

Michael Büchner

PICHE – Persistente Identifikatoren für alle Kulturgüter

Wie können Bücher, Archivakten, Filme oder Kunstwerke – kurz: Kulturobjekte aller Art – dauerhaft und eindeutig identifiziert und adressiert werden, und zwar unabhängig davon, ob sie digital oder in physischer Form vorliegen? Vertreterinnen und Vertreter der Deutschen Nationalbibliothek (DNB), der Deutschen Digitalen Bibliothek (DDB), des Bundesarchivs und des Landesarchivs Baden-Württemberg haben sich mit dieser zentralen Frage intensiv auseinandergesetzt. Mit »PICHE – Persistent Identifiers for Cultural Heritage Entities« wurde zunächst ein Fachkonzept entwickelt, das nun in einem Pilotprojekt an der DNB umgesetzt und evaluiert werden soll.

Kulturgüter eindeutig und dauerhaft bezeichnen

Ein Persistenter Identifikator (PI) ist eine Bezeichnung, die der dauerhaften Identifizierung einer beliebigen physischen oder digitalen Ressource¹ dient. In der Regel besteht ein PI aus einer standardisierten Zeichenfolge. Die Zuordnung zwischen einem PI und einer Ressource erfolgt einmalig und sollte zeitlich unbegrenzt gültig sein.

Beispiele für PI-Dienste im Bibliotheksumfeld sind die »National Bibliography Number« (URN:NBN:DE), die »Internationale Standardnummer für fortlaufende Sammelwerke« (ISSN) oder die »Internationale Standardbuchnummer« (ISBN).

Ein PI kann außerdem die Grundlage dafür bilden, um auf die bezeichnete Ressource beziehungsweise auf Informationen darüber online zugreifen zu können. Dieser Zugriff erfolgt über einen Resolvingdienst. Ein Resolvingdienst speichert die Verbindung zwischen einem unveränderlichen PI und einer veränderlichen Uniform Resource Locator (URL), die wiederum zur Ressource beziehungsweise zu Informationen über die Ressource im Internet führt.

Grundlage für weiterführende und nachhaltige Nachnutzung von Kulturgütern

Bereits innerhalb von Kultur- und Wissenseinrichtungen sind stabile und eindeutige lokale Identifikatoren für die reibungslose Arbeit mit Beständen wichtig. Sie ermöglichen es, physische oder digitale Ressourcen eindeutig zu benennen, sie aufzufinden und mit ihnen zu arbeiten. Diese Identifikatoren können unterschiedlicher Art sein – lokale Identifikatoren können beispielsweise fortlaufende Nummern sein, aber auch Aktenzeichen oder sogar Zeichenfolgen, die auf elaborierten Signatursystemen beruhen – und garantieren die Stabilität und Verwendbarkeit des lokalen Bestands. Außerdem können mit Hilfe dieser Identifikatoren Zusammenführungen, Änderungen oder gar Löschungen nachverfolgt werden.

Global eindeutige Identifikatoren sind hingegen über die Grenzen einzelner Einrichtungen² hinweg gültig. Sie ermöglichen eine weltweite Interoperabilität, Vernetzung und verlässliche Zitierung von Kulturgütern.

Notwendigkeit von Persistenten Identifikatoren

Für die Wichtigkeit, global eindeutige und idealerweise uniforme PI für Bestände aus Archiven, Bibliotheken und Museen sowie aus Mediatheken, Forschungseinrichtungen und Denkmalpflegeeinrichtungen zu vergeben, fehlt jedoch oft das nötige Bewusstsein.

Die Arbeit der DDB hat gezeigt, dass Einrichtungen, die digitale Objekte liefern, nicht immer stabile und eindeutige Identifikatoren zur Verfügung stellen können. Dies hat für die DDB und ihre Funktionen, die unter anderem eine spartenübergreifende Vernetzung der Kulturobjekte ermögli-

chen sollen, weitreichende Folgen: Objekte in der DDB können in solchen Fällen nicht stabil zitiert werden, bei einer Aktualisierung der Metadaten können Duplikate entstehen, die durch Benutzer hinzugefügten Informationen zu Objekten können nicht dauerhaft mit ihnen verknüpft und verfügbar gemacht werden, und die Verlinkungen zurück zur Darstellung auf der Website der liefernden Einrichtung sind nicht stabil.

