Persistente Identifikatoren

Eindeutige Kennzeichnung einzelner Objekte durch PIDs

- Was ist ein PID?
- Muster für PIDs
- Registrierungsagenturen
- Vorteile eines PID

Was ist ein PID?

> Ein persistenter Identifikator (PID) ist eine eindeutige Kurzkennzeichnung für ein digitales Objekt. Er ist ein Aspekt für gute Datenqualität nach den FAIR-Prinzipien.

Ein PID ist eine Kennzeichnung für digitale Objekte, Dokumente oder Ressourcen im Internet. Dieser **Identifikator** beschreibt das Objekt **eindeutig, einzigartig** und **persistent**.

> **Identifikator** – Bezeichnung eines Objekts
> **eindeutig** – unmissverständliche Zuordnung
> **einzigartig** – einmalige Verwendung, identifiziert genau ein Objekt
> **persistent** – dauerhaft speicher- und abrufbar, unverändert über lange Zeiträume

Ein Identifikator bezeichnet ein digitales Objekt wie ein Name eine Person. Da es aber auch möglich ist, dass zwei Personen zufällig denselben Namen besitzen, hilft bei einem Objekt ein **persistenter Identifikator** (*persistent identifier, PID*) als eindeutige Referenz. Der persistente Identifikator eines Objekts lässt sich daher eher mit der Sozialversicherungsnummer vergleichen.

PIDs sind oft in einem Format, die eine Auflösung über das Internet ermöglichen, was bedeutet, dass man durch den PID auf das zugehörige Objekt zugreifen kann. Sie sind eine Kombination aus alphanumerischen Zeichen und verlinken meist auf eine Kontextseite mit Informationen über das Digitalisat und einem Link auf das Digitalisat selbst.

Verwaltet werden die PIDs von Registrierungsagenturen, die sicherstellen, dass die PIDs eindeutig und dauerhaft sind. Beispiele für bekannte PIDs sind DOI (Digital Object Identifier) für wissenschaftliche Publikationen, URN (Uniform Resource Identifier) für verschiedene Ressourcen oder HANDLE als allgemeines System für die Verwaltung von PIDs. Auch die ISBN (Internationale Standardbuchnummer) für Bücher kann man als eine Art PID verstehen.

Komponenten eines PIDs

Ein PID setzt sich aus zwei Bestandteilen zusammen:

- einem Code für die Identifikation des Objekts (Name), z. B. `Q131397`
- einem Code für die persistente Darstellung des Objekts und/oder relevanter Daten (Link), z. B. `https://www.wikidata.org/wiki/Q131397`

Durch die Verbindung von Identifikator des Objekts sowie einer Basis-URL kann das Objekt eindeutig bei Wikidata aufgerufen werden.

Wie URN und URL zusammen die ? URI ergeben, besteht auch der PID aus einem Namen und einem Link zum Objekt.

Ein digitales Objekt kann schließlich mehrere Identifikatoren und/oder PIDs besitzen. Wichtig ist, dass alle Objektkennungen in den Metadaten angeführt werden.

Ein digitales Objekt kann mehrere verschiedene (persistente) Identifikatoren besitzen.

Mehrere verschiedene Identifikatoren und PIDs können dasselbe Objekt beschreiben. Etwa hat die Venus von Willendorf auf Wikidata viele persistente Identifikatoren. Im Link der jeweiligen Identifikatoren kann die verwendete Struktur abgelesen werden.

- Wikidata: Q131397
- VIAF: 185215410
- Deutsche Nationalbibliothek / Gemeinsamen Normdatei GND: 4187499-7
- Französische Nationalbibliothek / ARK: 15950169p
- Library of Congress: n99030260

Zweck eines PIDs

Der Hauptzweck eines PIDs ist es, Informationen für die Identifizierung, Verifizierung und (digitale) Lokalisierung des beschriebenen Objekts bereitzustellen. Anstelle direkt auf das Objekt zu verlinken, kann der PID auch auf eine Kontextseite verweisen, auf der beschreibende Informationen und ein Link zum Medienobjekt zu finden sind. Dabei ist es auch wichtig, dass sich das referenzierte Objekt nicht verändert oder allenfalls eine Versionierung festgehalten wird.

