

Persistent Identifiers:

Weighing the Benefits of In-house Systems Versus External Registries

Joanna McCloud

jmcclou1@jhu.edu
785-248-1316

Johns Hopkins University
Museum Studies Digital Curation Program

Submitted December 15, 2015
Revised final version May 22, 2018

Abstract

Cultural heritage institutions are struggling with how they will manage their digital assets for long-term visibility, accessibility, and preservation. To date, being able to share digital content and associated metadata for re-use has been at times a haphazard process, resulting in the loss or inaccessibility of many valuable objects. Persistent identifiers (PIDs) assigned to digital assets are unique, actionable, and effective in correcting this problem. The challenge has been to develop a system or systems that are inexpensive and easy to use by content owners. Some institutions use external systems, while many institutions have chosen to develop their own. Through analysis of the literature about PIDs, expert interviews, and link reliability testing, the advantages and disadvantages of both locally developed systems of creating, assigning, and maintaining PIDs versus the services of external registries such as Persistent Uniform Resource Identifiers (PURLs), Digital Object Identifiers (DOIs), and Archival Resource Keys (ARKs) and the EZID service are examined.

Table of Contents

- I. Introduction
- II. Research design and methodology
 - a. Literature analysis
 - b. Interviews with experts
 - b. Persistent link testing
- III. Analysis of the literature on PIDs
 - a. How to keep digital assets on the vine
 - b. What should be assigned a persistent identifier?
 - c. Persistent Uniform Resource Locator
 - d. Digital Object Identifier
 - e. Archival Resource Key
 - f. Reluctance to using external registries
- IV. Interviews with experts
 - a. In-house systems
 - b. External registries
 - c. Digital Object Identifiers
 - d. Archival Resource Keys and EZID
- V. Persistent link testing
- VI. Conclusion
- VII. References
- VIII. Appendices

Persistent Identifiers:

Weighing the Benefits of In-house Systems Versus External Registries

Introduction

An important goal of museums, archives, and libraries today is to digitize their collections--not only for preservation, but to make them available to other cultural heritage institutions, scholars, and the public for research and re-use. Aside from legacy data, the creation of new data is increasing at an alarming rate, making the management of digital assets a tremendous problem.

In addition to this, the National Science Foundation (NSF) recommends identifying research data that is generated through grants to promote sharing, give credit for the intellectual effort of the investigators, and maintain persistent access. This complies with an Executive Order given in 2013, expecting Federally funded research data to be made available to the public within a year of publication (<https://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>). Identification of data is typically done through persistent identifiers (PIDs).

In 2016 the National Information Standards Organization (NISO) explored best practices related the use of PIDs in research. This was a component of a larger study of Alternative Assessment Metrics (Altmetrics), working toward standardization of guidelines in this area (NISO RP-25-201x2B, p. 1). Altmetrics are the measurement of qualitative data that provides information on how objects are discovered, used, and the impact they make. This data is useful for grant applications, shows stakeholders the return on their investment, inspires new research topics, and identifies potential collaborators (NISO RP-25-2016 Outputs of the Alternative Assessment Metrics Project at

https://groups.niso.org/apps/group_public/download.php/17091/NISO%20RP-25-2016%20Outputs%20of%20the%20NISO%20Alternative%20Assessment%20Project.pdf). The public version of the discussion of PIDs in research and scholarly works, “Persistent Identifier Landscape Analysis-July 2016,” can be found at <https://goo.gl/1Ka76p>. The final guidelines have yet to be published.

Persistent identifiers are different than file names and are usually comprised of a series of numbers and/or letters that are permanently assigned to digital objects. They are part of the metadata and stay with them throughout the lifecycle, regardless of whether files are renamed or moved to a different location, and even remain as a “tombstone” (or marker) if they are deleted. Dr. Joe Tennis, Assistant Professor at The Information School of the University of Washington believes that PIDs are "an absolute must" to metadata sharing and are now considered best practice for preservation and management (personal communication, July 20, 2015).

One of the biggest difficulties in implementing actionable PIDs is the varied systems that are employed in minting them. Juha Hakala, Senior Advisor at The National Library of Finland points out that "In spite of the claims for the contrary, there is not a single standardized PID system out there" (2013, p. 4). PIDs are created from in-house (or local) systems or by recording objects with an external registry. There are benefits and disadvantages to both. Using an in-house system can be less expensive and organizations have complete control over what is done, but unless the PID is maintained and is actionable, the objects are virtually undiscoverable. Also, if the assets are transferred to another department within the organization or to another institution that uses a different system altogether, integration may be difficult, if not impossible.

There are external registries that mint PIDs such as Persistent Uniform Resource Identifiers (PURLs) managed by the Online Computer Library Center, Inc. (OCLC), Digital

Object Identifiers (DOIs) overseen by the International DOI Foundation (IDF), and the California Digital Library's Archival Resource Keys (ARKs) that are assigned through EZID. The registries provide technical support, redirect to objects globally, and in some cases back up the data as well. The downside is broken links, differences in the systems hampering migration, and the cost to obtain and maintain identifiers.

