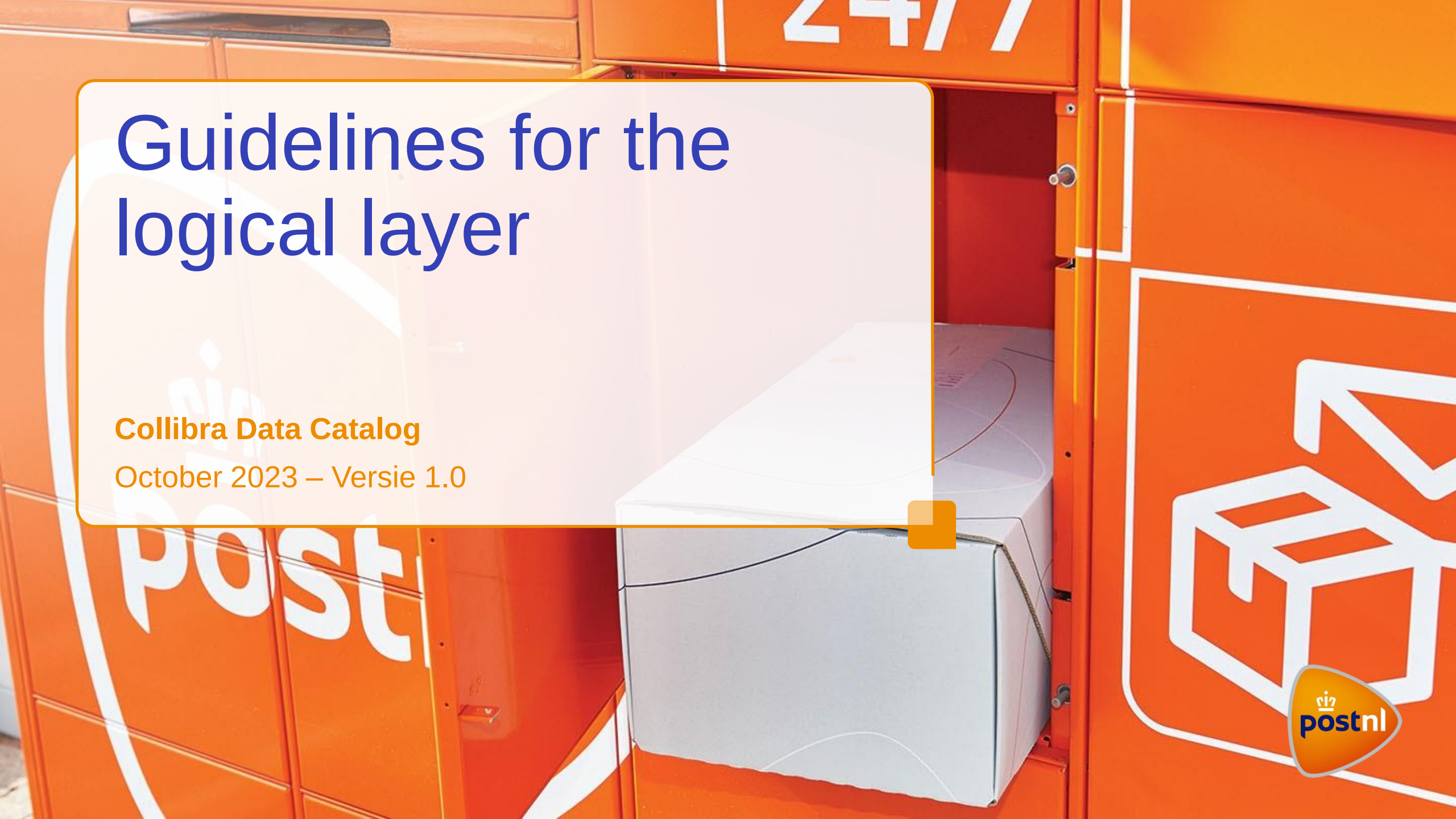


Guidelines for the logical layer

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