PIDs gibt es für verschiedene Arten von Ressourcen; hauptsächlich werden zwei unterschiedliche Kategorien verwendet: PIDs für Objekte (Daten, Objekte, Publikationen …) und PIDs für Personen und Organisationen.

Unterschied zu lokalen Identifikatoren

Der entscheidende Unterschied zwischen PIDs und „normalen“ Identifikatoren ist, dass PIDs global verwendet werden und somit global eindeutig sein müssen. „Normale“ Identifikatoren sind oftmals nur lokal oder im eigenen System eindeutig. Die Verlinkung auf die Darstellung des digitalen Objekts kann dabei auch ein Dateipfad, der von außen nicht zugänglich ist, oder ein nur temporär verfügbarer Link sein. Wenn die Verlinkung sich ändert, kann mit einem regulären Identifikator das Objekt nicht mehr aufgefunden werden. Somit sind „normale“ Identifikatoren weniger zuverlässig für eine langfristige Auffindbarkeit und Identifizierung des Objekts.

Bekommt ein digitales Objekt einen (weltweit) eindeutigen Identifikator von der Institution und funktioniert die öffentliche Verlinkung auf dessen Darstellung auf unbestimmte Zeit, wirkt der normale Identifikator wie ein persistenter Identifikator.

Linktipps

- Kulturpool Forum: Metadaten (Anmeldung nötig)
- National Information Standards Organization: PID Forum
- Nationalarchiv Niederlande: Persistent Identifier Guide

er
S
ei

Muster für PIDs

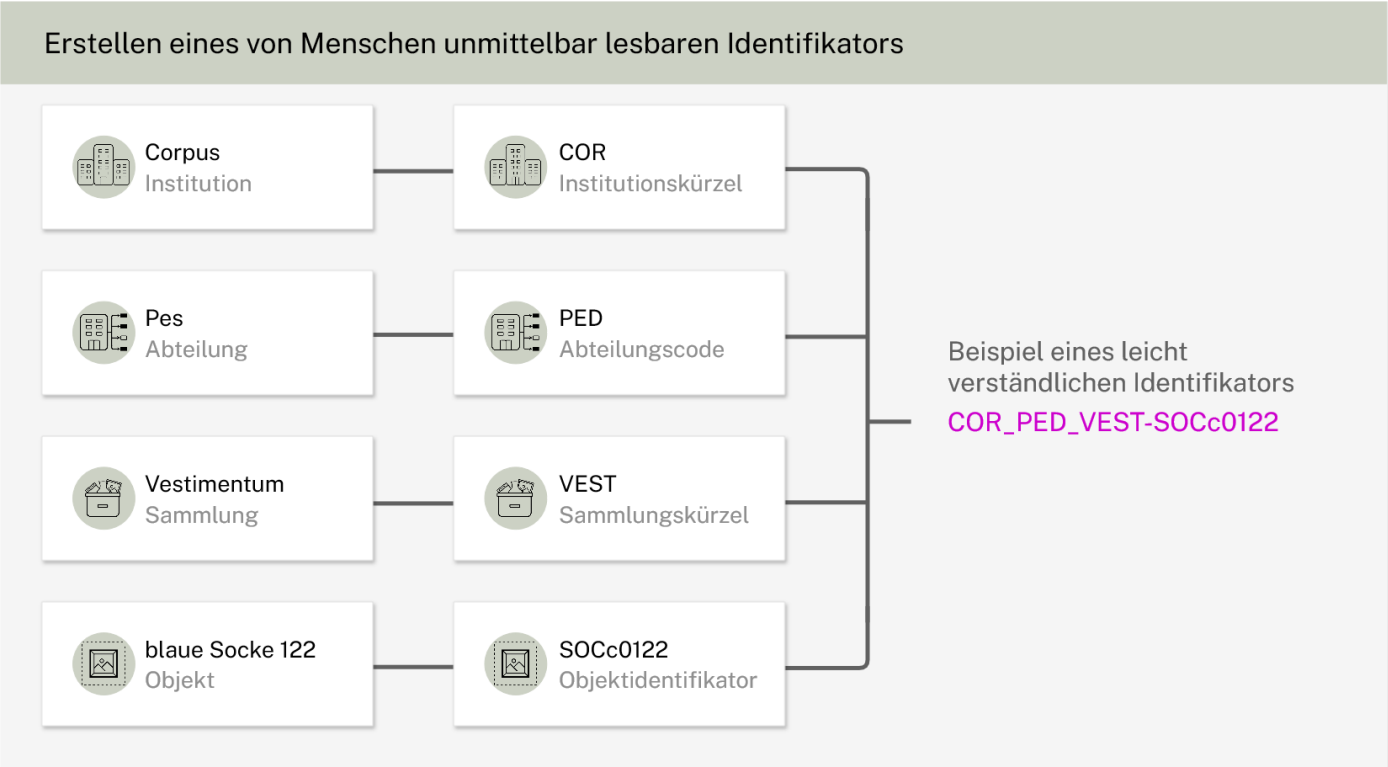
Für das Generieren eines PIDs braucht es eine eindeutige Folge von alphanumerischen Zeichen für jedes Objekt. Es empfiehlt sich dafür eine Struktur, die gleichzeitig für Menschen leicht lesbar und verständlich ist und das Objekt (zumindest) national eindeutig beschreiben kann. Eine eindeutige Identifikationsnummer lässt sich etwa nach dem hier vorgestellten Muster zusammenstellen:

- das Länderkürzel
- das (interne) Kürzel der Institution
- die Abteilung (oder ein Kürzel dafür)
- die Sammlung (oder ein Kürzel dafür)
- die intern vergebene Objektnummer

Am Beispiel einer Socke lässt sich ein solcher Identifikator erstellen:

- Länderkürzel = AT
- Institution = Körper (COR)
- Abteilung = Fuß (PED)
- Sammlung = Kleidungsstück (VEST)
- Objektnummer = SOCc0122

Daraus ergibt sich somit `AT_COR_PED_VEST-SOCc0122` als eindeutiger Identifikator.



Am Beispiel einer Socke lässt sich ein Identifikator nach einem immer gleichen Muster leicht erstellen. Grafik: Kulturpool, CC0.

Es kann in Institutionen vorkommen, dass verschiedene Abteilungen (oder Sammlungen) unterschiedliche Konventionen zur Erfassung eines Objekts haben. Aufgrund ihrer historischen Bedeutung für die Bestandserfassung können diese intern genutzten Identifikatoren in eine neue, eindeutige Anordnung einbezogen werden. Natürlich kann auch hier stattdessen einfach eine (neue) fortlaufende Nummer gewählt werden.

Tipps zum Erstellen eines PID

- Für die möglichst globale Eindeutigkeit kann das Kürzel des Landes vorangestellt werden.
- Ebenso kann ein eingetragenes Kürzel der Institution dazu beitragen, dass der Identifikator wirklich eindeutig ist.
- Eventuelle Bestandteile eines Objekts (z. B. gebrochene Teile) oder Angaben zu einer spezifischen Zusammengehörigkeit mehrerer Objekte können auch durch das Hinzufügen aufeinanderfolgender Zeichen am Ende der Objektnummer angegeben werden.
- Objekte können mehrere Identifikatoren (und auch mehrere PIDs) besitzen. Eine spätere Registrierung eines Objekts in einer PID-Datenbank ist somit kein Problem. Hat ein Objekt mehrere Identifikatoren, sollten idealerweise alle in den Metadaten zum Objekt angegeben werden.

Die Identifikatoren werden im Kulturpool im Metadatenfeld `«dc:identifier»` angegeben, wobei die Anzahl unbeschränkt ist. So können auf einfache Weise verschiedene Identifikatoren demselben Objekt zugeordnet werden.