This paper will address the following questions: What are the current best practices regarding PIDs? What are the advantages and disadvantages of local systems and external registries such as those that provide PURLs, DOIs, and ARKs? What is the comparison between external systems in functionality and reliability? What are the barriers that contribute to the lack of their adoption?

Research Design and Methodology

Three different methods of research were used in this project:

- Analysis of pertinent literature on the topic of persistent identifiers.
- Interviews with experts in the field to highlight the current best practices.
- Testing 20 random links to PIDs over approximately 2 and a half years to determine the reliability of PURLs, DOIs, and ARKs.

Literature Analysis

Keeping digital assets alive on the vine.

Susan Lyons, a Documents/Reference Librarian at Rutgers University Law School, discusses the explosive growth of using PIDs for scholarly citations in her article "Persistent Identification of Electronic Documents and the Future of Footnotes." She says "In 1994, there were just four instances of web citations in three law review articles. By 2003 there were at least 96,946 citations to the web in law review footnotes" (2005, p.68). The need for better data

management is a crisis now for many organizations because of the current and expected increase in volume of digital material. An expert in big data, Rachel Wolfson explains that "...the amount of data continues to double in size every two years" (2015).

Digital assets regularly disappear. One of the problems attributing to this phenomenon is link rot—where URLs are used as an identifier and the link dies. When this happens, the object associated with it essentially dies too. Sumeer Gul, Iram Mahajan, and Asifa Ali from the Department of Library and Information Science, at the University of Kashmir discuss link rot in their article "The Growth and Decay of URLs Citation: A Case of an Online Library and Information Science Journal." They point out that "...studies have shown that up to 23% of Uniform Resource Locator (URL) references became inactive after one year and up to 52% after five years" (2014, p. 29). When objects are deleted or moved to new locations, there is no way to track what happened to them or where they are currently residing, unless the PIDs are used and maintained. There could also be loss through reorganization where the filename is altered, or the mission of the institution changes and the maintenance of some objects are no longer a top priority (Lunghi, Cirinnà & Bellini, 2013).

Persistent Identifiers maintain their integrity through technological and cultural changes, as well as ownership. Mark Chillingworth is one of many who feel that PIDs could be a strong solution to hidden metadata. In his article "Identity Check" he says "The information industry believes persistent identifiers will deliver improved information linking, naming, storing, search and access" (2004, p. 1), and that the result will improve user experiences. Persistent Identifiers can help digital assets remain alive and accessible, but not only do the identifiers need to be persistent, they must also be unique and actionable.

What should be assigned a PID?

It is an important decision choosing which objects should have PIDs assigned to them. In many cases the sheer volume of digital assets that many institutions possess prohibits giving all objects in their collections PIDs. Therefore, it is necessary to select representative groups. "A PID should be assigned only to resources that are stable, significant for the related user community, and suitable with the scope of the identification system" (Lunghi et. al., 2013, p. 377).

Persistent Uniform Resource Locator (PURL)

The Online Computer Library Center (OCLC) (<https://purl.oclc.org/docs/index.html>) developed Persistent Uniform Resource Locators, or PURLs, in 1995 as hyperlinks for online catalogs (Lyons, 2005). They are designed to identify printed resources, cataloging systems, and Web pages. A PURL is a URL that uses the HyperText Transfer Protocol (HTTP) for redirection (Emma Tonkin, 2008). When searching for objects using PURLs, the process is complete when the Internet browser accesses the appropriate server with the content. Hans-Werner Hilse and Jochen Kothe describe the permanence of PURLs in their article "Implementing Persistent Identifiers: Overview of Concepts, Guidelines and Recommendations" by saying that while "One can change what a PURL resolves to, one cannot change or delete the PURL itself" (2006, p. 32).

There are three components to a PURL string such as

<http://purl.fdlp.gov/GPO/gpo54200>:

- the protocol (example: http://)
- a resolver address (example: pur.fdlp.gov/GPO)
- the name for the resource (example: gpo54200)

While the path of a URL is case sensitive, this is not true for PURLs. Also, PURLs are unique only in the part of the string that contains the resolver address (Hilse & Kothe, 2006).

The resolver is the portion of the identifier that completes the search action.

The PURL system is a central registry that requires participating institutions to install a local server and monitor any changes in the assets to which their PURLs point. Because PURLs include a URL, they break just like URLs do and must consistently be checked. It is a high-maintenance solution, but unlike DOIs they do not require a resolution service.

Clifford Lynch, Director of the Coalition for Networked Information, says that PURLs are designed to work by looking in a database for the currently listed location of an object and connecting that way. He also notes an inherent problem with the system saying, "PURLs will probably no longer work as new protocols appear that supersede HTTP, and as content migrates to access through such successor protocols" (1997-1998, p. 5). PURLs may not be versatile enough to overcome technological changes in the future.

