nestor

Kriterienkatalog zur Prüfung der Vertrauenswürdigkeit von PI - Systemen

Entwurf zur öffentlichen Kommentierung

nestor-Arbeitsgruppe Standards für Metadaten, Transfer von Objekten in digitale Langzeitarchive und Objektzugriff

nestor-materialien 13

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Kriterienkatalog zur Prüfung der Vertrauenswürdigkeit von PI - Systemen

Entwurf zur öffentlichen Kommentierung

nestor Arbeitsgruppe Standards für Metadaten, Transfer von Objekten in digitale Langzeitarchive und Objektzugriff

nestor-materialien 13

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Herausgegeben von

nestor - Kompetenznetzwerk Langzeitarchivierung und
Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland

nestor - Network of Expertise in Long-Term Storage of Digital Resources

http://www.langzeitarchivierung.de

Projektpartner

Bayerische Staatsbibliothek, München
Bundesarchiv
Deutsche Nationalbibliothek (Projektleitung)
FernUniversität in Hagen
Humboldt-Universität zu Berlin - Computer- und Medienservice / Universitätsbibliothek
Institut für Museumsforschung, Berlin
Niedersächsische Staats- und Universitätsbibliothek, Göttingen

nestor Arbeitsgruppe Standards für Metadaten, Transfer von Objekten
in digitale Langzeitarchive und Objektzugriff

c/o Niedersächsische Staats und Universitätsbibliothek Göttingen
Jens Ludwig
Papendiek 14
D-37073 Göttingen
Tel.: +49 (0)551-3912121
E-mail:<ludwig@sub.uni-goettingen.de>

URN: urn:nbn:de:0008-20080710140
http://nbn-resolving.de/urn:nbn:de:0008-20080710140

Autor

Niklaus Bütikofer, InfoMemory GmbH

Weitere Mitwirkende

Karsten Huth, Bundesarchiv
Katja Hüther, Deutsche Nationalbibliothek
Dr. Christian Keitel, Landesarchiv Baden-Württemberg
Dr. Nikola Korb, Deutsche Nationalbibliothek
Jens Ludwig, Niedersächsische Staats- und Universitätsbibliothek Göttingen
Christa Schöning-Walter, Deutsche Nationalbibliothek
Sabine Schrimpf, Deutsche Nationalbibliothek
Tobias Steinke, Deutsche Nationalbibliothek

Gefördert im Rahmen der Initiative INS 2008

Kriterienkatalog zur Prüfung der Vertrauenswürdigkeit von PI-Systemen

Version 1.2

Inhaltsverzeichnis:

1 Einleitung

1.1 Zweck

Der vorliegende Kriterienkatalog zur Prüfung der Vertrauenswürdigkeit von Persistent Identifier-Systemen (PI-systeme) soll Anbietern und Nutzern von Persistent Identifiern (PI) helfen, über lange Zeiträume und nicht genau vorhersehbare Veränderungen hinweg digitale Objekte identifizierbar, referenzierbar und verfügbar zu halten, indem er aus der Sicht der Langzeitarchivierung relevante Anforderungen und Kriterien formuliert.

1.2 Kontext

PI-Systeme stehen in einem sehr dynamischen Kontext. Drei Entwicklungstendenzen sind hervorzuheben.

(1) Es gibt heute eine Reihe verschiedener PI-Systeme, die alle dieselben Grundfunktionen anbieten. Sie unterscheiden sich im Wesentlichen in der Strukturierung und Kodierung der Identifier, in der benutzten Technologie, im Geschäftsmodell und in den Zusatzdiensten, die sie in zunehmendem Umfange anbieten, um für Informationsanbieter und Nutzer attraktiver zu werden.

(2) Potentiell kann alles, was unterscheidbar und benennbar ist, mit einem PI versehen werden (von digitalen Dokumenten über einzelne Metadatenelemente zu Prozeduraufrufen). Das kann zu sehr hohen Mengen an Resolveranfragen führen. Die Skalierbarkeit eines PI-Systems und die Möglichkeit, bestimmte Objekttypen, welche grosse Resolverkapazitäten beanspruchen, auszuschliessen, ist deshalb wichtig.

(3) Auch PI-Systeme sind Änderungen unterworfen und umfassen ggf. eine große Anzahl von Akteuren. PI-Systeme, insbesondere die Betreiber von Resolver-Diensten, sind daher gezwungen, die Vertrauenswürdigkeit von PIs und Datenquellen zu beurteilen und den Nutzern transparent zu machen.

1.3 Anwendungsbereich

Das vorliegende Dokument bezieht sich nur auf die Grundfunktionen von PI-Systemen und nur auf die Situation, in welcher PIs ausserhalb einer bestimmten Institution veröffentlicht werden. Das impliziert, dass die in den PIs identifizierten Objekte auch für berechtigte Dritte ausserhalb einer bestimmten Institution zugänglich sind oder nach Ablauf einer bestimmten Frist zugänglich werden.

Bei der Ausarbeitung des vorliegenden Dokuments standen die Archive und Bibliotheken als Anwendungsfelder und die Langzeitarchivierung und Langzeitreferenzierung als hauptsächliche Anwendungszwecke im Vordergrund. Es ist deshalb wahrscheinlich, dass aus der Perspektive anderer Anwendungsfelder andere Gewichtungen vorgenommen werden und zusätzliche Faktoren ins Spiel kommen.

2 Begriffliche Grundlagen

Um begriffliche Klarheit und Genauigkeit zu erreichen, sind im Folgenden die zentralen Begriffe definiert und beschrieben. Damit werden gleichzeitig auch die konzeptionellen Grundlagen des Kriterienkatalogs dargelegt.[1]

Identifier: Ein Name (name), der eindeutig mit einem Objekt (thing) verknüpft ist. Man kann sagen, ein Name identifiziert ein Objekt, wenn der Name nur mit einem Objekt verknüpft ist. Ein Namen wird durch eine Zeichenkette repräsentiert, ein Objekt durch ein oder mehrere Exemplare.