Nachnutzung bestehender PI-Dienste

Weltweit eindeutige und uniforme PI für Kulturgüter könnten über eine gemeinsame Verständigung aller Kultursparten ihren Weg in die Alltagspraxis finden. Bestehende PI-Dienste sind allerdings zu meist auf konkrete Anwendungsfälle spezialisiert – ein PI-Dienst für Ressourcen aus dem Bestand von Bibliotheken beispielsweise kann von Archiven oder Museen nicht unmittelbar nachgenutzt werden, da dort andere Anforderungen zum Tragen kommen.

Die Intention vom PICHE-Projekt ist es, einen solchen spartenübergreifenden Dienst für PI aufzubauen, mit dem Ressourcen aus dem Verantwortungsbereich beziehungsweise dem Bestand jeder Kultur- oder Wissens Einrichtung eindeutige und dauerhafte Bezeichner erhalten können. Und das außerdem unabhängig davon, ob diese bereits in digitaler Form vorliegen beziehungsweise ob sich ihre digitale Repräsentation in Zukunft verändern wird oder nicht. Bestehende PI-Dienste gehen zu meist von der Unveränderlichkeit der digitalen Repräsentation einer Ressource aus. Dies steht allerdings im Widerspruch zur ständigen technischen Weiterentwicklung. So werden inzwischen von immer mehr Museumsobjekten 3D-Digitalisate angefertigt, die weitaus aussagekräftiger sind als die bislang üblichen digitalen Fotografien. Der PICHE-Dienst zeichnet sich dadurch aus, dass er solche Veränderungen abbilden kann.

Das Konzept für den PI-Dienst wurde zunächst maßgeblich mit Archiven entwickelt. Die Vorarbei-

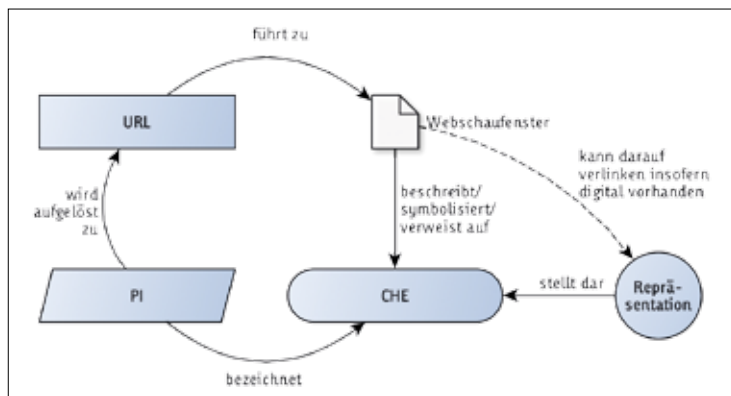
ten für das PICHE-Pilotprojekt gehen bis in das Jahr 2013 zurück, als gemeinsam mit Vertreterinnen und Vertretern der DDB und der Archiv-Community mit der Erarbeitung eines Fachkonzepts für einen spartenübergreifenden und materialunabhängigen PI-Dienst begonnen wurde. Bereits im November 2014 wurde das Fachkonzept in einer ersten Version veröffentlicht und wird seitdem stetig mit Vertreterinnen und Vertretern aller Kultursparten und unter der Schirmherrschaft der DDB weiterentwickelt.

CHE sind beliebige Kulturgüter aus Kultur- und Wissens Einrichtungen aller Sparten

Kulturgüter im Kontext von PICHE heißen Cultural Heritage Entities³ (CHE). Eine CHE ist eine intellektuelle Einheit, die zumeist als abgeschlossen betrachtet wird. Sie ist der zentrale Bezugspunkt in PICHE und kann eine konzeptionelle Mehrstufigkeit⁴ aufweisen. Eine CHE wird durch Metadaten (z.B. Titel, Autor, Erstellungsdatum) beschrieben, durch einen PI eindeutig bezeichnet und befindet sich in der Regel im Verantwortungsbereich einer bestimmten Kultur- oder Wissens Einrichtung.