In 2005 when Susan Lyons wrote about PIDs, the United States Government Printing Office (GPO) was the largest user of PURLs. At that time the GPO had been assigning PURLs for seven years and were adding 14,000 to 15,000 of them each year. To keep up with the changes, the GPO has an electronic archive in place that redirects objects removed from the main website, so that assets can always be found (2005). Other large institutions such as the National Library of Australia, the Florida Center for Library Automation, and the National Center for Biomedical Ontology use PURLs as well.

Digital Object Identifier (DOI)

Liu Wang, Assistant Professor and Electronic Information Coordinator at Western Kentucky University, Bowling Green states that Digital Object Identifiers (DOIs) (www.doi.org)

were featured at the 1998 Frankfurt Book Fair as a method of integrating intellectual content and the web (2007). It was standardized as ISO 26324 in 2012 (Neumann & Brase, 2014).

Used in conjunction with the Handle System, DOIs are resolved to their destination. The Handle System (<http://www.handle.net>) was developed by the Corporation for National Research Initiatives (CNRI) as a naming system and became available in 1994. It is considered the "underlying *communication* technology for managing and resolving DOIs" (Hilse & Kothe, 2006, p. 21). The objectives of handles are to be unique, persistent, proficient at recognizing multiple instances of an object, and provide international support of all known printable characters. The software for the system is free but institutions must register with CNRI establishing a Naming Authority [entity that is responsible for assigning PIDs], and sign a license agreement (Hilse & Kothe, 2006).

DOIs are overseen by the International DOI Foundation (IDF), a not-for-profit member organization that includes the DataCite and CrossRef agencies. The foundation is run by officials that are elected from participating institutions. Affiliates can also become Registration Agencies (RAs), assign DOIs, and manage the content that the handles use (Hilse & Kothe, 2006). The cost of registering publishing names depends on the services purchased and which RA is chosen, as each one has the authority to use its own business model and may work on a cost recovery basis. At the beginning of 2018 more than 148 million DOIs had been assigned (www.doi.org/faq.html).

As previous director of IDF, the late Dr. Norman Paskin outlined the DOI system by saying that it brings together the following attributes to strings:

- syntax specifications-the DOI string
- a resolution module-the handle

- a metadata module-associates descriptive metadata with the object and
- social infrastructure-support through policies and RAs (2010, p. 2).

The DOI string always begins with the number 10 (to designate that it is a DOI) and then is divided into two more sections. Example:

- doi: 10.12026/april2015

After the number 10 is the Publisher ID and the Item ID, which are separated by "/" (Wang, 2007).

Other large DOI publishers include the American Psychological Association (APA), the Elton B. Stephens Company (EBSCO), Massachusetts Institute of Technology (MIT) Press, and the American Medical Association (AMA) (Lyon, 2005).

In "An Update on Uniform Resource Locator (URL) Decay in MEDLINE Abstracts and Measures for its Mitigation," Erick Ducut, Fang Liu, and Paul Fontelo say that DOIs can identify almost any object. Images, audio and video assets, software, ebooks, journals, printed music, and databases are all within the realm of DOIs (2008, p.2). They can also be applied to any desired level of granularity," as well as...a whole book and also to every chapter, every illustration, photograph, or table" (Wang, 2007, p.162).

Not only are DOIs versatile and automatic, they also help with content management. It is possible for DOIs to remain effective, even when objects are moved or ownership changes. If either one of these events occur, the Publisher ID is all that will need to be altered to keep assets visible; the Item ID will remain the same. Liu Wang points out other benefits of using DOIs. She says that they can increase the "...usage of acquired resources, [provide] expanded access to content not owned, and enhanced localized thinking" (2007, p.163). Museums, archives, and

libraries not only will meet Federal requirements for funding but will also be able to share objects with each other and the world at large.

DOIs identify the object itself, not its location, and each version of an object can be assigned its own DOI. Paul Miller, an analyst and consultant, says they "can be resolved to multiple physical manifestations of that intellectual property, overcoming the difficulty of identifying and describing the existence of a resource at more than one location" (2000, p. 3). Although versatile and effective, unless changes are updated in the system, the handle will fail for assets that are renamed or moved to new locations. In the article "A Policy Checklist for Enabling Persistence of Identifiers," Nick Nicholas, Nigel Ward, and Kerry Blinco say that there is a "longstanding problem of URIs breaking (no longer retrievable), when the resource pointed to it is moved or no longer maintained" (2009, What do you mean by persistent? para 2).

DataCite is another DOI agency (<https://www.datacite.org/>). Janna Neumann and Jan Brase discuss the qualities of DataCite in their article "DataCite and DOI Names for Research Data." Since 2009 DataCite has been using DOIs to provide easy access to research data and as an effective and easy way to handle citations. A search tool of indexed metadata is included for DOI names registered through DataCite, which is made available through the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). They provide an application programming interface (APID) for harvesting (2014). Members include Purdue University, the British Library(BL), the California Digital Library(CDL), and 27 others globally (<https://www.datacite.org/>).