Identität: Zwei oder mehr Objekte sind dann identisch, wenn sie in ihren wesentlichen Merkmalen übereinstimmen. Welche Merkmale wesentlich sind, hängt vom Zweck bzw.

[1] Die hier verwendeten Begriffe und Definitionen lehnen sich stark an die im australischen PILIN-Projekt ausgearbeitete Ontologie an: http://resolver.net.au/hdl/102.100.272/RPLZ54PQH

vom Kontext ab, in dem zwischen identischen und nicht identischen Objekten unterschieden wird.[2] Für die Zwecke der inhaltlichen Informationsarbeit bspw. sind die MS-Word-Version und die PDF-Version eines Dokumentes identisch. Für den Zweck der Darstellung auf einem Bildschirm sind sie dagegen nicht identisch, weil dafür in der Regel unterschiedliche Software benötigt wird und unterschiedliche Rechenvorgänge ablaufen.

Name: Ein Name wird durch eine Zeichenkette repräsentiert. Er gehört zu einem identifizierbaren Namenssystem (Kontext). Ein Name setzt sich im Kontext dieses Dokuments immer mindestens aus der Zeichenkette des Namenssystems und aus der Zeichenkette des Namens zusammen. Die Form der Zeichenkette muss den Regeln entsprechen, welche das Namenssystem setzt (Kodierungsschema, encoding scheme). Bestimmte Resolvingverfahren können Namen verlangen, die nach einem bestimmten Kodierungsschema aufgebaut sind.

Objekt (Thing): Ein Objekt kann alles sein, worüber gesprochen werden kann, insbesondere alles, was unterschieden und mit einem Identifier versehen werden kann. Es können also z.B. statische oder dynamische Objekte, Dokumente oder Prozeduren und auch aggregierte Objekte oder Teilobjekte identifiziert werden.

Der vorliegende Kriterienkatalog ist auf die Langzeitarchivierung ausgerichtet; er beschränkt sich auf Informationsobjekte, die im Prinzip abgeschlossenen und statisch sind. Langzeitarchivierung orientiert sich heute jedoch am OAIS-Referenzmodell, welches mit dem Konzept des Archival Information Package (AIP) als logische Einheit arbeitet. AIPs sind als gesamtes Paket über längere Zeiträume immer dynamisch, da sie bspw. ihre eigene ‚preservation history', die sich laufend ergänzt, mit sich führen und da sie, um erhalten zu werden periodisch in neue Formen überführt werden müssen. Jeder Betreiber einer Datenquelle muss Regeln bestimmen, nach denen er in diesem Spannungsfeld zwischen der Erhaltung des zu überliefernden Kerns und notwendiger Veränderung arbeitet.

Persistenz: Persistenz heisst in diesem Kriterienkatalog, dass ein Identifier auf Dauer eindeutig mit einem Objekt verknüpft bleibt. Ein persistenter Identifier wird dementsprechend nur einmal vergeben und bleibt ohne zeitliche Beschränkung im jeweiligen Resolversystem, welches das entsprechende Namenssystem und den entsprechenden Namensraum bedient, registriert, auch wenn das zugehörige Objekt nicht mehr existiert.

Persistent Identifier-System: Ein PI-System ist eine auf einander bezogene Kombination von

- Definitionen,
- Regeln (policies),
- Diensten (Services) und
- Datenquellen,

welche für die Verwaltung und Nutzung von persistenten Identifiern verwendet werden.

Resolver: Ein Resolver ist ein System, das Identifier registriert und auflöst, indem es auf Anfragen Informationen über die Verknüpfung des Identifiers mit den Angaben zum derzeitigen Lagerort des Objekts (Verknüpfungsdaten; derzeit URL) zurückgibt. Der Resolver wird von einem Betreiber als Dienst unterhalten. Das Resolving kann durch mehrere koordinierte Betreiber oder über koordinierte Sub-Resolver in mehreren Schritten durchgeführt werden.

Datenquelle (Data Source): Eine Datenquelle ist ein System zur Speicherung, Verwaltung und Vermittlung von Daten. Eine Datenquelle wird von einem Betreiber als Dienst unterhalten. Eine Datenquelle ist im Kontext dieses Dokuments typischerweise ein OAIS[3].

---

[2] Vgl. Norman Paskin, On Making and Identifying a "Copy", D-Lib Magazine 2003, Volume 9 Number 1, DOI: 10.1045/january2003-paskin

[3] Gemäss ISO 14721 (2003) Space data and information transfer systems — Open archival information system — Reference model

Verknüpfungsdaten (Association Data): Verknüpfungsdaten stellen die Verknüpfung zwischen dem Identifier und dem Objekt in einer Form dar, die im Resolver registriert werden kann und bei einer Anfrage definierte Aktionen auslöst. Die Verknüpfungsdaten enthalten Informationen, die benötigt werden, um auf die identifizierten Objekte zuzugreifen. Für die Vertrauenswürdigkeit des PI-Systems ist es unerheblich, ob die Verknüpfungsdaten nur aus einer URL bestehen, aus komplexen Anweisungen an das System der Datenquelle oder aus irgendeiner Verfahrensanweisung für künftige Datenübertragungssysteme, solange sie die Verknüpfung mit dem Objekt realisieren können. Das Resolversystem überlässt es der Datenquelle, wie sie das dem Identifier zugeordnete Objekt innerhalb ihres Systems auffindet und zur Verfügung stellt.

Ein einzelner Identifier kann mehrere Objekte identifizieren und deshalb mehreren Instanzen von Verknüpfungsdaten zugeordnet sein. Die im Resolver enthaltenen Verknüpfungsdaten können unterschiedliche Prozesse auslösen:

- Sie können bspw. über ein Http-Redirect eine Anfrage nach dem Objekt an die Datenquelle senden. Die Datenquelle ihrerseits kann das betreffende Objekt oder nur Metadaten über das Objekt zurückliefern. Die Datenquelle kann auch einen Dialog mit dem Nutzer initiieren, in welchem dessen Berechtigung abgeklärt wird, oder in welchem ein Bezahlungsvorgang abgewickelt wird, bevor das Objekt selber geliefert wird.
- Sie können, wenn mehrere Objekte unter einem Identifier registriert sind, zunächst einen Auswahldialog mit dem Nutzer auslösen und erst nach erfolgter Auswahl eine Anfrage an die jeweilige Datenquelle schicken.