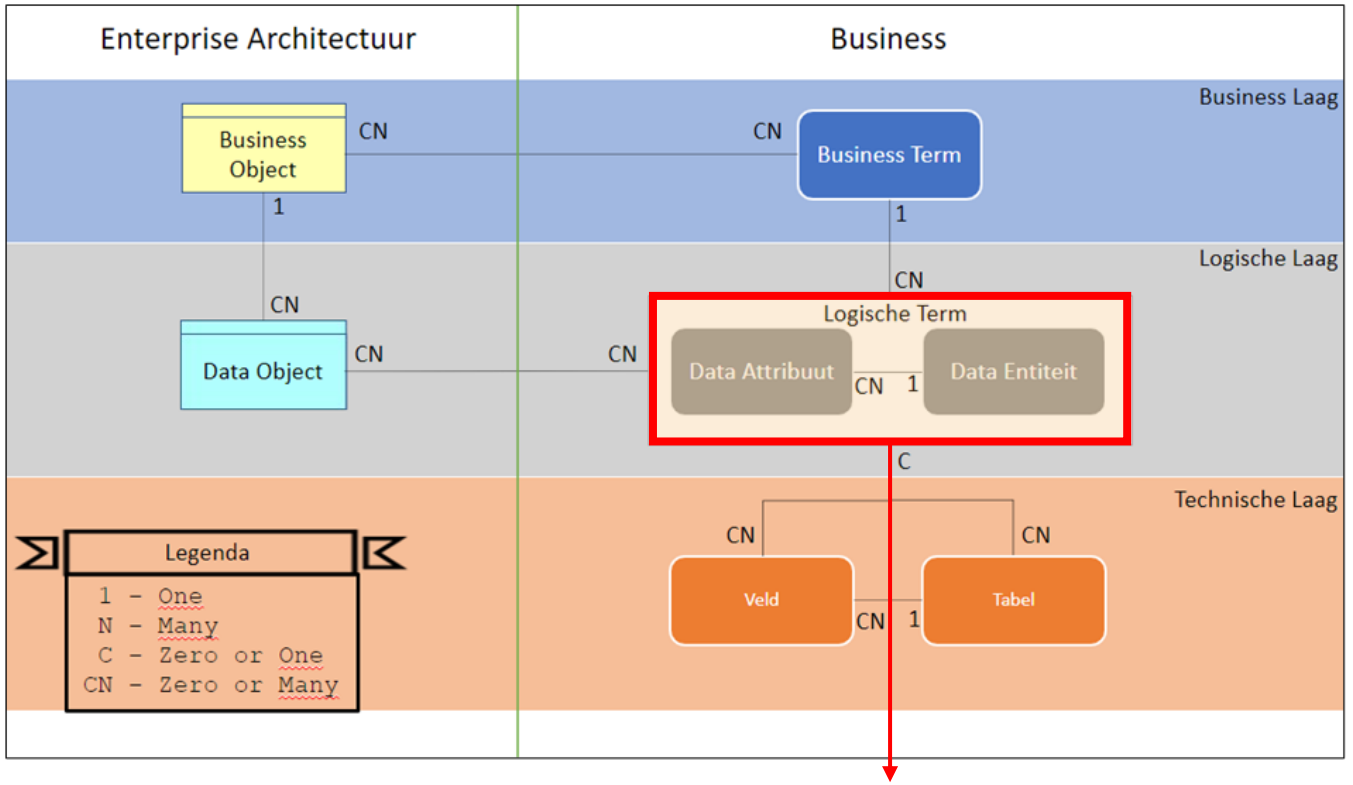
Three layers: business, logical and technical

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.