Persistent links to online academic journals are created through the CrossRef agency of DOI. According to their 2016-17 Annual Report at <https://www.crossref.org/annual-report/> there are over 90 million records available in the database. A few of the numerous global member

organizations include the German Max Planck Institut fur Europaische Rechtsgeschichte (MPIeR), the Society of Antiquaries of Scotland, and the Mayo Clinic Scientific Press in the United States. In 2018 the annual membership fee is \$275 to \$50,000, based on the revenue of the publishing organization. There are additional charges for content which is categorized by type (journal articles, working papers, datasets, chapters, etc.), depositing DOIs, and for other services such as Similarity Check and Crossmark. High volume discounts are offered on some items and updating metadata is free (<http://www.crossref.org/fees>).

DOIs have gained a large user base. As of 2015, the IDF consisted of 30 members from 16 countries (www.doi.org). "There are already tens of millions of registered DOIs and the resolving mechanism is being used a few million times each month" (Hilse & Kothe, 2006, p. 24). Jan Brase and Irina Sans from the German National Library of Science and Technology, along with Michael Lautenschlager from the German Climate Computing Centre, in their article "The Tenth Anniversary of Assigning DOI Names to Scientific Data and a Five-Year History of DataCite," point out that there are 350 data centers from all over the world now working with DataCite and more than four million DOI names have been registered with them (2015).

Archival Resource Key (ARK)

The California Digital Library (CDL) maintains the Archival Resource Key (ARK) which was developed in 2001 by John Kunze. ARKs are permanent URLs and are part of a complete software and protocol system for resolving PIDs to their objects. They are versatile, transparent, can identify most anything, free to assign, and are designed to be integrated into other identifier systems and vice versa. Additional advantages are that ARKs resolve globally and can be hosted on local servers or moved to other servers.

According to their website at <https://wiki.ucop.edu/display/Curation/ARK>, there are four distinct categories of objects to which ARKs are given:

- " digital objects – documents, databases, images, software, websites, etc.
- physical objects – books, bones, statues, etc.
- living beings and groups – people, animals, companies, orchestras, etc.
- intangible objects – places, chemicals, diseases, vocabulary terms, performances, etc.

(<https://confluence.ucop.edu/display/Curation/ARK>)"

Emma Tonkin in her article "Persistent identifiers: Considering the options," says that ARKs are different than typical URLs and they are "used to retrieve three things:

- the object itself
- its metadata (by appending a single "?")
- and a [maintenance] commitment statement from its current provider (by appending "??")

[to the end of an ARK]" (2008, p. 6)

Like DOIs, ARKs use a Name Assigning Authority Number (NAAN), but it has three main parts that then make it actionable. Sébastien Peyrard and Jean-Philippe Traroni from the Bibliothèque nationale de France (BnF), along with John Kunze from the California Digital Library discuss these components in their article "The ARK Identifier Scheme: Lessons Learnt at the BnF and Questions Yet Unanswered." An ARK is an identifier consists of the following parts:

- identifier scheme ("ark:")
- NAAN
- the ARK name itself (2006)

Example: <http://chach.edu/ark:/NAAN/4489-xth-3887/mjrm>

The Bibliothèque nationale de France (BnF) adopted the ARK system in 2006. It wanted something that could be integrated into its local system and housed on its server. It nixed DOIs because of the fee for the identifiers, and because to use handles it would have to install the software for them. PURLs were out of contention due to the identifiers being held externally. What sold BnF on ARKs was that they could be implemented and maintained on their local server (Peyrard, et al., 2006).

It was not a completely smooth implementation process.

"First, the use of a reverse proxy conflicted with the IT operations requirements. Second, to detect, change and redirect the requests, the ARK resolver had to implement some domain-specific rules. This was dangerous for the security and maintainability of the whole system" (Peyrard et al., 2006, p. 90).

Two years down the road, the BnF developed a more generic system that added in three modules designed to fix the problems. For the most part, institutions using ARK PIDs do not have these types of problems, but it demonstrates the need for closer examination of the technical components of potential systems before adopting one (Peyrard et al., 2006).

EZID service

In addition to ARKs, CDL offers EZID. This is a self-serve service for generating PIDs that include ARKs, DataCite DOIs, and CrossRef DOIs. Not only is it possible to mint identifiers, but they can be updated here. For educational and non-profit institutions there are five levels of services available with annual subscription fees charged in 2018 as follows for ARKs only:

- Associate/Undergraduate granting institution- \$300
- Masters granting institution- \$600

- Research institution or non-profit organization- \$1,500
- Research group, department, or team-\$500
- Consortium of three or more institutions or organizations will receive a 20% discount on fees listed in the first three levels (<http://ezid.cdlib.org/home/pricing>)

This includes the creation and maintenance of up to one million identifiers per year.