Metadaten: Ein Resolver muss eine Reihe von Metadaten verwalten, um seine Dienste aufrecht erhalten zu können. Er muss insbesondere Daten unterhalten, die ihm erlauben, die Berechtigung der Betreiber von Datenquellen zur Registrierung von Identifiern und zum Nachführen von Verknüpfungsdaten zu prüfen. Bei mehrfachen Verknüpfungsdaten muss er über Metadaten verfügen, die eine Auswahl ermöglichen.

Auf der anderen Seite muss die Datenquelle diejenigen Metadaten unterhalten, die es ihr ermöglichen, die Verknüpfungsdaten im Resolversystem korrekt nachzuführen.

Regeln: Damit das PI-System wie erwartet funktioniert, müssen Regeln formuliert und eingehalten werden, sowohl auf der Seite des Resolverdienstes wie auch auf der Seite der Datenquelle. Die für einen zuverlässigen Betrieb notwendigen Regeln müssen unter den Beteiligten vereinbart werden und nach aussen hin transparent sein.

Vereinbarung: Da die Vertrauenswürdigkeit eines PI-Systems sowohl vom Resolver-Dienst wie auch vom Dienst der Datenquelle abhängt, ist eine für alle Nutzer transparente Vereinbarung notwendig, in welcher sich beide auf die Regeln verständigen, die einen vertrauenswürdigen Betrieb sicherstellen sollen.

Zugriffssystem: Die Datenquelle verfügt über ein Zugriffssystem, welches es ermöglicht, auf das identifizierte Objekt zuzugreifen.

Vertrauenswürdig (trustworthy): Eigenschaft eines Systems, gemäß seinen Zielen und Spezifikationen zu operieren (d.h. es tut genau das, was es zu tun vorgibt bzw. was seine Betreiber versprechen, dass es tut)[4]. Aus der Sicht eines Benutzers ist ein System vertrauenswürdig, wenn seine Erwartungen erfüllt werden.

Von einem PI-System wird insbesondere erwartet, dass seine Kernfunktionen (siehe unten) zuverlässig ausgeführt werden und dauerhaft zur Verfügung stehen und dass dieselben Identifier immer mit denselben Objekten verknüpft sind.

Die Dienste eines PI-Systems sind für einen Benutzer selber nicht genau nachprüfbar. Insbesondere kann der Benutzer nicht überprüfen, ob der Resolver ihn zum richtigen Objekt führt. Dies ist nur möglich, wenn er über weitere identifizierende Informationen über das Objekt (Metadaten) verfügt (z.Bsp. Autor, Titel, Datum, …). Auch dann kann er nicht sicher sein, ob das Objekt verändert worden ist, er kann lediglich prüfen, ob diejenigen Informationen, die er über das Objekt hat, immer noch zutreffen.

[4] erweitert nach INS Projekt 2007, S. 35

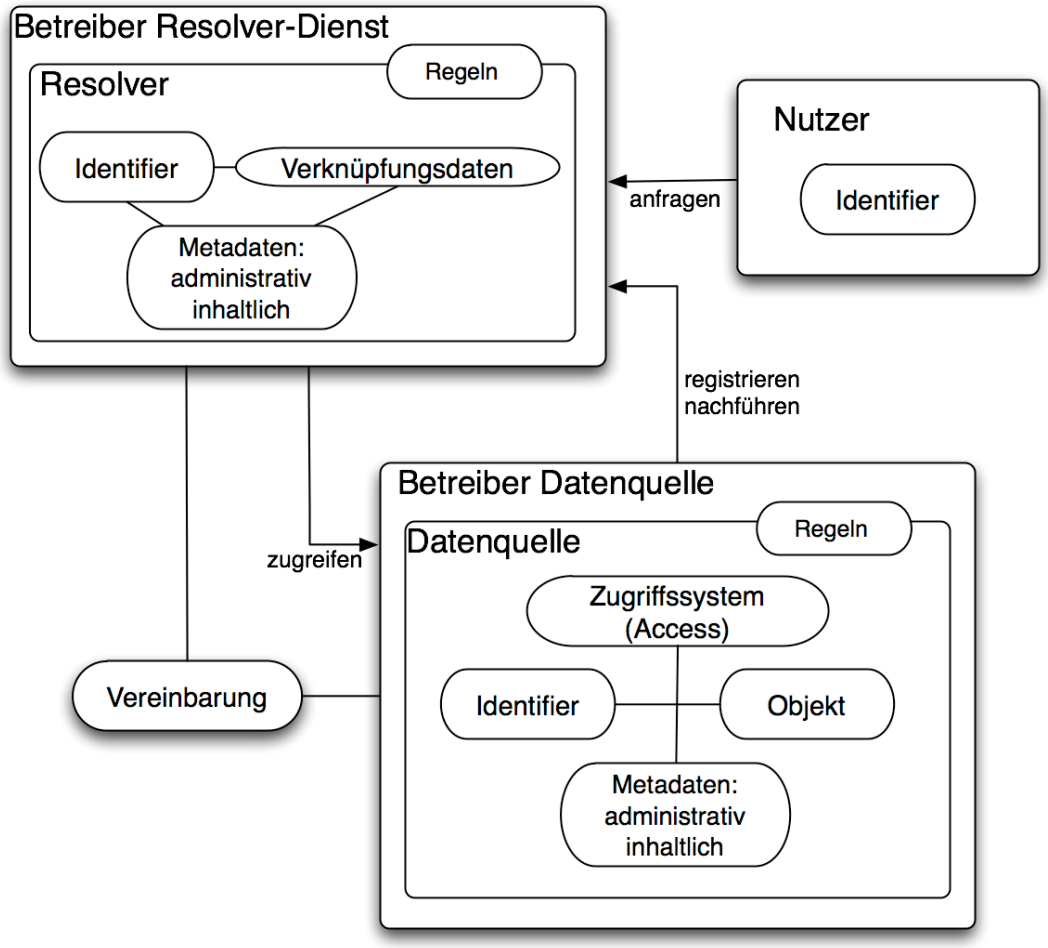

Abbildung 1: Übersicht PI-System

Kernfunktionen (Core Service): Die Kernfunktionen eines Persistent Identifier Management Systems schliessen ein:

- Namensvergabe regeln: Eindeutige Namen (Identifier) können vom Resolver-System zur Verfügung gestellt werden. Die Namensvergabe kann aber auch durch Zuteilung eines Namensraumes an Dritte, insbesondere an Datenquellen, delegiert werden.
- Registrieren (register): einen Identifier und die zugehörigen Verknüpfungsdaten im Resolver eintragen und dabei überprüfen, ob dieser nicht bereits registriert ist.
    a. Identifier wird vom Resolver-System vergeben:
    Die Datenquelle sendet ihre interne Objekt-Identifikation und die Verknüpfungsdaten an den Resolver. Dieser sendet den vergebenen Identifier zusammen mit der Objekt-Identifikation der Datenquelle an diese zurück.
    b. Identifier wird von der Datenquelle innerhalb eines ihr zugeteilten Namensraumes vergeben:
    Die Datenquelle sendet den Identifier und die Verknüpfungsdaten zusammen mit Authentifizierungsinformationen der Datenquelle, damit ihre Berechtigung, den Identifier zu vergeben, geprüft werden kann.
- Nachführen (update): die Verknüpfungsdaten zu einem Identifier auf einen neuen aktuellen Stand bringen.
- Auflösen (resolve): auf Anfrage hin die Informationen aus den Verknüpfungsdaten über den Zugriff auf das mit dem Identifier identifizierte Objekt zurückgeben.

Mehrwertdienste (value-added services):

Ein PI-System kann insbesondere zusätzlich …

- Informationen (Metadaten) über das Objekt, mit welchem die Identifier verknüpft sind, aufnehmen und zur Verfügung stellen (z.Bsp Metadaten zur Rechteverwaltung oder zu Form und Version des Objektes);
- Verknüpfungen mit anderen Identifiern, die bzw. deren zugeordnete Objekte in bestimmten Beziehungen zueinander stehen, aufnehmen und zur Verfügung stellen.

Zusatzdienste dürfen die Kernfunktionen nicht beeinträchtigen.

3 Herausforderungen an die Vertrauenswürdigkeit und Persistenz

Das vertrauenswürdige Funktionieren eines PI-Systems kann in der zeitlichen Entwicklung durch eine Reihe von Ereignissen und Veränderungen beeinträchtigt werden.

3.1 Beeinträchtigung der Vertrauenswürdigkeit im laufenden Betrieb

Bei der Ausführung der Kernfunktionen von PI-Systemen können neben den üblichen Bedrohungen von Informatiksystemen die folgenden Probleme auftreten:

Beim Registrieren von Identifiern im Resolver-System:

(1) Die Mehrfachregistrierung eines identischen Objektes unter verschiedenen Identifiern durch dieselbe oder durch unterschiedliche Datenquellen kann durch das Resolversystem nur erkannt und bei Bedarf verhindert werden, indem es zusätzliche identifizierende Metadaten über das Objekt speichert, verwaltet und bei jeder Neuregistrierung eine entsprechende Doublettenprüfung durchführt. Die Mehrfachregistrierung ist in Bezug auf die Vertrauenswürdigkeit des PI-Systems kein Problem, wenn man sich bewusst ist, dass man nicht von unterschiedlichen Namen auf unterschiedliche Objekte schliessen darf (siehe auch (13) unter 3.2).

(2) Die mehrfache Registrierung von nicht identischen Objekten unter demselben Namen kann durch das Resolversystem ebenfalls nur verhindert werden, wenn es zusätzliche identifizierende Metadaten über das Objekt speichert und verwaltet oder generell nur ein einziges Set von Verknüpfungsdaten für jeden Namen/Identifier zulässt.

(3) Die Verknüpfungsdaten sind nicht korrekt und erlauben keinen Zugang zum Zugriffssystem der Datenquelle. Das Resolversystem kann routinemässig überprüfen, ob die Verknüpfungsdaten bei ihrer Aktivierung Fehlermeldungen produzieren und die Datenquellen entsprechend informieren, sofern es eine korrekte Adresse der jeweiligen Datenquelle kennt.

Beim Nachführen (update) der Verknüpfungsdaten:

(4) Durch Veränderungen am System der Datenquelle werden die Verknüpfungsdaten im Resolversystem ungültig und der Betreiber der Datenquelle unterlässt es, die Verknüpfungsdaten im Resolversystem rechtzeitig nachzuführen. [Wie Problem (3)]

(5) Neue Verknüpfungsdaten zu einem Identifier werden registriert, aber die alten ungültigen Daten werden nicht gelöscht. [Wie Problem (3)]

(6) Die Datenquelle verliert im Rahmen von internen Veränderungen die interne Verknüpfung zwischen Objekt und dem Identifier bzw. den Verknüpfungsdaten im Resolver-System, so dass nicht mehr bekannt ist, bei welchem Identifier welche Verknüpfungsdaten nachgeführt werden sollten.

(7) Nicht autorisierte Personen melden irreführende Verknüpfungsdaten zu bestimmten Identifiern an das Resolver-System.