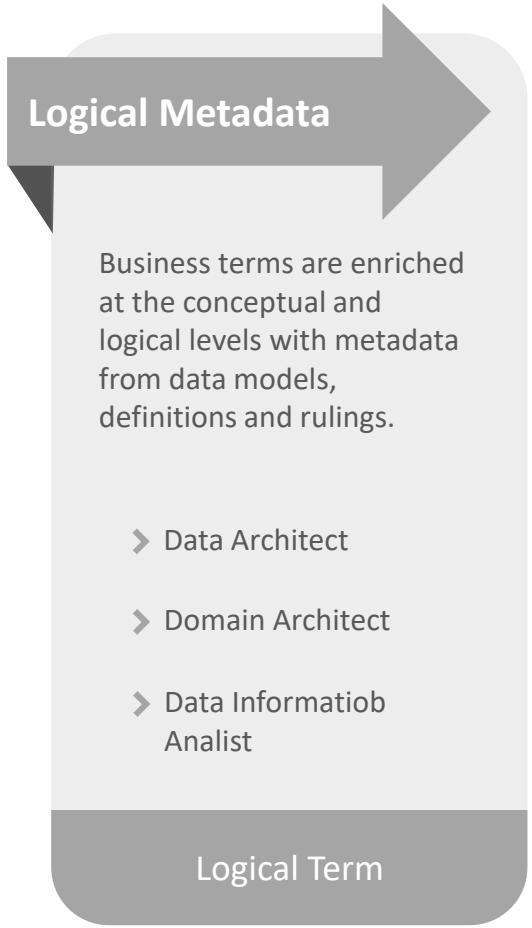


The focus in this document is specifically on the 'Logical Layer.' In other words, what is a logical term and what is the associated metadata?

Logical terms

Basic principles (1/3)

- *"A term used to establish a relationship between the Business terms and the tables and fields as they exist in the systems."*
- This involves the words and terms that can be used to simplify technical table and field names.
- In the Data Catalog, a logical term can arise when there is a need to link a business term to the technical layer (or just the logical layer). When there is no relationship between the business layer and the other layers, there is no need to create a logical term.
- If a business term closely aligns with the technical layer, you can choose to keep the business term and the logical terms the same. This applies to the name and potentially the associated metadata as well.



Logical terms

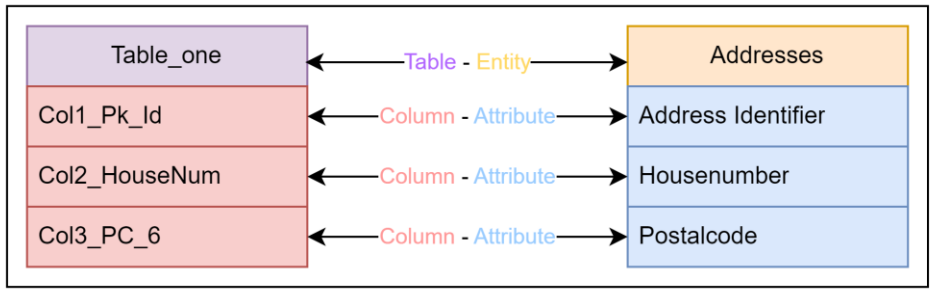
Basic principles (2/3)

- A logical term can take two forms: in the form of an entity or in the form of an attribute.
- Logical terms are used in the logical data model. This model describes the structure and relationships between the logical terms. Logical data models are documented in ARIS and are modeled using the Entity Relationship Diagram (ERD).
- The logical data model can be seen as a simplification/clarification of the existing physical data model but is technology-independent. However, the LDM (Logical Data Model) can serve as a guideline for how the actual physical data model will look.
- Because physical data models often use system names for tables and columns that are too technical for 'ordinary' readers, the need for the logical layer is immediately indicated.
- **Note**: Entities and attributes are loaded into ARIS separately from each other. The mutual relationship between entity and attribute is not included. ARIS remains the System of Record (SoR) for the models.

Logical terms

Basic principles (3/3)

- A logical term can be a data entity or a data attribute.
  - A data entity is the logical counterpart of a table in the technical layer.
  - A data attribute is the logical counterpart of a field in the technical layer.
- Just as a table consists of one or more fields, an entity consists of one or more attributes.



- Because a data entity consists of at least one or more data attributes, a relationship can only be established from the business layer to an entity. From there, the associated attributes are linked.
- If an entity consists of a single attribute, both can have the same name.

