



# Policy for persistent identifiers in the data space

<b>Revision</b>	Version 1.1
<b>Last updated</b>	3 December 2024
<b>Author(s)</b>	Hugo Manguinhas, Maša Škrinjar, Nuno Freire, Ariadna Matas
<b>Contributor(s)</b>	Europeana Foundation staff
<b>Dissemination Level</b>	Public

## Version history

Revision no.	Date	Person	Organisation	Description
V0.1	18 Mar 2023	Hugo Manguinhas, Maša Škrinjar, Nuno Freire	Europeana Foundation	First draft
V0.2	19 Apr 2023	Hugo Manguinhas, Maša Škrinjar	Europeana Foundation	Second draft after review
V1.0	20 Apr 2023	Antoine Isaac, Nuno Freire, Ariadna Matas	Europeana Foundation	Final review
V1.1	3 Dec 2024	Hugo Manguinhas, Maša Škrinjar, Antoine Isaac, Nuno Freire	Europeana Foundation	Integrate feedback from survey and EAF

*How to cite this:*

*Policy for persistent identifiers in the data space (V1.1). Europeana Initiative. December 2024.  
URL: <https://pro.europeana.eu/post/policy-for-persistent-identifiers-in-the-data-space>.  
CC BY-SA*

# Table of contents

<b>Introduction</b>	<b>4</b>
<b>Purpose</b>	<b>4</b>
<b>Principles</b>	<b>5</b>
1. Characteristics of a PID	5
2. Assigning PIDs to resources and managing change	5
3. Dissemination, use and interoperability of PIDs	6
4. Sustainability and Governance of PIDs	6
<b>Definitions used in this document</b>	<b>7</b>
<b>References</b>	<b>8</b>

## Introduction

When different organisations use different identifiers for the same cultural heritage resource, or even within the same organisation, different identifiers are employed over time, this results in challenges when attempting to unambiguously identify and locate the resource and its associated information. URLs pointing to resources are vulnerable to becoming broken or inaccessible, leading to link rot, which disrupts all forms of (re)use and negatively impacts traffic towards the organisation's website as well as its SEO ranking among search engines. Using persistent identifiers to provide stable references to resources is a crucial step towards ensuring reliable identification, location and access. Persistent identification plays a vital role in enhancing the reuse of cultural heritage resources by providing more change resistant and citable identifiers that facilitate discovery and proper attribution.

The adoption of persistent identifiers by cultural heritage institutions with digital collections shows commitment towards interoperability and data access that in turn builds trust in the institution and the data it creates. This practice supports the dissemination and reuse of cultural heritage objects across academic, cultural, scientific, and commercial domains, helping to make them more accessible to audiences across the data spaces and contributing to the broader goals of knowledge exchange and innovation.

**Definition:** Persistent identifier (PID) is an association between a sequence of characters and a specific resource. The term "persistent" refers to the identifier's role in ensuring continued access to the resource and the data associated with it for the foreseeable future. Organisations identifying resources can implement in-house PID solutions, or rely on existing PID service providers. Some of the most commonly used PIDs systems are [Archival Resource Key \(ARK\)](#), [Digital Object Identifier \(DOI\)](#), [National Bibliography Numbers \(NBN\)](#), [Persistent Identifiers for eResearch \(ePIC\)](#).

## Purpose

This document is intended for participants in the European data space for cultural heritage, especially data partners that are assigning identifiers to resources being shared in the data space.

The persistence of identifiers relies on a combination of technological and organisational endeavours. It depends on (technical) services built around them and ongoing commitment to sustainability from operators of PID systems or infrastructures. At the same time it is equally important for organisations or individuals overseeing PID implementation and its management to demonstrate responsibility and commitment.

In the context of the data space, it is worthwhile acknowledging that collaborating with various stakeholders whose identification practices and infrastructures are not entirely within the data space steward's control, entails a level of uncertainty. To ensure better

trustworthiness of the data space, this policy establishes clear expectations for data partners seeking to implement and contribute (P)IDs. Their identifiers will be acknowledged as persistent as long as they adhere to the principles outlined in this document and demonstrate compliance through transparent documentation of their approaches in their own institutional policies, which are understood as an important indicator of institution's commitment to PID implementation and its continuous administration. At the same time, being able to develop an understanding of the procedures and workflows implemented by data partners is crucial for the data space steward and other stakeholders, as it enables the mitigation of potential risks and, ultimately, enhances the trustworthiness of the PIDs available in the data space.

## Principles

### 1. Characteristics of a PID

**P1:** A PID is a URI that follows a formally defined and recognisable identifier scheme capable of supporting an identification space large enough to accommodate existing and future needs.

