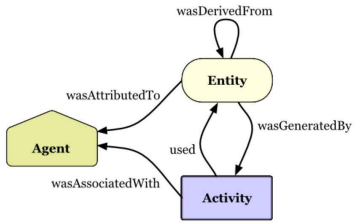
Overview

- Design a standard provenance model for the PDS4 IM that improves interoperability with other archives. A standard model, PROV-DM: The PROV Data Model, is available from the W3C Provenance Group. It was developed to provide the foundation for trust on the Web.
- Provenance is crucial for establishing the authenticity of science data. It is the documented history of the data, detailing its origin, ownership, and any changes made.
- Provenance helps to validate the authenticity of the data by providing information about where and when the data was created. Well-documented provenance also helps scientists determine the scientific value of the data. In some cases, it is essential for resolving legal disputes over ownership of and access to the data.
- PDS4 should have a general provenance model so that data providers can capture provenance for the science data that they are submitting to the archive.

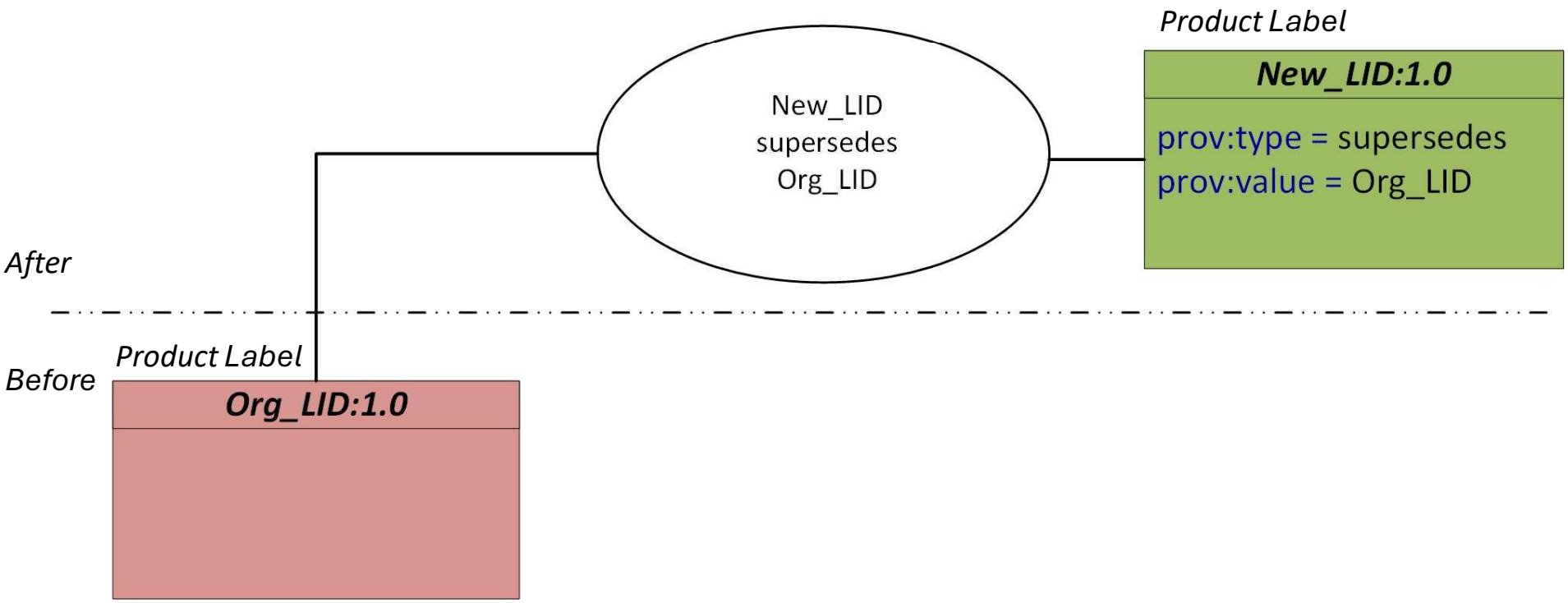


Preferred: +2

Superseded LIDs

Use Case 3 – Uni-Directional

LDD in first product with new LID

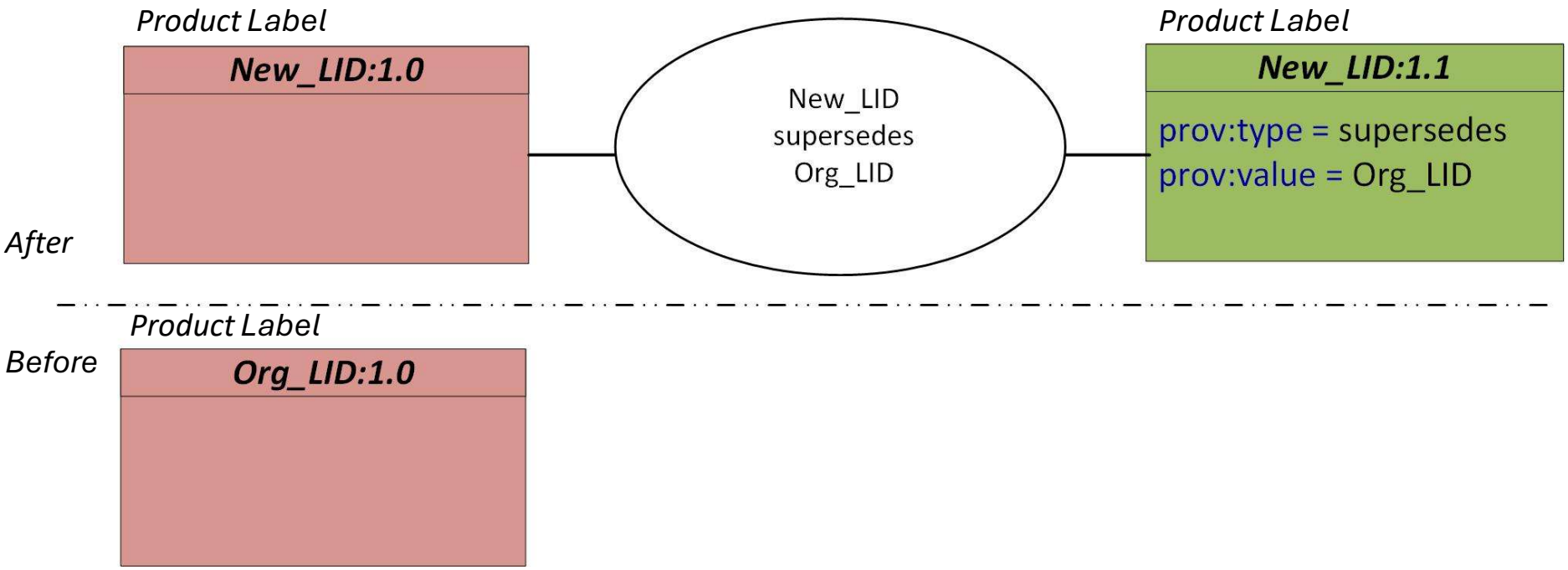


Preferred: +1

Superseded LIDs

Use Case 4 – Uni-Directional

No LDD in first product with new LID



Example

prov:SupersededLID
  prov:title = Insight superseded LID
  prov:description = An Insight product superseded LID and its replacement

  prov:Entity
    prov:title = urn:nasa:pds:insight-ifg-mars...sol0212-20190702t010324-20190703t014257...
    prov:description = Product LID change due to change in product start/stop times ...

    prov:Attributes
      prov:attribute = Supersedes
      prov:value = urn:nasa:pds:insight-ifg-mars...sol0212-20190702t011217-20190703t015130...

    prov:Attributes