Reluctance to using external registries

Not everyone agrees that external systems are the way to go. Nikos Askitas is the Director of Data and Technology at the International Zeolite Association (IZA), a private not-for-profit organization in Germany. According to their website at <http://www.iza.org/en/webcontent/about/index>, IZA performs national and international labor market research and is associated with the University of Bonn.

Askitas recognizes that using URLs to locate objects and cite them is ineffective and feels that PIDs are a good solution, but he does not like the idea of a registration agency such as the IDF retaining them. "...you should be able to take your objects elsewhere if you for some reason want to **without** breaking functionality of existing references. Otherwise you must rename your objects, but this violates persistency in more than one way" (2010, p.6). He sees PIDs as a solution to the problem of hidden metadata and thinks that technology is not sufficiently developed to be a fail-safe answer. He says that "Persistency on a technical/digital level is nothing but the attempt to install a centralized global webmaster" (2010, p. 7), and that we need to be safeguarding ourselves against obsolescence instead.

Interviews with experts

Interviews were conducted with the following experts: September 30, 2015 with Leslie Johnston (Director of Digital Preservation at the National Archives and Records Administration)

and Mary Elings (Head of Digital Collections at the Bancroft Library of the University of California-Berkeley); October 7, 2015 with Joan Starr (EZID Service Manager); February 23, 2017 with Ivey Glendon (staff member in Metadata Analysis and Design, Collections Access and Discovery at the University of Virginia); February 26, 2017 with Ardie Bausenbach (Senior Automation Planning Specialist at the Library of Congress); and March 7, 2017 with Dr. Dave Fearon, consultant for the Johns Hopkins University (JHU) Data Management Services, which is part of the Sheridan Libraries.

These interviews provide perspectives on the current best practices of using PIDs for discovery, access, and preservation of digital objects, the challenges faced in implementing identifier systems, in-house versus external ones (which is preferred and why), and how reliable resolution is for PURLs, DOIs, and ARKs.

In-house systems

For a large government agency such as the United States National Archives and Records Administration (NARA), using a system developed in-house appears to be effective. Leslie Johnston, the Director of Digital Preservation, says that data (especially from libraries) comes with PIDs generated from their Digital Asset Management System (DAMS) and do not necessarily need another ID. This data is not published for the public therefore NARA does not use DOIs or the handle system.

The agency has been collecting digital data since the 1970s. There are 200 agencies that submit 60 different types of digital files to the Digital Preservation department. To maintain authenticity, her office never changes the PIDs. Whatever PIDs come in from each agency is kept as is. There are other IDs aside from these that are used for accession and transfer. The challenge is keeping up with the volume of objects that need to be tracked. There is an ever-

increasing number of PIDs (8-10 per object) that pertain to geographic information, author, content, location, legal information, data regarding the digitization process, and others (personal communication, September 30, 2015).

The University of Virginia Library (UVA) uses a combination of in-house, open source, and commercial products for content and metadata management, which possess a variety of benefits and drawbacks. Ivey Glendon who works with Metadata Analysis and Design, Collections Access and Discovery at the University described the metadata advantages and shortcomings of using an in-house solution compared to open source or commercially available products.

For instance, systems such as TrackSys are customizable by the institution to meet the needs of distinct and local collections. "A drawback for locally developed systems, and a frequent one, is that staff turnover can make working with a local solution challenging, potentially leaving the department without a way to continue working with it." Also, neglect of keeping changes updated is common (personal communication February 23, 2017).

For open source software solutions and commercial products, UVA has adopted the open source Avalon system for streaming media asset management and subscribes to the commercial products ArchivesSpace (<http://www.lyrasis.org/lyrasisdigital>) and the add-on license SharedShelf through ArtStor (<http://internal.artstor.org/shared-shelf/s-html/shared-shelf-home.shtml>). While commercially produced systems charge a membership fee, they can be preferred solutions, given the benefits of strong community and developer support. For these reasons, Glendon favors out-of-the-box systems that can handle multiple types of metadata (personal communication, February 23, 2017).

At the Library of Congress (LC), Ardie Bausenbach works as a Senior Automation Planning Specialist. She says that the Library of Congress relies on many types of persistent identifiers, one of which is the use of actionable handles to manage digital objects. They resolve locally through the Library's handle server (<http://hdl.loc.gov>), a node of the CNRI handle server network; when needed, these identifiers can also be resolved through other handle servers on CNRI's distributed network. "In addition to handles, the Library provides LCCN Permalinks for bibliographic and authority records in the LC Online Catalog, as well as dereferenceable URLs through the Library's Linked Data Service (<http://id.loc.gov>) for access to commonly found Library standards and controlled vocabularies" (personal communication, February 26, 2017).

