

# British Library Persistent Identifier Policy

## 1. Purpose

The purpose of this policy is to outline the approach to the use of identifiers and persistent identifiers (PIDs) in internal and external facing British Library systems, to define a consistent set of requirements for PIDs to use when managing and developing services and to articulate the responsibilities in applying and reviewing these principles.

**Definition:** A persistent identifier provides a long lasting digital reference to an entity, whether it is physical or digital. Persistent identifiers are a core component in providing reliable, long-term digital access to collections.

The British Library encourages the use of PIDs across its collections and collection metadata. It recognises the role PIDs have as a component in sustainable, open infrastructure and in enabling interoperability and use of Library content and resources. PIDs also support the Library's content strategy and its goal of connecting rather than collecting as PIDs support long term and reliable access to resources.

The Library uses multiple classes of PIDs to different extents in different roles. The general roles that the Library may take in regard to a given PID are defined in Section 3. While not all existing services may meet the requirements described in this policy, it provides a benchmark against which they can be measured and aspire to develop.

## 2. Scope and Exclusions

This policy **covers**:

- Identifiers that are used to refer to standard descriptive information about an entity (e.g. item, person, place) or the entity itself, whether used for internal or external identification of entities
- Definition of a set of principles (Section 4) and requirements of PIDs in use and for consideration for new types of identifiers
- Provision of a benchmark against which existing and future identifier use can be measured

This policy **excludes**:

- Definition of namespaces or the Library's approach to them, which is covered separately by the British Library's draft namespace guidance (see Definitions)
- Definition to whom the Library provides PID services, for example DataCite consortium members and ISNI members
- Identifiers which are assigned temporarily in pursuit of Library functions but which are ultimately deleted

## 3. Use of Identifiers at the British Library

The British Library has an interest in a range of PIDs types and schemes. The following describe the roles the Library may take with regard to a given PID:

- U. 1. **Create:** The Library mints these identifiers, uses them to identify entities within the collection and provides them to others for their purposes, e.g. ISNI, ARKs, ISIL, ISSN, DOI

- U. 2. **Contribute:** The Library uses these identifiers, builds services based on them and has a role in their maintenance, but does not create them. Maintenance can include contributing to a central registry managed elsewhere and building services include using the identifier as an index term in Library systems, e.g., LCSH, LC/NACO
- U. 3. **Consume:** The Library uses these identifiers in their workflows but does not create or maintain them, e.g. ISRC, ISBN

While this policy describes the Library's preferred characteristics for a PID, it recognises that the Library cannot dictate the characteristics of all identifiers in use within the Library. Identifiers created within the Library (U. 1) are obliged to meet this policy as far as possible, and preference is given to identifiers that adhere to the principles in Section 4 of this policy for U. 2 and U. 3.

The Library also uses many identifiers within its systems that are not persistent, e.g. shelfmarks and database identifiers, or where the identifiers are not applied uniquely or persistently as desired. While the Library aspires that all entities should also have a PID, this does not mean that existing non-persistent identifiers must be removed or deleted from items; in some cases, records of these identifiers will be maintained.

Where the Library creates identifiers that do not meet the requirements described in this policy, for example due to historical reasons or resource constraints, this should be reviewed as part of ongoing developments. At a minimum, they should adhere to the principles set out in Section 4 of this policy, if not the requirements described in Section 5.

Number of identifiers in Library systems:

- As of August 2021, the number of ARKs held in BL systems is 1,988,794,503, an estimated 1,750,000,000 of those are active
- The Library has created 1,689 DOIs across the Research Repository and its other services
- In 2019, there were 2,426,878 articles indexed in Explore with DOIs associated with them
- In 2020, 5,593,921 of 10,091,682 (55.4%) distinct name strings in the Library's catalogue data were matched against an ISNI record

## 4. Principles

In its use of identifiers, the British Library adheres to the following principles, which describe the qualities PIDs created, contributed or consumed by the Library must have.