      prov:attribute = Reason
      prov:value = Replacement

Reasons for LID Supercession

How do we capture the reason for the LID supersession in the provenance?

LID supercession should no longer happen but some possible reasons for consideration are:

1. Replacement: The original LID no longer matches the product/entity.
2. Duplication: An error, system issue, or oversight led to a duplicate product.
3. Merged: Multiple products/entities were combined into one to reduce redundancy.
4. Obsolescence: The original product/entity is no longer in use.
5. Policy Changes: New policies require changes to the LID structure.

LID Logic

1. In PDS4, we define the concept of a "product" and allow for multiple versions of that product.
2. Each version of a product can be referred to as a PDS4 Product.
3. Every PDS4 Product is assigned a LID (Logical Identifier) and a VID (Version Identifier), which together form a unique identifier known as the LIDVID.
4. The LIDVID serves as the unique identifier for a PDS4 Product within the PDS system.
5. This structure implies that a LID identifies a unique **set** of PDS4 Products, representing all versions of a specific product.
6. Likewise, a different LID represents a distinct **set** of PDS4 Products for a different product.
7. If two PDS4 Products have different LIDs but are considered versions of the same product, this situation logically results in two **sets** of versions of that product. This situation is inconsistent and conflicts with point 5.
8. The purpose of the "superseded" LDD (Local Data Dictionary) is to extend the Information Model (IM) to help resolve the inconsistency and communicate the relationship between the two **sets** to the PDS4 system, services, and software.