Eine Repräsentation ist eine logische Einheit von beliebigen digitalen Dateien und/oder beliebigen physisch vorhandenen Gegenständen, die zu einer angemessenen Darstellung der CHE notwendig sind. Jede CHE hat mindestens eine Repräsentation, das physische oder digitale »Original«. Damit ein PI vergeben werden kann, muss für die CHE außerdem ein so genanntes Webschaufenster vorhanden sein.

Das Webschaufenster ist eine im Webbrowser darstellbare Webseite, die mit einer CHE verbunden ist und Informationen über diese zur Verfügung stellt. Dazu zählen zumindest beschreibende Metadaten und Links zu Repräsentationen. Die Bereitstellung, die Gestaltung und der Inhalt des veränderlichen Webschaufensters liegen im Verantwortungsbereich der Kultur- oder Wissens Einrichtung, die die beschriebene CHE in ihrem Bestand hat.



Schematische Darstellung für die Verwendung von PI für beliebige CHE

Nachnutzung bestehender Standards und Strukturen

Ein Bestandteil des Fachkonzepts ist die genaue Form des PI. Hierfür werden bereits etablierte Lösungen nachgenutzt: Universally Unique Identifier⁵ (UUID) definieren die Form von (statistisch gesehen) weltweit eindeutigen PI.

Ein wesentliches Kriterium beim Aufbau des neuen spartenübergreifenden PI-Dienstes ist außerdem die Nachnutzung bestehender Strukturen und Workflows an der DNB. Diese betreibt mit dem URN:NBN:DE-Dienst bereits erfolgreich einen PI-Dienst auf Basis von Uniform Resource Names⁶ (URN). PICHE wird zukünftig von der DNB betrieben werden und soll sich daher ergänzend in den bestehenden URN-Dienst einfügen.

Verantwortung auf mehrere Schultern verteilen

Ein spartenübergreifender PI-Dienst kann nur durch eine gut funktionierende, kooperative Organisationsstruktur zufriedenstellend zur Verfügung gestellt werden. Dafür muss Kultursparten- beziehungsweise Community-spezifisch Verantwortung übernommen werden, mit der vor allem fachspezifische Aufwände (z. B. Unterstützung von Anwendern, Mitarbeit in unterschiedlichen Gremien) einhergehen.

Das PICHE-Fachkonzept formuliert dafür drei Typen⁷ von Akteuren, die in einer verteilten Struktur angeordnet sind:

Die Betreiberin ist die DNB in Kooperation mit der DDB. Zusammen sind sie für die Koordination, die Realisierung und den Betrieb des PICHE-Dienstes verantwortlich. Dies gilt insbesondere auch für den verwendeten URN-Namensraum. Die DNB stellt außerdem die technische Infrastruktur (z. B. den Resolvingdienst oder entsprechende Verwaltungswerkzeuge) bereit.

Partner sind Kultur- oder Wissenseinrichtungen in Deutschland, die für eine Kultursparte und/oder für eine Community die koordinierende Verantwortung übernehmen. Sie treten als erster Ansprechpartner und Fachstelle für ihren Anwenderkreis auf. Partner spezifizieren Best Practises und sind in ihrem Verantwortungsbereich für den reibungslosen Betrieb des PI-Dienstes verantwortlich. Das Bundesarchiv tritt im Pilotprojekt als erster Partner auf. Anwender von PICHE sind Kultur- oder Wissenseinrichtungen, die CHE besitzen, dafür PI vergeben möchten und sich dazu an einen Partner wenden. Es ist möglich, dass Kultur- oder Wissenseinrichtungen zugleich als Partner und Anwender fungieren.

Für den produktiven Betrieb impliziert diese Aufteilung, dass sich Kultur- und Wissenseinrichtungen, die ihre Ressourcen mit PICHE persistent bezeichnen wollen, nicht direkt an die Betreiberin, die DNB wenden werden, sondern an einen für ihre Kultursparte beziehungsweise Community zuständigen Partner. Für Betreiber und Partner wird auch die DDB eine unterstützende Rolle einnehmen.