Beim Auflösen (resolving):

(8) Vorausgesetzt, die Daten im Resolversystem sind korrekt, wofür in erster Linie die Datenquellen verantwortlich sind, kommen als Fehlerquellen, welche die Vertrauenswürdigkeit beeinträchtigen können, praktisch nur technische Ursachen oder absichtliche Sabotage in Frage. Technisch ist das Resolving in seiner Kernfunktion ein relativ einfacher Vorgang, welcher im Rahmen der bei Informatiksystemen üblichen Sorgfaltspflichten und Schutzmassnahmen abgesichert werden kann.

3.2 Herausforderungen durch mittel- und langfristige Veränderungen

(9) Das PI-System als Ganzes oder einzelne Datenquellen ändert das Schema, nach dem die Identifier gebildet werden.

Die ist bspw. der Fall, wenn die Identifier bedeutungstragende Elemente enthalten, die verändert werden („sprechende PIs“). Die Folgen sind:

(1) Alle Zitierungen, die den alten Identifier verwenden, werden ungültig. Es sei denn der Resolver führt eine Konkordanz, über welche die alten Identifier automatisch mit den neuen verknüpft werden.

(2) Alle Datenquellen müssen die Identifier in ihrem System anpassen.

(3) Allfällige Zusatzfunktionen, die auf einem bestimmten bspw. hierarchisch aufgebauten Schema beruhen, werden unter Umständen obsolet, die Grundfunktionen werden dadurch aber nicht notwendigerweise in Frage gestellt.

Eine blosse Ergänzung der Identifier ist möglich, wenn …

(1) diese Ergänzungen in einer Form angebracht werden, die es erlaubt, die Ergänzungen von der bisherigen Zeichensequenz des Identifiers abzutrennen und wenn

(2) das Resolversystem die alten Identifier auflösen und den Anfragenden zu den neuen Identifiern weiterleiten kann.

Die Änderung von Identifiern lässt sich auf Dauer wohl kaum verhindern. Das Führen einer Konkordanzliste ist deshalb für PI-Systeme nötig um ein persistentes Resolving zu ermöglichen. Konkordanzlisten müssen nicht notwendigerweise im Resolversystem geführt werden, sie können auch von der einzelnen Datenquelle erstellt werden. Die Datenquelle kann auch den alten und den neuen Identifier jeweils mit identischen Verknüpfungsdaten, die auf dasselbe Objekt zeigen, unterhalten.

Globale Änderungen von Identifiern bergen immer das Risiko von Fehlern in sich, welche aber mit guter Vorbereitung und entsprechenden Tests praktisch ausgeschlossen werden können. Da in den Datenquellen diese Änderungen ebenfalls nachvollzogen werden müssen, stellt eine globale Änderung vor allem hohe Anforderungen an die Koordination aller am System Beteiligten.

(10) Das mit einem Identifier verknüpfte Objekt ist nirgendwo mehr verfügbar.

Solche Fälle sind unvermeidbar. Von vertrauenswürdigen PI-Systemen darf erwartet werden, dass sie den Identifier weiterhin registriert halten und eine qualifizierte Rückmeldung geben, die sich von einer technischen Fehlermeldung unterscheidet (z.Bsp. „Objekt nicht mehr verfügbar“).

(11) Das Objekt ist verändert worden. Die alte Version ist nicht mehr verfügbar, die neue Version gilt im Kontext der Datenquelle als nicht mehr identisch mit der alten.

(A) Der Namen bleibt gleich.

Weil Benutzer eines vertrauenswürdigen PI-Systems erwarten, dass gleiche Identifier immer mit identischen Objekten verknüpft sind, müssen sie auf die Veränderungen am Objekt aufmerksam gemacht werden. Verfügt die Datenquelle über ein OAIS-konformes System, dann sind solche Veränderungen

im AIP (Preservation history) verfügbar. Die Datenquelle muss diese dem Benutzer allerdings auch in angemessener Form anzeigen.

Taucht die alte Version wieder auf, kann sich ein Mehrfachvorkommen nicht identischer Objekte unter demselben Identifier ergeben, das den Nutzern in einem vertrauenswürdigen System transparent gemacht werden muss.

(B) Die neue Version hat einen eigenen Identifier.

Der Name der alten Version muss weiterhin auflösbar bleiben. Die Verknüpfungsdaten werden ersetzt durch einen Verweis auf den Namen der neuen Version. Der Benutzer muss darüber informiert werden, dass es sich beim referenzierten Objekt um eine neue Version handelt. Dies kann durch den Resolver-Dienst oder durch die Datenquelle geschehen. Die Betreiber des Resolver-Dienstes und der Datenquelle vereinbaren entsprechende Regeln.

(12) Das Objekt ist verändert worden. Die alte Version ist weiterhin verfügbar.

(A) Das Objekt ist unter einem zweiten Namen registriert worden.

Dieser Fall ist in Bezug auf die Vertrauenswürdigkeit kein Problem. Als Zusatzdienstleistung wäre erwünscht, dass Benutzer vom Resolver-Dienst oder von der Datenquelle auf die Existenz einer Vorgänger- oder Nachfolger-Version hingewiesen werden.

(B) Die neue Version wird in Form zusätzlicher Verknüpfungsdaten unter demselben Namen registriert.

Dieser Fall ist in Bezug auf die Vertrauenswürdigkeit ein Problem, wenn der Benutzer nicht ausdrücklich darauf hingewiesen wird. Dieser Fall muss durch klare Regeln über die PI-Vergabe bei neuen Versionen vermieden werden.

(13) Ein Objekt wird mehrmals mit verschiedenen Namen registriert.

Dieser Fall ist in Bezug auf die Vertrauenswürdigkeit des PI-Systems kein Problem, wenn man sich bewusst ist, dass man nicht von unterschiedlichen Namen auf unterschiedliche Objekte schliessen darf.

(14) Die technischen Verfahren des Resolverdienstes ändern sich grundlegend.