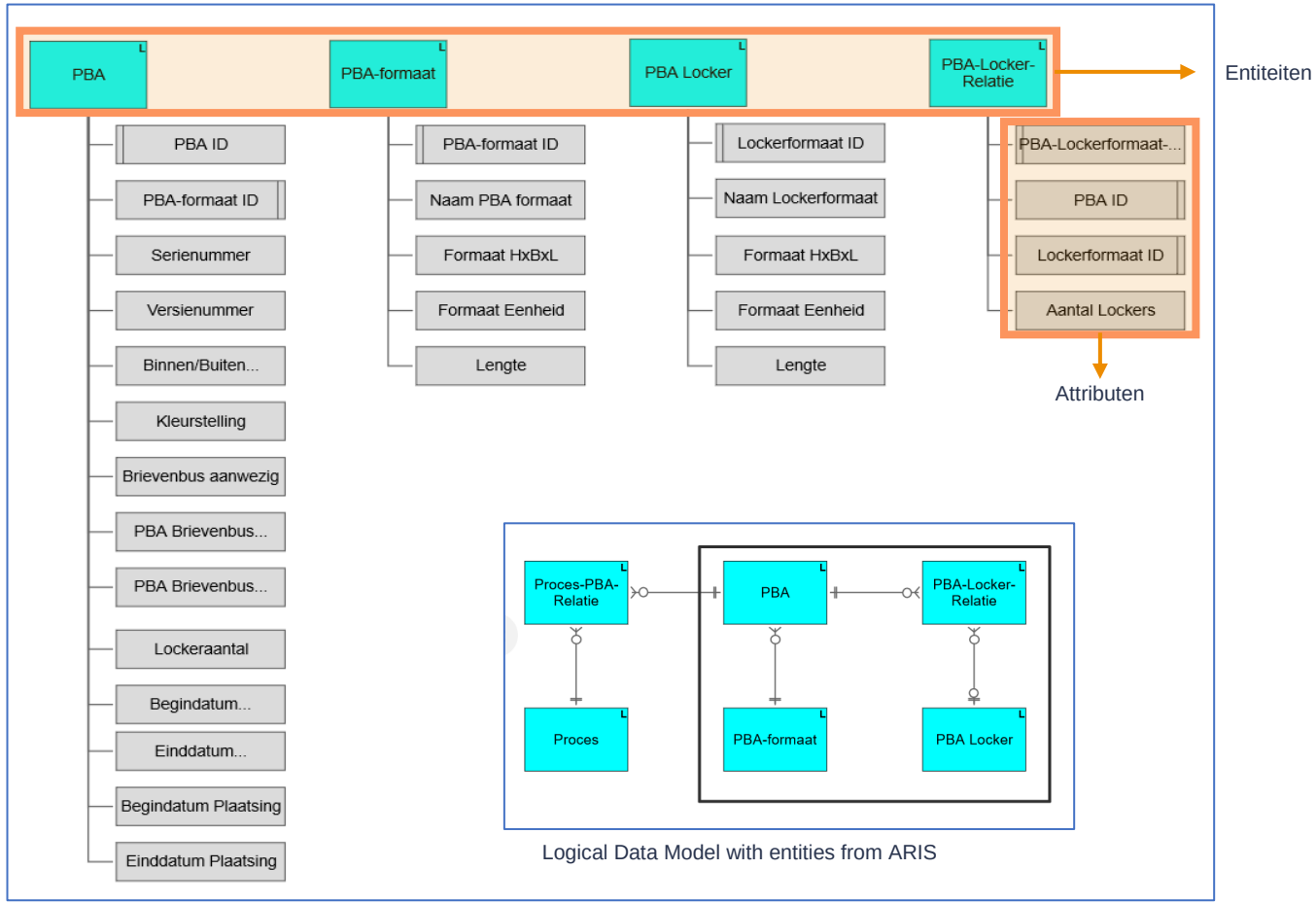
Logical terms

Creation

- An initial batch of logical terms is imported from ARIS. These terms were previously created for various data models, systems, and applications.
- This initial batch of logical terms will be insufficient to populate the entire logical layer because:

1. Not every logical term is present in ARIS.
2. Systems may have been further developed or modified, but these changes may not have been reflected in ARIS yet.

- For this reason, it is also possible to manually add logical terms to the logical layer.