Handles help the Library support efficient control, identification, and reliable resolution of these digital assets, as well as other electronic content such as webforms and web pages as well. This ensures the integrity of the authenticity of the PIDs and their resolution. LC handles are consistently kept up-to-date for digital assets, so when content moves, it always remains accessible (personal communication, February 26, 2017).

The Library's handle server does not contain descriptive metadata for the Library's digital objects; for most LC handles, this metadata is maintained in LC Online Catalog records and in the Library's archival EAD finding aids, which contain the actionable handle links to LC's digital objects. Because handles do not change, these links in LC's widely distributed catalog records and finding aids ensure that national and international web references in educational content, bibliographies, blogs, publications, and sites such as Wikipedia, can access the Library's digital content no matter where it resides on LC systems (personal communication, February 26, 2017).

The Library of Congress, for example, "publishes and distributes many of its cataloging

records and cataloging-related tools and resources" nationally and internationally through the Cataloging Distribution Service (CDS) (<https://www.loc.gov/cds/>). These records are circulated within 24 hours of catalog updates. "Handles enable the Library to decouple the creation and updating of its catalog records from workflows that generate the Library's web content presentations." Individual catalog records are assigned a Library of Congress Control Number (LCCN), which facilitates persistent retrieval through the LCCN Permalink service (<https://lccn.loc.gov>). Catalog records can be accessed in web searches of the LC Online Catalog, through machine Z39.50/SRU searches (<https://www.loc.gov/z3950>), and through serializations provided by the Linked Data Service (<http://id.loc.gov>) (personal communication, February 26, 2017).

As unique content identifiers, handles enable intellectual entities to be associated with multiple representations over time. When content moves, LC updates its handles to reflect the current presentation URL. In some cases, a user's location (onsite or remote) affects whether they can view content. For some material, handles are assigned to each version of a digital object. Library policies and procedures govern handle assignment and maintenance. "As a core infrastructure component for access control and management of digital assets, handles provide the Library with semantic interoperability for its digital content across application environments" (personal communication, February 26, 2017).

External registries

Dr. Dave Fearon is a consultant for JHU Data Management Services, which is part of the Sheridan Libraries. He says that the University uses Dataverse to generate DOIs in some cases and that "We are also supporting our institution's use of the Open Science Framework, which

can also generate DOIs for “registered” projects that become public access” (personal communication March 7, 2017).

Dataverse and Open Science Framework are both open source repositories that offer DOI minting services. Dataverse is being developed at Harvard University’s Institute for Quantitative Social Science (IQSS) in collaboration with contributors worldwide (<https://dataverse.org>).

Open Science Framework is a product of the Center for Open Science, a non-profit organization that promotes the integrity and sharing of scientific research (<https://osf.io/>).

Archival Resource Key and EZID Services

Joan Starr, the EZID Service Manager at the University of California, Office of the President, described during an interview held October 7, 2015, the typical clients who use ARKS. They are often not-for-profit, for-profit, and international institutions who pay a subscription fee on a cost recovery basis. Research institutions usually get their own DOI prefix and then sponsor accounts within the university. This is an affordable way for researchers, scientists, and scholars to be able to get credit for their work. It also promotes re-use of content and makes it easier to calculate the impact of studies.

Starr feels that DOIs are better suited for publication citations because they are permanent. ARKs on the other hand have flexible metadata guidelines and can be deleted. Both are globally actionable IDs and to keep them working, any changes to the target URL must be kept current. This can be done through the EZID interface. If content has been deleted, the PID owner can use EZID to set up a "tombstone" as appropriate.

The Bancroft Library is part of the University of California, Berkeley, where Mary Elings is the Head of Digital Collections. She uses a combination of PIDs from ARKs (for finding aids)

and METs as part of the ingest process. The Bancroft mints their PIDs from an ARK system that is maintained locally, when they are not publishing to CDL.

While ARK is a smooth-running system, there can be difficulties. If an ARK is keyed in incorrectly, the object will be overwritten by the automatic system and subsequently go missing until the problem is found and fixed (personal communication, September 30, 2015).

Persistent Link Testing

Twenty PIDs for scholarly articles, archived websites, web pages, journals, objects, and others were chosen randomly for each system: PURLs, DOIs, and ARKs and tested December 7 and 14, 2015, October 28, 2016, and March 4, 2018. As can be seen in the table below, all three types performed less than 100% the first two test runs, with PURLs initially resolving 80% of the time and maxing at 85%, DOIs and ARKs at 95%. On the third run all three systems had improved. PURLs showed a 95% resolving rate and DOIs and ARKs all performed as expected. It appears that broken links were repaired, thus supporting visibility and accessibility.

Resolving Success Rates			
Type	December 2015	October 2016	March 2018
PURLs	80%	85%	95%
DOIs	95%	95%	100%
ARKs	95%	95%	100%

See Appendices A through C for testing details.