Das Pilotprojekt: Ein erster Schritt in die Praxis

Ein wesentliches Ziel des Pilotprojekts PICHE ist die konzeptionelle Klärung aller organisatorischen Fragen für den Betrieb des Dienstes und insbesondere die Ausarbeitung und Schärfung des Rollenmodells sowie der Rollen- und Aufgabenverteilung. Fachliche und politische Kommunikationsstrukturen müssen – wo etablierte Netzwerke nicht nachgenutzt werden können – geschaffen werden. Da der Bedarf an spartenübergreifenden Netzwerken auch in anderen Bereichen zunimmt, ist eine Zusammenarbeit mit den entsprechenden Gremien und Projekten wichtig⁸. Die abschließende Erprobung des

Dienstes soll zusammen mit dem Bundesarchiv (in einer Doppelrolle als Partner und Anwender), der DDB und der DNB erfolgen. Das Projekt soll bis Mitte 2018 abgeschlossen sein.

Das dem Pilotprojekt zugrundeliegende Fachkonzept wird (auch auf Basis des Abschlussberichtes)

von einer hoffentlich wachsenden und weiterhin so engagierten Community und auch Autorengemeinschaft stetig und ständig ergänzt und präzisiert werden, so dass der PI-Dienst zukünftig in einer gefestigten, respektierten und etablierten Struktur betrieben und genutzt werden kann.

Weiterführende Informationen

PICHE-Webseite: <<https://wiki.dnb.de/display/che>>

Projektwebseite: <<http://www.dnb.de/DE/Wir/Projekte/Laufend/piche.html>>

Mailingliste: che@listserv.dfn.de (<<https://www.listserv.dfn.de/sympa/subscribe/che>>)

Anmerkungen

- 1 Beispiele für physische Ressourcen sind museale Objekte oder Archivakten, für digitale Ressourcen PDF- oder Bilddokumente bis hin zu 3D-Modellen.
- 2 Die Grenzen für lokale PI sind nicht immer auf Einrichtungsebene zu finden. Auch Erfassungssysteme, die von unterschiedlichen Einrichtungen genutzt werden, können beispielsweise lokale PI vergeben, die dann innerhalb dieses einrichtungsübergreifenden Erfassungssystems eindeutig sind. Ein globaler eindeutiger PI dagegen wurde mit an Sicherheit grenzender Wahrscheinlichkeit weltweit nur ein einziges Mal vergeben und kann unter allen Umständen weiterverwendet werden (z. B. bei einer Migration in ein anderes Erfassungssystem). Somit spielt bei einem globalen PI nicht nur die Eindeutigkeit eine Rolle, sondern insbesondere auch die definierte und stabile Form.
- 3 Domainspezifisch wird eine CHE etwa als »Ressource« oder »dokumentarische Erschließungseinheit« bezeichnet. Bei der Erarbeitung des Fachkonzepts war es eine Herausforderung, Begrifflichkeiten zu finden, die über die verschiedenen Sparten hinweg durchgängig benutzt werden können und keine andere sparten- oder community-spezifische Bedeutung haben.
- 4 In ihrer Binnenstruktur kann eine CHE aus weiteren CHE zusammengesetzt sein. Dies ist beispielsweise bei hierarchischen Beständen im Archivbereich von besonderer Bedeutung.
- 5 Siehe RFC 4122 (A Universally Unique Identifiers [UUID] URN Namespace, 2005).
- 6 Siehe RFC 1737 (Functional Requirements for URN, 1994) und RFC 2141 (URN-Syntax, 1997).
- 7 Das Fachkonzept nennt noch eine weitere Rolle: »Ein Benutzer (auch Nutzer oder User) verwendet einen PI, um eine CHE zu referenzieren beziehungsweise zu zitieren (beispielsweise als Autor einer Publikation) oder in einem digitalen Netz zuverlässig zu dieser CHE beziehungsweise zu einem Webschaufenster der CHE zu gelangen (beispielsweise als Benutzer des Portals der Deutschen Digitalen Bibliothek).«
- 8 Insbesondere sind die aktuellen Aktivitäten zur Öffnung der Gemeinsamen Normdatei (GND) und der Erschließungsregeln (RDA) inklusive der zugehörigen Gremienstrukturen für andere Kultursparten zu nennen. Gerade das dazu geplante Modell der GND-Kooperative, bei der eine GND-Zentrale als Geschäftsstelle agiert sowie Agenturen einen bestimmten Teilnehmerkreis koordinieren, erscheint auch für PICHE – gegebenenfalls in leicht veränderter Form – als zweckmäßig.