Solange diese Änderungen keinen Einfluss auf die Form der Identifier haben, solange der Resolverdienst die erwarteten Grundfunktionen ausführen kann und solange der Resolverdienst mit den Benutzern und den Datenquellen über die üblichen standardisierten Kanäle Daten austauschen kann, spielt es keine Rolle, auf welchen technischen Grundlagen der Dienst funktioniert. Das PI-System muss in seinen Grundfunktionen unabhängig von einem spezifischen technischen System sein.

Änderungen des Resolvingverfahrens können aber durchaus eine Änderung der Identifier bedingen. Vergleiche dazu (9).

(15) Das PI-System wird aufgegeben und der Resolver-Dienst eingestellt.

Da anzunehmen ist, dass nicht alle der heute existierenden Systeme langfristig überleben werden, ist dies ein realistisches Szenario. Ein vertrauenswürdiges PI-System muss deshalb seine Kerndaten in einem offenen Standardformat exportieren können und ein Szenario für die Systemnachfolge ausgearbeitet haben.

Grundsätzlich ist folgende Möglichkeit denkbar:

Alle Objekte, die im aufgelösten System registriert waren, werden in einem neuen System mit einem neuen PI registriert. Je nach Regeln der Namensgebung des neuen Systems kann die Zeichenkette im alten PI als Zeichenkette in den neuen PI übernommen werden (analog der Integration der ISBN-Nummern in bestimmten PI-Systemen). Andernfalls muss eine Konkordanz zwischen altem und neuem Identifiern mit entsprechendem Auflösungsdienst eingerichtet werden.

4 Kriterienkatalog

Vorbemerkung:

Die nachfolgenden Kriterien beziehen sich auf die globalen Anforderungen der Vertrauenswürdigkeit und der Persistenz, wobei Persistenz eine Voraussetzung für die Vertrauenswürdigkeit im Anwendungsbereich der Langzeitarchivierung ist.

Der Katalog ist in die drei Bereiche *organisatorischer Rahmen*, *Umgang mit Objekten* sowie *Infrastruktur und Sicherheit* aufgeteilt. Der Bereich *organisatorischer Rahmen* ist übergreifend und enthält auch Kriterien, die auf den Umgang mit Objekten und die Infrastruktur Auswirkungen haben.

„Öffentlich“ heisst in diesem Kriterienkatalog: Mindestens für alle an einem PI-System Beteiligten, also den Betreibern von Resolverdiensten und Datenquellen und ggf. eigenständigen Trägern frei zugänglich, aber nicht notwendigerweise für alle Nutzer.

4.1 Organisatorischer Rahmen

4.1.1 [Trägerschaft] Die Betreiber der Resolverdienste und der Datenquellen im PI-System sind wichtige Institutionen (Keyplayer) im Anwendungsbereich.

Je mehr wichtige Institutionen hinter einem PI-System stehen (und es auch selber benutzen), umso grösser die Verbreitung und umso kleiner das Risiko, dass es aus einem geringen Grunde aufgelöst wird. Die Namen der beteiligten Institutionen sind öffentlich. Die Betreiber der Resolverdienste und der Datenquellen können sich in einer rechtlichen Form als Trägerschaft zusammenschließen und die Verteilung der Verantwortung kann verschiedenen Modellen folgen. Sie kann kooperativ sein oder auch nur aus einer einzigen zentralen Institution bestehen.

zB: die International DOI Foundation (IDF) mit Verlagen und Forschungsinstitutionen als Mitgliedern (http://www.doi.org/idf-member-list.html)

4.1.2 [Verbindlichkeit] Die Betreiber der Resolverdienste verpflichten sich in rechtlich bindender Form und mit langen Kündigungs- bzw. Austrittsfristen, das PI-System auf Dauer zu unterhalten.

Auch wenn rechtliche Verpflichtungen immer gekündigt werden können, bilden sie Vertrauen und lange Kündigungs- bzw. Austrittsfristen verschaffen Zeit, um eine Migration in ein neues System in die Wege zu leiten. Die eingegangenen rechtlichen Verpflichtungen sind in ihren Kernelementen öffentlich.

zB. die Stiftungsurkunde der International DOI Foundation (IDF)

4.1.3 [Betreiber] Die Betreiber von Resolverdiensten sind in rechtlich bindender Form auf die Einhaltung der Definitionen, Grundsätze und Regeln verpflichtet.

Die Resolverdienste können als Organisationseinheit in eine Trägerinstitution eingebunden sein, sie können aber auch als selbständige Organisationen, die über Verträge und Kontrollmechanismen eingebunden sind, betrieben werden.

zB. der URN-NBN-Resolverdienst als Organisationseinheit der Deutschen Nationalbibliothek

4.1.4 [Betreiber] Der Resolverdienst ist eine Hauptaufgabe des Betreibers.

Der Betreiber kann neben dem Resolverdienst weitere Hauptaufgaben wahrnehmen, aber der Resolver muss gleichgewichtig mit anderen Aufgaben betrieben werden bzw. der Betrieb darf nicht unter anderen wichtigeren Hauptaufgaben leiden.

Ist der Betreiber als unselbständige Organisationseinheit in eine grössere Institution eingebunden, muss der Betrieb eines PI-Systems durch die zugeteilten Aufgaben abgedeckt sein.

zB: der URN-NBN-Resolverdienst der Deutschen Nationalbibliothek

4.1.5 [Geschäftsmodell] Die Finanzierung des Betriebs des Resolverdienstes ist nachhaltig gesichert.

Das Geschäftsmodell und die Finanzierungsquellen sind öffentlich und die Einkommensquellen fliessen stetig. Ein hoher Anteil der Einnahmen im Gesamtbudget ist vertraglich zugesichert oder über Kundenbindung abgesichert. Die finanziellen Ergebnisse sind öffentlich.

Die Transparenzanforderungen können vermutlich nur durch gemeinnützige Organisationen erfüllt werden. Eine finanzielle Reservebildung sollte aber möglich sein. Der Resolverdienst sollte für Nutzer kostenlos sein.

zB. Die Finanzierung erfolgt vollständig durch festgelegte Beiträge der Trägerinstitutionen oder der Betreiber der Datenquellen. Die Trägerschaft hat das Recht, diese Beiträge an einen veränderten Bedarf anzupassen.