Conclusion

There are a tremendous number of digital objects being processed through cultural history institutions daily and the maintenance of these assets is foremost on the minds of those who are responsible for their care. Current best practices show that the assignment and use of persistent identifiers not only meet Federal requirements, but also keep assets visible, shareable, and authenticated.

Locally developed systems work for very large institutions such as NARA and the Library of Congress because they have the resources available and policies in place to fund continued improvement and assure that maintenance will be performed consistently. Smaller organizations who try to implement locally customized systems set themselves up for inherent problems such as URL decay, conflicting schemas, technical difficulties, and unreliable maintenance which presages the death of some of their assets.

PURLs, being comprised of URLs and prone to breaking, have shown a reliability of 80 to 95 percent. They use the HyperText Transfer Protocol for redirection, which will at some point be replaced with new technology that may render this PID ineffective.

DOIs are reliable 95 to 100 percent of the time and are more suitable for "fixed" items that will not be changed. The cost of DOIs through CrossRef is easy to estimate, but a bit of research is necessary regarding cost of registering a publishing name because the various RAs use their own business model for this.

California Digital Library's ARKs also resolves at 95 to 100 percent. They are versatile and can identify most anything. They are affordable as well and the self-serve aspect of EZID may make this system even more attractive.

While there are subscription fees for using external registries, the value of a more standardized system, the unique and actionable qualities of PIDs outweigh the problems of in-house systems. In the long run it could be worth the investment, particularly for smaller institutions. But regardless of whether a locally developed system is adopted, or an institution goes with an external registry, mapping updates must be done for objects to remain alive and accessible. Well maintained persistent identifiers are essential in the management of digital assets and in seeing them through future changes in metadata standards, technology, and culture.

References

- Askitas, N. (2010, June 28). What makes Persistent Identifiers persistent? In *Working Papers Series des Rates fur Sozial-und Wirtschaftsdaten*, (147), 1-8. Retrieved from http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_147.pdf
- Brase, J., I. Sens, & M. Lautenschlager. (2015, January/February). The tenth anniversary of assigning DOI names to scientific data and a five-year history of DataCite. In *D-Lib Magazine*, 21(1/2). doi: 10.1045/january2015-brase
- Chillingworth, M. (2004, July/August). Identity check. In *Information World Review*, 204. Dataverse. Retrieved from <https://dataverse.org>
- Ducut, E., F. Liu, & P. Fontelo. (2008, June). An update on Uniform Resource Locator (URL) decay in MEDLINE abstracts and measures for its mitigation. In *BMC Medical Informatics and Decision Making*, 8(23), 1-8. doi:10.1186/1472-6947-8-23
- Gul, S., I. Mahajan, & A. Ali. (2014). The growth and decay of URLs citation: A case of an online Library and Information Science journal. In the *Malaysian Journal of Library & Information Science* 19(3), 27-30.
- Hakala, J. (2010). Persistent identifiers-an overview. In *The KIM Technology Watch Report*, 1-16. Retrieved from <http://metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/>
- Hilse, H.W., J. Kothe. (2006). *Implementing Persistent Identifiers: Overview of concepts, guidelines, and recommendations*. Consortium of European Research Libraries, European Commission on Preservation and Access, 17-35. Retrieved from http://webdoc.sub.gwdg.de/edoc/ah/2006/hilse_kothe/urn%3Anbn%3Ade%3Agbv%3A7-ISBN-90-6984-508-3-8.pdf

IZA World of Labor. (2015). Retrieved from <http://www.iza.org/en/webcontent/about/index>

Lunghi, M., C. Cirinnà, & E. Bellini. (2013, January). Trust and persistence for internet resources. In the *Journal of Law Information and Science* 4(1), 375-389. doi: 10.4403/jlis.it-5494

Lynch, C. (1997-1998). Identifiers and their role in networked information applications. In *Cause/Effect: A practitioner's journal about managing and using information resources on college and university campuses*. Retrieved from <http://net.educause.edu/ir/library/html/cem/cem97/cem9743.html>

Lyons, S. (2005). Persistent identification of electronic documents and the future of footnotes. In the *Law Library Journal* 97(4), 681-693. Retrieved from http://www.aallnet.org/mm/Publications/llj/LLJ-Archives/Vol-97/pub_llj_v97n04/2005-42.pdf

Miller, P. (2000, June). Metadata: I am a name and a number. In *Ariadne* (24). Retrieved from <http://www.ariadne.ac.uk/issue24/metadata>

National Information Standards Organization. (2016, July). NISO RP-25-201x2B. Retrieved from https://groups.niso.org/apps/group_public/download.php/17091/NISO%20RP-25-2016%20Outputs%20of%20the%20NISO%20Alternative%20Assessment%20Project.pdf

Neumann, J. & J. Brase. (2014, July 7). DataCite and DOI names for research data. In *Journal of Computer-Aided Molecular Design*. doi: 101007/s10822-014-9776-5