PIDs in den Metadatenstandards

Mögliche Übermittlung des PID in einzelnen Metadatenstandards:

- ABCD: *UnitGUID*
- ABCDEFG: *UnitGUID*
- CRM: *Identifier E42*
- DarwinCore: *occurenceID*
- DublinCore: *identifier*
- EML: *DarwinCore occurenceID*
- ISAD(G): *Template field identifier*

di
e
s

Registrierungsagenturen

Es ist für die Nutzung im Internet oftmals von Vorteil, den persistenten Identifikator bei einer **Registrierungsagentur** (auch: Verzeichnisdienst, Repositorium, Resolver) registrieren zu lassen, um das Objekt wirklich global einzigartig und dauerhaft auflösbar zu machen.

Die Registrierungsagentur übernimmt die Speicherung der Daten zum Objekt und das Verwalten der Identifikatoren. Oft kann der Identifikator selbst ausgewählt werden und es werden von der Registrierungsagentur weitere Kürzel vorangestellt, um die Eindeutigkeit zu gewährleisten. Veränderungen und Aktualisierungen zum Objekt sind oft selbst verwaltbar, der PID bleibt dabei stets unverändert. Es ist empfehlenswert, direkt nach einer Registrierung des digitalen Objekts bei einem Verzeichnisdienst den PID auf der eigenen Website und in den Metadaten zu vermerken. Somit kann der PID so schnell wie möglich zum Zitieren oder Verlinken verwendet werden.

Persistente Identifikatoren, die bei einem System registriert werden, folgen üblicherweise einem Schema ähnlich zu diesem:

- die Webadresse der Registrierungsagentur
- ein Kürzel des Dienstes (z. B. ARK oder URN)
- ein Präfix für die Institution
- ein Suffix für das digitale Objekt

PIDs und Registrierungsagenturen

Verschiedene PIDs sind über weite Bereiche bekannt.

- Das Handle-System ermöglicht die dauerhafte Identifizierung von Internetressourcen und erlaubt Institutionen, eigene PID-Dienste zu betreiben. Dabei werden Kennungen lokal registriert und aufgelöst, was Kostenvorteile bietet, aber technische Wartung erfordert. Das Handle-System bietet keine direkte Registrierung von Metadaten.
- ARK (Archival Resource Key), betrieben von den California Digital Library Identifier Services, ist ebenfalls ein Handle-basiertes und dezentralisiertes System. Es unterstützt die Registrierung von Metadaten, allerdings ohne festen Standard.
- Wissenschaftliche Artikel werden oft mit einem DOI (Digital Object Identifier) der International DOI Foundation referenziert, der für eine Vielzahl elektronischer Objekte wie Fachartikel, Datensätze und Grafiken vergeben wird. Verschiedene Registrierungsagenturen bieten DOI-Registrierungs- und Auflösungsdienste an, die je nach geografischer Region oder Ressourcentyp variieren. Die Registrierung von Metadaten ist möglich.
- Die ORCID ID dient der eindeutigen Identifikation von Personen.
- Das Research Organization Registry (ROR) vergibt dauerhafte und eindeutige Identifikatoren für wissenschaftliche Organisationen.

Linktipps

- Remco van Veenendaal: Choosing a Persistent Identifier Type for Your Digital Objects, Digital Preservation Coalition (PDF) Dieser Bericht bietet einen Startpunkt für Institutionen mit Kulturerbe. Hier werden Fragen zu den persistenten Identifikatoren sowie deren Bedeutung und Verwendung betrachtet und beantwortet.
- Towards a National Collection: Which persistent identifier does what? PID functionalities Auf dieser Seite finden Sie eine Zusammenstellung zum Vergleich verschiedener Identifikatoren und deren Funktionalitäten.
- Titia van der Werf: Persistent identifiers in cultural heritage Generelle Fragen zu persistenten Informationen, z. B.: „Wie entscheidet man sich für einen PID?“, sowie eine Übersicht für PIDs für kulturelle Objekte.

e
st
ell

Vorteile eines PID

Im digitalen Raum steigt die Bedeutung einer genauen und zuverlässigen Identifizierung von Ressourcen. Genau hier bietet ein PID besonders Unterstützung. Etwa können Quellenangaben in Artikeln durch die Angabe des PIDs unterstützt werden, wodurch ein referenzierter Artikel schneller gefunden werden kann. Sie vereinfachen die Verwaltung, das Teilen und das Referenzieren der Ressource im digitalen Raum.