### Persistence

- P. 1. A PID must be never be deleted, but may be marked as deprecated if required
- P. 2. A PID must be usable in perpetuity to identify its associated entity
- P. 3. A PID must only describe one entity and must never be reused for different entities
- P. 4. A PID must have established versioning processes and procedures in place; these may be defined locally by the Library as a creator or by the PID provider
- P. 5. A PID must have established governance mechanisms, such as contracts, in place to ensure the standards of use of the PID are met and continue to be met
- P. 6. A PID must resolve to metadata about the entity available in both a human and machine readable format
- P. 7. A publicly accessible PID must be resolvable via a global resolver (see Definitions)
- P. 8. A PID must have an operating model that is sustainable for long-term persistent use

### Established user community

- P. 9. A PID must have an established user community, which has adopted it as a standard, either through an organisation such as the International Organization for Standardization (ISO) or as a *de facto* standard through widespread adoption; the Library will support and develop the use of new types of PIDs where there is a defined and recognised use case which they would address

### Interoperable

- P. 10. A PID must be able to link with the other identifiers in use at the Library through open metadata standards and the capability to cross-reference resources

### New PID types or new use

- P. 11. New types of PIDs should only be considered for use in the Library where there is a defined need which cannot reasonably be met by a combination of PIDs already in use
- P. 12. Any new PID type used by the Library should meet the requirements described in this policy
- P. 13. Where a PID type is emerging and does not have an established community, the Library can seek to influence its development in line with principles for open and sustainable infrastructures

## 5. Requirements for creation, contribution and consumption of PIDs at the British Library

These requirements outline the Library's responsibilities in using PID services and creating PIDs. While the Library uses identifiers which do not meet all of these requirements, they are included for future work and developments.

- R. 1. The Library aspires to assign PIDs to all resources within its collections, both physical and digital, and associated entities, in alignment with the guiding principles of the Library's [content strategy 2020-2023](#).
- R. 2. The Library has varying levels of involvement in different PID schemes, but all PIDs created by the Library must meet the requirements described in this section and the Library prefers the use of PIDs which meet the principles outlined in Section 4
- R. 3. Identifiers created by the Library must have an opaque format, i.e. not contain any semantic information within them, to ensure their longevity
- R. 4. A PID must resolve to information about the entity to which it refers
- R. 5. The Library must have a process to specify the granularity at which PIDs are assigned and how relationships between PIDs for component and overarching entities are managed
- R. 6. The Library must have a process to manage versioning including changes, merges and retirement of entities
- R. 7. Standard descriptive information about an entity, e.g. creator, should have a PID
- R. 8. All metadata associated with a PID should comply with Collection Metadata Licensing Guidelines
- R. 9. Where a PID referring to a citable resource resolves to a webpage, that webpage should display a suggested citation including the hyperlink to the PID to encourage ongoing use of the PID outside the Library

## 6. Responsibilities

- This policy is co-owned by Collection Metadata and Architecture in Technology
- Collection Metadata assumes responsibility for the choice of identifiers and their deployment within cataloguing systems
- Architecture is responsible for creating and maintaining the Library's infrastructure so that identifiers can be created and retain persistence in perpetuity
- This policy is approved by the Collection Metadata Strategy Group
- Other departments (e.g. Content and Research Services, curatorial teams and Digital Research) consume and provide PID services, and therefore have an interest in the use and management of PIDs and should be consulted in the adoption or changes to the Library's use of PIDs
- In most instances, the business owner of the collection system within which an entity is recorded assumes responsibility for the continued resolution of PIDs within that system in conjunction with colleagues from Technology
- When developing new services it is the responsibility of the project's Senior Responsible Owner to ensure that this policy is followed
- The Library's Architecture Review Board will monitor technical compliance to this policy
- The business user group for a particular system, where it exists, e.g. IAMS (Integrated Archive and Manuscript System) User Panel, will oversee operational and process compliance to this policy
- The Collection Metadata Strategy Group will oversee technical and operational compliance where there is no business user group and will provide an escalation point for user groups where needed

## Review

This policy will be reviewed by the Collection Metadata Strategy Group and other stakeholders in 2023 (2 years after approval)

## Definitions

- **Entity:** An entity refers to a 'thing', which may include a resource, a person, a corporate body, a location etc.
- **Global resolver:** See Resolvability. A global resolver allows an entity to be resolved through a single domain internationally, i.e. you do not need to know the domain of the host institution. Examples include "n2t.net" for ARKs, and "doi.org" for DOIs.
- **Machine readable:** Where the identifier and its metadata are structured in a way that can be processed by computers with no human interaction.
- **Namespace:** A group of related elements that each have a unique name or identifier. The domain name syntax
- **Persistent identifier (PID):** A long lasting (perpetual) digital reference to an entity.
- **Resource:** a digital or physical collection item regardless of format or medium
- **Resolvability:** The ability to access information about the object through the identifier. A resolver can be thought of as an index, which keeps an updated location of the entity so that the entity can continually be accessed via the PID. In most cases a resolver will translate a PID into an actionable HTTP(S) URL.