4.1.6 [Rechte] Die Betreiber der Resolverdienste verfügen über sämtliche notwendigen Rechte am PI-System und am Resolversystem.

Die Betreiber der Resolverdienste legen die Herkunft und die Rechtslage der wesentlichen Systembestandteile offen. Sie verfügen über mindestens die unbeschränkten Nutzungsrechte an diesen Bestandteilen.

Eine Trägerorganisation hat die Zeichenkette seines Namenssystems und die Domain des Resolverdienstes rechtlich weltweit schützen lassen.

4.1.7 [Transparenz] Die wesentlichen Systembestandteile sind veröffentlicht.

Der Aufbau des PI-Systems, die verwendeten Definitionen, Datenmodelle, Regeln und Technologien sind öffentlich dargestellt.

zB. die DOI Webseite (www.doi.org) oder die URN-NBN-Webseite (www.persisten-identifier.de)

4.1.8 [Neutralität] Das PI-System favorisiert auf keiner Ebene die Verknüpfung mit Objekten bestimmter Datenquellen und macht die mehrfache Verfügbarkeit von Objekten dem Nutzer gegenüber transparent.

Zu einem Dokument existieren mehrere Exemplare, beispielsweise bei einem Archiv und bei einem kostenpflichtigen Portal. Der Resolverdienst legt gegenüber dem Nutzer offen, dass das gesuchte Objekt an mehr als einem Ort verfügbar ist und ermöglicht diesem, beide Verknüpfungen nacheinander zu aktivieren.

zB. Der Resolverdienst bietet dem Benutzer eine Auswahl an Verknüpfungsdaten zu vorhandenen Exemplaren an, aus denen er frei wählen kann.

4.1.9 [Exit-Strategie] Die Betreiber der Resolverdienste und der Datenquellen verfügen über eine Strategie, wie nach einer Einstellung des Resolverbetriebs die Auflösbarkeit der vergebenen PIs sichergestellt werden soll.

Die Betreiber der Resolverdienste und der Datenquellen verfügen über eine Strategie für den Fall, dass der Betrieb eines Resolverdienstes eingestellt werden muss, damit die Auflösung der vergebenen PIs sichergestellt werden kann.

zB. Die Betreiber kann darlegen, dass die PIs grundsätzlich durch andere Organisationen, insbesondere durch Resolverdienste anderer PI-Systeme aufgelöst werden können. Sie verfügt über Absichtserklärungen von Betreibern anderer PI-Systeme, im Bedarfsfalle die Resolverdienste zu übernehmen. Sie legt dar, wie die vergebenen PIs im Bedarfsfall über eine einfache Anwendung in neu vergebene PIs übersetzt werden können.

Sie verpflichtet sich, die URL des Resolverdienstes und allfällige andere, für die weitere Auflösung der PIs notwendigen Rechte entschädigungslos an Auffangorganisationen zu überlassen.

4.1.10 [Datenquellen] Die Betreiber von Datenquellen, die PIs für ihre Objekte vergeben wollen, verpflichten sich vertraglich gegenüber den Betreiber von Resolverdiensten, die Grundsätze und Regeln des PI-Systems einzuhalten.

Der Inhalt der Vereinbarungen und die Liste der Betreiber von Datenquellen, mit welchen eine Vereinbarung besteht, sind öffentlich.

Datenquellen, welche keine Gewähr bieten für die Einhaltung der Grundsätze und Regeln, kann eine Vereinbarung verweigert oder nachträglich gekündigt werden.

4.2 Umgang mit Objekten

4.2.1 [Skalierbarkeit] Das Kodierungsschema des Namensraumes erlaubt eine beliebig grosse Anzahl von Namensvarianten.

zB. Namen, die aus Laufnummern bestehen, welche nach oben offen sind.

4.2.2 [Eindeutigkeit] Der Identifier lässt eindeutig erkennen, zu welchem PI-System er gehört.

Da die Eindeutigkeit nur innerhalb eines Namenssystems gewährleistet werden kann, muss das jeweilige Namenssystem erkennbar und über den zugehörigen Resolverdienst auflösbar sein.

zB. Der Identifier umfasst als Bestandteil auch den Namen des PI-Systems (Etikett des Namenssystems)

4.2.3 [Eindeutigkeit] Die Betreiber von Resolverdiensten treffen angemessene Vorkehrungen, um die Mehrfachvergabe eines PIs für nicht identische Objekte zu verhindern.

Wird die Vergabe der Identifier an Betreiber von Sub-Resolvern oder Datenquellen delegiert, wird diesen ein klar abgrenzbarer und skalierbarer Namensraum zugewiesen.

zB. hierarchisch erweiterbare Unternamensräume wie beim URN:NBN-Schema

Einschränkung der Berechtigung zur Namensvergabe auf diejenigen Datenquellen, die in einem bestimmten abgrenzbaren Bereich die grösste Anzahl Objekte aufbewahren. Nur diese Datenquellen bieten die Objekte über den Resolverdienst an. Diese Datenquellen führen ein öffentliches Register, das genügend andere identifizierende Metadaten enthält. Dies ermöglicht es weiteren Datenquellen, zu überprüfen, ob einem Objekt aus ihrem Bestand bereits ein PI zugeordnet worden ist.

zB. Einschränkung der Namensvergabe für sogenannte Helvetica auf die Schweizerische Nationalbibliothek, welche gemäss ihrem Auftrag Vollständigkeit anstreben muss.

Der Resolverdienst verlangt von den Datenquellen zu jedem Objekt weitere identifizierende Metadaten, welche eine verlässliche Doublettenprüfung ermöglichen

Die PI-Vergabestelle trifft analog interne Vorkehrungen, um eine Mehrfachvergabe von PIs für nicht identische Objekte zu verhindern.

4.2.4 [Gültigkeit] Der Resolverdienst prüft periodisch die Gültigkeit der Verknüpfungsinformationen.