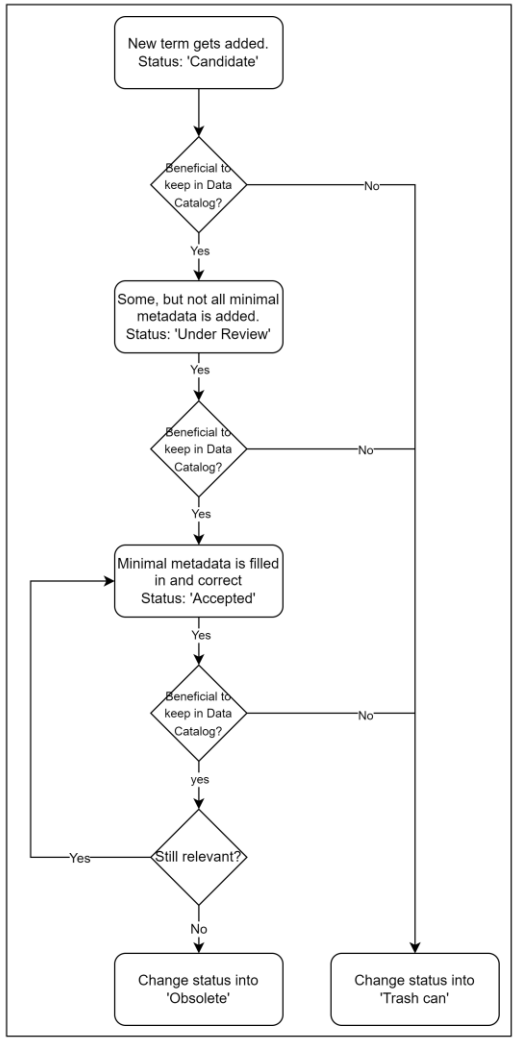
Example of entities with their corresponding attributes from ARIS.

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

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

- The red-colored fields are mandatory.
- The black-colored fields are optional.
- The blue-colored fields depict the relationships between the three layers and/or business terms.
- Fields 10 through 12 are currently out of scope.
- Fields 22 through 25 are automatically generated by Collibra itself (standard functionality).

Each of these fields will be further explained on the following slides.

Captured metadata

Explanation of the fields (1/5)

Metadata	Description	Guideline	Mandatory
Definition	A clear description of the 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
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
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 (2/5)

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 (3/5)

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 (4/5)

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
Data Type (attribute)	Indicates the type of data expected in the data attribute.	• Selection list with the values: Text, Date, DateTime, Numeric, Code, Measure, Amount, Identifier, Indicator, Quanity, Value, Percent.	Optioneel
Language	The language in which a term has meaning.	• Selection list with values: NL (Dutch), EN (English), DE (German), and FR (French).	Mandatory
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
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

