EN: This field will contain the English translation of the definition.
- 'Translation' Relationship: This field establishes the relationship between the NL (Dutch) business term and the English version.

In addition, the current 'Synonym' field is no longer used to establish the relationship between NL (Dutch) and EN (English) business terms. Instead, the 'Relation' field is used.

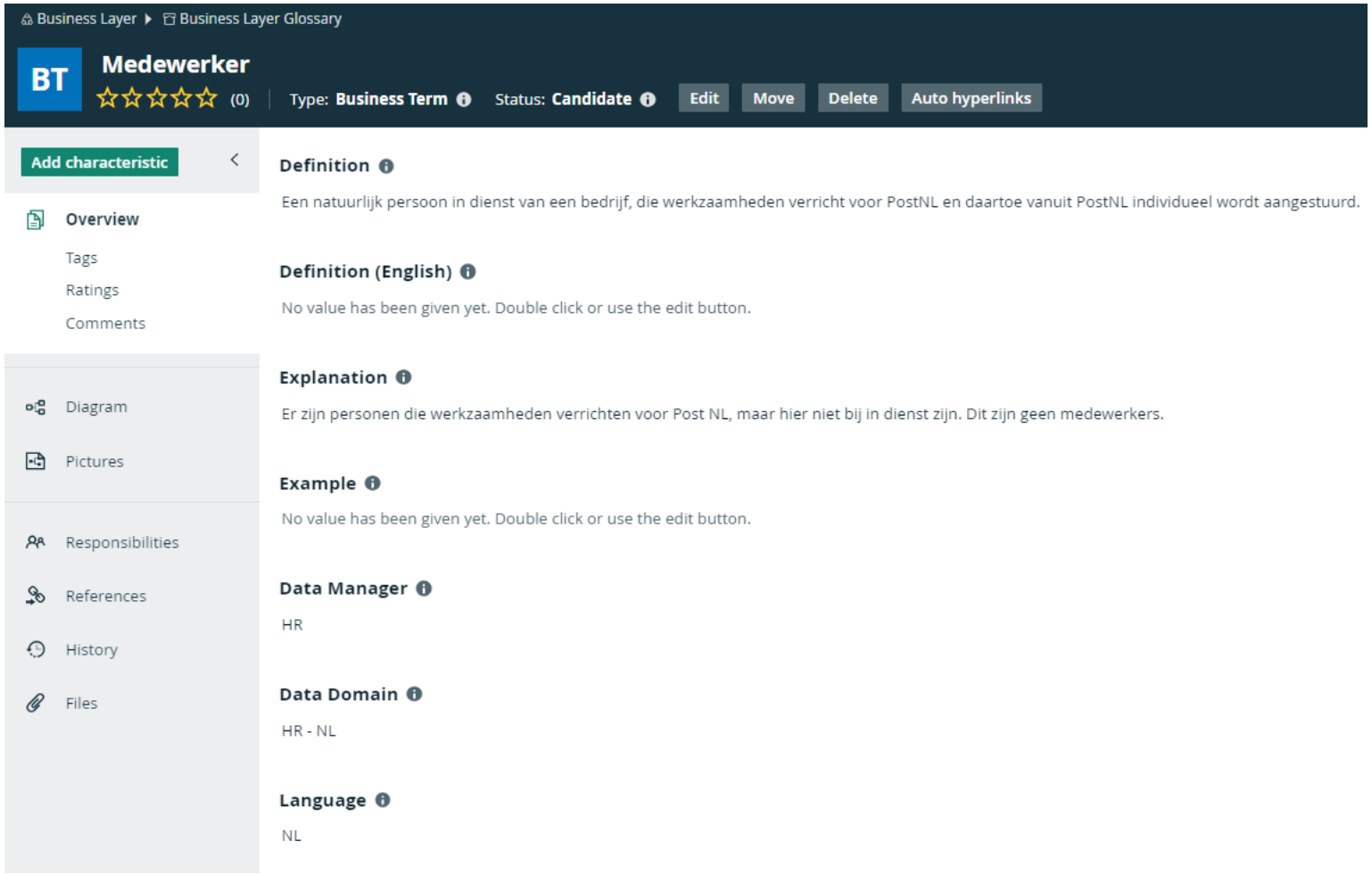
For most Business Units (BU's), this means that they enter all terms only in Dutch and fill in the 'Definition_EN' field with the English translation. For E-Commerce Operations and BU's dealing extensively with English speakers, this means that all business terms are entered in both Dutch and English, and they are linked through the 'Translation' relationship. It is the responsibility of the respective DMO (Data Management Organization) to maintain this relationship.

It is mandatory that the metadata is entered in the same language as the business term. This means that an English term receives English metadata, and a Dutch term receives Dutch metadata.

Captured metadata

An example

- A few example of captured metadata for the business term ‘Medewerker’.



Relations

Relations with entities on the logical layer (1)

- When a business term also exists on the application and system level (technical layer), it should be indicated through a 'relationship.'
  - Example: The business customer found in SalesForce or the retail location found in BLS.
- In most cases, it is not possible to directly link a business term to a technical term (tables and fields). For this purpose, the logical layer is needed. It contains logical terms that serve as an intermediary to show the relationship. Additionally, there may be system fields at the technical layer that are difficult to understand. In these cases, the logical term provides the solution.
- A business term can be linked to a Data Entity or a Data Attribute on the logical layer.
  - Data Entity is the logical counterpart of a table.
  - Data Attribute is the logical counterpart of a field.

Relations

Relations with entities on the logical layer (2)

- In consultation with a Data Architect, Domain Architect, or Data Information Analyst, a Business Layer Curator determines what is needed to display the complete lineage.
  - Relationships can be established with Data Entities and Data Attributes imported from Aris.
  - Relationships can be established with new Data Entities and Data Attributes that are not in Aris.
- It may happen that no new logical terms are needed to link the business layer to the technical layer. In this case, the business term is related to a Data Entity at the logical layer that has the same name and metadata as the business term.

Relations

Relations with enterprise objects

- As previously described, a business term can be the same as a business object or a data object. In these cases, the business term has the same name and/or definition, but the responsibility for it remains with the business.
- In these cases, the business term can refer to the object on which it is based. You do this by establishing a relationship between the business term and the object. Even when the business terms do not exactly match the objects, it is still possible to refer to an object.