Der Resolverdienst aktiviert regelmässig die Verknüpfungsdaten und prüft, ob eine Fehlermeldung zurückkommt. Er gibt den Betreibern der Datenquellen eine entsprechende Rückmeldung.

4.2.5 [Sicherheit] Der Resolvingdienst erlaubt nur berechtigten Datenquellen, PIs zu registrieren und nachzuführen.

Vor jedem Registrierungs- und Nachführungsvorgang wird die Datenquelle authentifiziert und auf ihre Berechtigung überprüft.

zB. Verwendung von digitalen Signaturen beim Datenaustausch.

4.2.6 [Transparenz] Die Datenquelle legt ihre Regeln im Umgang mit ihren Objekten und der PI-Vergabe dar.

Die Regeln der Datenquelle legen fest:

- Welches Objekte mit PIs versehen werden und nach welchen Regeln deren Aufbewahrungsdauer festgelegt wird;
- welche Veränderungen des Objektes ein neue Version mit neuem PI zur Folge haben;
- welche alten Versionen weiterhin aufbewahrt werden;
- ob PIs von gelöschten Objekten weiterhin aufgelöst werden und eine qualifizierte Rückmeldung an Nutzer auslösen (zB. „Objekt am 5.10.2008 gelöscht, überarbeitete Version unter PI XY verfügbar“)

Die Regeln sind für alle Nutzer einsehbar, damit sie das Ergebnis des Resolving-Vorgangs beurteilen können.

4.2.7 [Transparenz] Nutzer erhalten eine qualifizierte Rückmeldung über die Verfügbarkeit und Veränderungsgeschichte.

Ist ein Objekt nicht direkt zugänglich erhält der Nutzer entweder von der Datenquelle oder vom Resolverdienst, die ihm Auskunft über die Vorhandensein und Zugänglichkeit des Objekts geben.

zB. „Kein Objekt zum PI XY verfügbar“; „Objekte nur für berechtigte Nutzer XY zugänglich“.

Ist ein Objekt seit der Vergabe eines PIs in wesentlichen Merkmalen verändert worden liefert die Datenquelle einem Nutzer Angaben zur Veränderungsgeschichte des Objekts (v.a. für die Identität des Objekts relevante Informationen)

4.2.8 [Einfachheit] Die Namen von PIs sind einfach aufgebaut.

Es kommt vor, dass PIs durch Menschen abgeschrieben werden müssen. Lange und komplizierte PIs verursachen dabei häufig Fehler. Zu diesem Zwecke sollen PIs kurz und strukturiert sein sowie nur aus Zeichen des ASCII-Zeichensatzes bestehen.

zB: urn:nbn:ch:bel-9478

4.2.9 [Einfachheit] Die Auflösung eines PI muss für den Benutzer einfach und barrierefrei durchgeführt werden können.

Der Resolverdienst kann auf einfache Weise benutzt werden. Die Benutzungsoberfläche folgt anerkannten Standards der Usability und der Barrierefreiheit.

4.3 Infrastruktur und Sicherheit

4.3.1 [Standortunabhängigkeit] Der Zugriff auf den Resolverdienst ist über verbreitete öffentliche Netzwerke möglich.

zB. Zugang zum Resolverdienst über Internet und http-Protokoll-Dienste.

4.3.2 [Sicherheit] Der Betreiber des Resolverdienstes trifft angemessene und anerkannte Massnahmen der Informatiksicherheit.

PI-Systeme benutzen IT-Systeme, deren Vertrauenswürdigkeit in starkem Masse vom ordnungsgemässen Betrieb und den Sicherheitsvorkehrungen, die getroffen werden, abhängt. Für diesen Bereich gibt es anerkannte Standards wie ISO 17799, welche die Anforderungen im Detail darlegen, sowie eine Reihe von Audit-Möglichkeiten, um die Einhaltung der Standards überprüfen zu lassen. Der Betreiber des Resolvingdienstes wendet diese Standards an und lässt deren Einhaltung regelmässig durch externe Stellen überprüfen (Audit). Er veröffentlicht die Ergebnisse.

4.3.3 [Ausfalllösung] Der Betreiber unterhält eine Ausfallösung für seinen Resolvingdienst.

Der Betreiber des Resolverdienstes zeigt auf, wie er sicherstellt, dass sein Dienst dauernd (99%) zur Verfügung steht.

zB. Es stehen genügend Spiegelserver zur Verfügung, welche bei einem Ausfall eines Resolvingservers alle laufenden Anfragen bewältigen können.

4.3.4 [Technologieunabhängigkeit] Der Betreiber ist bei der Nutzung und beim Ersatz von Systemteilen unabhängig von einzelnen Dritten.

Es gibt keine für den Betrieb des Resolverdienstes notwendigen Systembestandteile, deren Nutzungsrechte von Dritten widerrufen werden können, deren Ersatz bei Defekten nicht gesichert ist und die nicht problemlos durch andere Produkte ersetzt werden können.

Der Betreiber legt die Herkunft seiner Systembestandteile offen.

zB. Verwendung von Opensource-Produkten und von weitverbreiteten Standard-Komponenten.

4.3.5 [Migrierbarkeit] Die für die Kernfunktionen notwendigen Daten können jederzeit in offen dokumentierten und standardisierten Formaten aus dem System exportiert werden.

Um die Übertragung der Daten in andere Systeme zu ermöglichen und zu erleichtern sind Exportfunktionen, welche offen dokumentierte Standardformate benutzen, notwendig.

zB. Export der Resolvereinträge in einem vom Betreiber dokumentierten XML-Format.

4.3.6 [Skalierbarkeit] Der Resolverdienst ist beliebig erweiterbar.

Der Resolverdienst ist beliebig erweiterbar, um mit wachsenden Mengen an registrierten PIs und mit einer wachsenden Anzahl von Anfragen umgehen zu können und akzeptable Antwortzeiten zu garantieren.