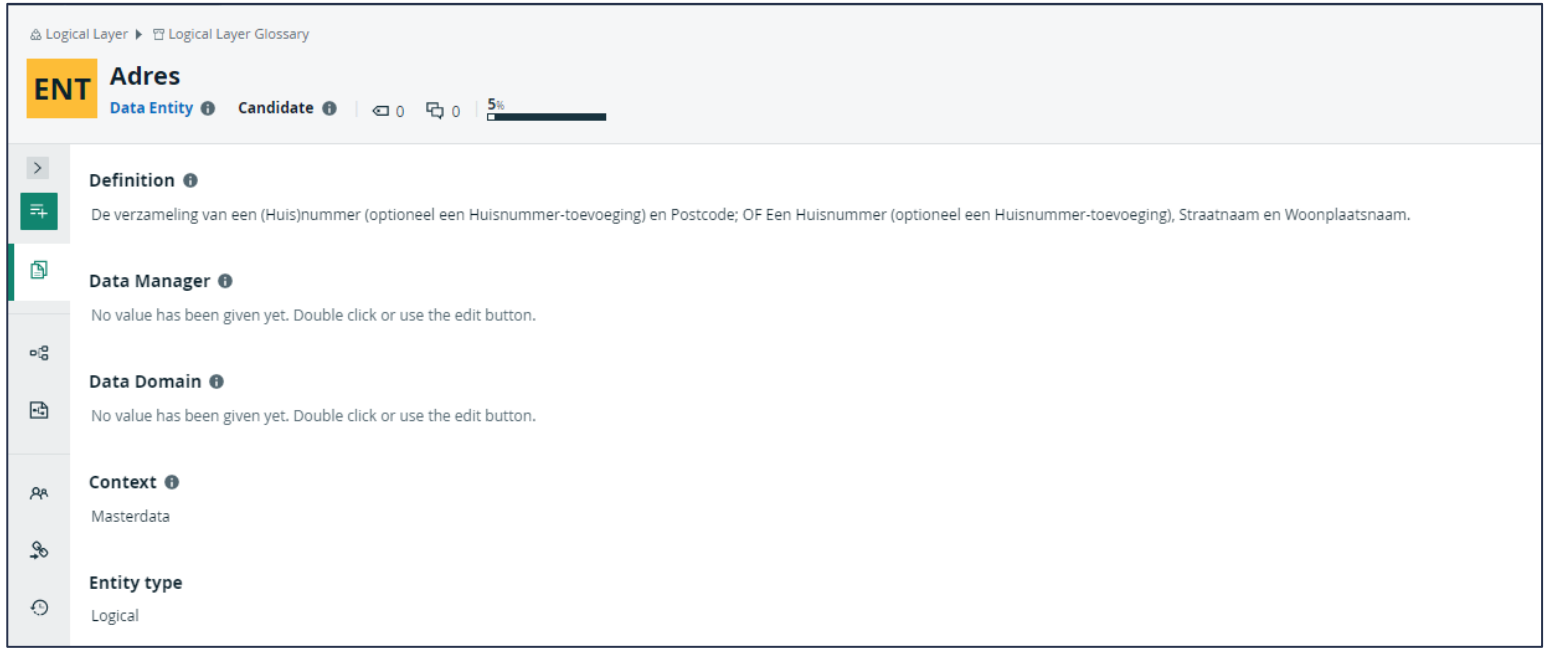
Explanation of the fields (5/5)

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
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

Vast te leggen metadata

Een voorbeeld

- A few examples of captured metadata for the logical term ‘Adres’.



- A term can have a different context in various appearances.

☐	Adres	Het adres van het gebouw waar PostN...	Masterdata
☐	Adres	Dit object bevat alle unieke adressen ...	Business Partner
☐	Adres	De verzameling van een (Huis)numme...	Masterdata
☐	Adres	De aanduiding van een bepaalde locat...	Commercieel Product
☐	Adres	Een adres is de aanduiding van een be...	Masterdata
☐	Adres	Adres volgens Basis Administratie Geb...	Masterdata
☐	Adres	De verzameling van gegevens voor de ...	Business Partner

Relations

Relation *from* the business layer

- When a Business term also exists at the application and system levels (technical layer), this should be indicated with a relationship.
- Example: The business customer found in Salesforce or the retail location found in BLS.

- It is not possible to directly link a Business term to a Technical term (tables and fields). The logical layer is required for this purpose. The logical layer 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 comprehend. In such cases, the logical term provides the solution.

- A logical term can be the same as a Business Term, Business Object, or Data Object. This can even result in identical definitions and other metadata. However, Architecture remains the owner and responsible party for Business and Data objects, including their associated metadata.

Relations

Relation with technical terms on the technical layer

- As described on the previous slide, it is not possible to directly link a Business term to a Technical term (tables and fields) or vice versa.
- When there is a need to make the technical implementation at the application and system levels (technical layer) transparent, a relationship must be created at the logical layer. Example: The technical implementation of a process location.
- A logical term can be linked to technical fields and technical tables on the logical layer.
  - Data Entity is the logical counterpart of a table.
  - Data Attribute is the logical counterpart of a field.
- Only when a system is loaded at the technical layer can it be linked to the related term (Entity or Attribute) in the data catalog.

Relations

Relation with Enterprise Objects

- Similar to business terms, logical terms can also have a relationship with a business object or a data object. In these cases, the logical term has the same name and/or definition, but responsibility remains with the business.
- In these cases, the logical term can reference the object it is based on. This is done by creating a relationship between the logical term and the object. Even when the logical terms do not exactly match the objects, a reference can still be made to an object.