**P2:** A PID is as opaque as possible, meaning that it contains the least information possible about the resource it identifies.

**P3:** A PID is unique, meaning that it identifies only one resource, it is not reused to identify other resources, and no other PID exists within its identifier scheme for the same resource.

**P4:** A PID is persistent, meaning it is never changed, nor deleted.

### 2. Assigning PIDs to resources and managing change

**P5:** PIDs identify resources that are intended to be stable in their definition.

**P6:** The assignment of PIDs is not restricted to any specific type of resource, whether physical, digital or conceptual, an individual, a subpart or an entire collection or even a version of a resource.

**P7:** When a resource undergoes significant changes to the extent that it becomes a new resource, the PID of the previous version of the resource is not used to identify the new resource.

**P8:** When circumstances result in a resource being deleted, the PID is kept and marked as deprecated.

### 3. Dissemination, use and interoperability of PIDs

**P9:** A PID resolves to a landing page (containing the PID, information about the resource and a means for accessing the resource) or to the resource itself depending on the type of resource, or ultimately to the PID record if neither are available.

**P10:** When a PID is deprecated, its resolution presents a tombstone page where a subset of the information about the resource is maintained and information about the reason for deletion is available.

**P11:** Metadata about the PID is maintained in a PID record and made accessible without restrictions in machine-readable formats.

**P12:** A PID is usable, without any legal, contractual or financial restrictions, in perpetuity to identify its associated resource.

**P13:** Users are informed on how to use PIDs for referring to resources sustainably.

### 4. Sustainability and Governance of PIDs

**P14:** The administration of PIDs is fulfilled by a trustworthy and long-lasting system infrastructure that guarantees quality, security, reliability, availability and performance.

**P15:** A single owner is committed to keeping the information in the PID record and the landing page up to date.

**P16:** The information about PID ownership is made public.

**P17:** A new PID is not assigned to the resource when the ownership of the resource changes while its content remains the same.

**P18:** A PID policy is created, maintained and made available by the owner of the PID, which defines the governance mechanisms and system infrastructure that guarantee compliance against the principles defined in this document.

**P19:** A PID policy is explicit, unambiguous and stable.

**P20:** A PID policy is documented and disseminated to all (internal & external) stakeholders.

## Definitions used in this document

- **Landing page:** a webpage that displays the PID, its ownership, metadata about the resource, and access to the resource itself.
- **PID assignment:** the process of associating (existing) PID with the resource it is intended to identify.
- **PID owner:** an individual or organisation who mints (process of generating a new PID according to the (pre)defined scheme) and assigns PIDs to their resources and takes the ownership of the PIDs. They are responsible and accountable for the governance of a PID, such as keeping the information in the PID record and the landing page up to date, as well as maintaining the PID policy and ensuring compliance with the principles set by this document.
- **PID policy created by a PID owner:** a policy specifying the conditions and scope of PID assignment and non-assignment (ie. which types of resources and granularity are covered), lifecycle and version management, indicating what information is kept available upon deprecation of the resource, as well as resolution, limitations, scheme and adherence to standards.
- **PID record:** a machine readable record containing administrative information for the PID and a subset of the metadata of the resource that is openly available. It is used to facilitate automation of processes and interoperability within and across PID service providers, as well as other systems.
- **PID scheme:** a comprehensive set of rules and standards defining various aspects of PIDs, such as their format and syntax. Overall, it provides a framework for ensuring the desired attributes of PIDs, like uniqueness and opacity.
- **PID service provider:** an entity that is operating and offering PID services and infrastructure (e.g. registration, management, resolution services) for PIDs within a specific PID system, but does not typically have ownership rights over PIDs.
- **PID system:** a framework that comprises technological infrastructure and processes designed for registering, managing and resolving PIDs.
- **PID user:** an individual or organisation intending to use a PID to refer (or resolve) to a resource.
- **Resolution:** mechanism through which resources are retrievable by their (persistent) identifier using a standard protocol.
- **Tombstone page:** a variant of a landing page that informs users about the unavailability of the resource, the reasons for its unavailability and guidance for future reuse.

## References

- Netwerk Digitaal Erfgoed [Persistent Identifier Guide](#)
- CEOS [Persistent Identifiers Best Practices](#) (Earth Observation mission data)
- Group of European Data Experts (GEDE) [PID report](#) (seems to be from CLARIN)
- GBIF [A beginners guide to PIs](#)
- CD2H [Best practices for Using Identifiers](#)
- Paper [Identifiers for the 21st century](#)
- Catch Plus project [report](#)
- BL PI [policy document](#), [v2 document](#)
- BnF PI [policy document \(FR\)](#)
- [PI Policy EOSC](#)