Warum sollte ein PID verwendet werden?

- **Eindeutige Identifizierung des Objekts**: PIDs sind wie ein Fingerabdruck für Ressourcen; sie helfen, diese klar zu identifizieren, auch wenn sich deren Beschreibung oder Links über die Zeit verändern. Zudem sind relevante Daten zu den Ressourcen durch den Zugriff auf Metadaten einfacher und länger auffindbar.
- **Zitieren und eindeutiges Zuordnen**: PIDs vereinfachen das richtige Zuordnen und Zitieren, wobei die richtige Institution als Halterin des Objekts einfach bestimmt werden kann. Dadurch kann die Institution akademische und ethische Standards aufrechterhalten und auch im weiteren Sinne die Zusammenarbeit und den Wissensaustausch fördern. Objekte können mühelos miteinander verbunden und verlinkt werden, wodurch eine ganzheitliche Erfassung der Metadaten vereinfacht wird.
- **Übersetzungen und Informationserhalt**: Durch die eindeutige Identifizierung der Objekte über den Identifikator können auch Übersetzungsfehler oder kleinere Änderungen der Wortbedeutung in anderen Sprachen verhindert sowie die Mehrdeutigkeit von Beschreibungen spezifiziert werden. Ebenso werden die Metadaten nicht in alle Sprachen übersetzt oder nicht mit Einträgen in kontrollierten Vokabularen versehen, sodass bei einer Suche nach diesem Objekt in anderen Sprachen verschiedene Informationen verloren gehen können. Ein PID stellt sicher, dass die wichtigen Daten und Metadaten dauerhaft zur Verfügung stehen.
- **Updates, Revisionen und Veränderungen der Website**: Registrierte PIDs erlauben Aktualisierungen, Überarbeitungen und Änderungen: Wenn etwa durch Veränderungen an Website-Strukturen alte Links nicht mehr aufgerufen werden können, kann ein Objekt nicht mehr angezeigt werden. Mithilfe eines PIDs kann die aktuelle Darstellung mit dem Objekt verbunden werden. Es können auch Versionierungen zu einem registrierten PID aufgenommen werden. Somit können historische Iterationen und neue Versionen gemeinsam gespeichert werden. Das trägt zur Nachhaltigkeit im digitalen Raum bei.
- **System für Verwaltung**: PIDs erlauben organisiertes und effizientes Verwalten und Aktualisieren der Daten, auch bei verlinkten Objekten. Die Identifikatoren vereinfachen das Verfolgen, die Veränderung und Archivierung der digitalen Objekte. Bei einer Verwendung eines PIDs, der von Menschen leicht lesbar ist, lässt sich leicht erkennen, an welchem Ort sich das Objekt befindet oder zu welcher Sammlung es gehört. Somit gehen keine wichtigen Informationen verloren, wenn es personelle oder institutionelle Veränderungen gibt.

Weitere indirekte Vorteile durch die Verwendung von PIDs:

- einfaches Darstellen und leichte Zugänglichkeit von weiteren Informationen, z. B. Nutzungsrechte für die Nachnutzung der Digitalisate
- gesteigerte Wahrnehmung durch die Zugänglichkeit der verknüpften Informationen über verschiedene Plattformen
- Verbindung zwischen dem physischen Objekt und dessen digitaler Information
- Investoren und Sponsoren können schneller den Umfang der Sammlungen erkennen

Im Sinne nachhaltiger Digitalisierung nach den FAIR-Prinzipien wird deshalb die Verwendung von PIDs empfohlen.