Nicholas, N, N. Ward, & K. Blinco. (2009, January/February). A policy checklist for enabling persistence of identifiers. In *D-Lib Magazine*. Retrieved from <http://dlib.org/dlib/january09/nicholas/01nicholas.html>

Open Science Framework. Retrieved from <https://osf.io>

Paskin, N. (2010). Digital Object Identifier (DOI) system. In the *Encyclopedia of Library and Information Sciences*, 3rd ed, 1-7. doi: 10.108/E-ELIS3-120044418

Peyrard, S., J.P. Tramoni, & J. Kunze. (2006). The ARK Identifier Scheme: Lessons learnt at the BnF and questions yet unanswered, 83-94. Retrieved from <http://dcpapers.dublincore.org/pubs/article/view/3704>

Stebbins, M. (2013, February 22). Expanding public access to the results of federally funded research. Retrieved from (<https://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>)

Tonkin, E. (2008, July). Persistent identifiers: Considering the options. Retrieved from <http://www.ariadne.ac.uk/issue56/tonkin/>

Wang, J. (2007, September 1). Digital object identifiers and their use in libraries. In *Serials Review* 33, 161-164. doi: 10.1016/j.serrev.2007.05.006

Wolfson, R. (2015). Data Driven Cultures. In *Big data now*, 1, 1. O'Reilly Media, Inc., Sebastopol.

Appendix A

R = resolved F = failed

PURL Link Testing			12/07/2015		10/28/2016		03/04/2018	
#	Object Name	PID Link	R	F	R	F	R	F
1	"OFR blog : the official blog of the Federal Register 2.0 project" (August 21, 2012 and January 13-February 4, 2014)	http://purl.fdlp.gov/GPO/gpo27786	X		X		X	
2	"National Water Program: Best Practices" by the United States Environmental Protection Agency, (2007-2009)	http://purl.fdlp.gov/GPO/gpo56556	X		X		X	
3	Gulf Coast Ecosystem Restoration Council website (2013-2015)	http://purl.fdlp.gov/GPO/gpo56382	X		X		X	
4	The National Archives blogs (January 13-February 3, 2014)	http://purl.fdlp.gov/GPO/gpo29577	X		X		X	
5	"Introduction to Persistent Uniform Resource Locators" (1996) by Keith Shafer, Stuart Weibel, Erik Jul, and Jon Fausey	http://purl.oclc.org/doc/s/inet96.html		X	X		X	
6	"Two paths to interoperable metadata" (2003) by Carol Jean Godby, Devon Smith, & Eric Childress	http://purl.oclc.org/dc2003/03godby.pdf		X		X	X	
7	"Journal for the education of the young scientist and giftedness" (2013) by the Bülent Ecevit Üniversitesi	http://bibpurl.oclc.org/web/75626	X		X		X	
8	"Ethics manual" (2012) by Lois Snyder	http://bibpurl.oclc.org/web/11556	X		X		X	
9	"China's iGeneration: Cinema and moving image culture for the twenty-first century" (2006) by Matthew D. Johnson, Keith B. Wagner, Kiki Tianqi, and Luke Vulpiani	http://bibpurl.oclc.org/web/67779	X		X		X	
10	Zepheira's RELAX-NG schema link on https://purl.org/docs/help.html	http://purlz.org/project/purl/documentation/attachPurls.rng		X		X		X

PURL Linking Test			12/14/15		10/28/16		03/04/18	
#	Object Name	PID Link	R	F	R	F	R	F
11	"Implementing persistent identifiers : overview of concepts, guidelines and recommendations" (2006) by Hans-Werner Hilse and Jochen Kothe	http://resolver.sub.uni-goettingen.de/purl/?isbn-90-6984-508-3	X		X		X	
12	"PURLs: Persistent Uniform Resource Locators" (1996) by Stuart Weibel	http://purl.oclc.org/docs/new_purl_summary.html		X		X	X	
13	Dublin Core Metadata Initiative home page	http://purl.org/DC	X		X		X	
14	Proposed crosswalk object in "A repository of metadata crosswalks" (2004) by Carl Jean Godby, Jeffrey A. Young, & Eric Childress	http://purl.org/net/mets_crosswalk_profile	X		X		X	
15	"Improving metadata quality: Augmentation and recombination" (2006) by Diane Hillman, Naomi Dushay, & Jon Phipps	http://purl.org/metadat aresearch/dcconf2004/papers/Paper_21.pdf	X		X		X	
16	"Describing and analyzing the recordkeeping capabilities of metadata sets" (2004) by Joanne Evans & Lori Lindberg	http://purl.org/metadat aresearch/dcconf2004/papers/Paper_27.pdf	X		X		X	
17	"Two paths to interoperable metadata" (2003) by Carol Jean Godby, Devon Smith, & Eric Childress	http://purl.oclc.org/dc2003/03godby.pdf	X		X		X	
18	"Bulletin of the German Historical Institute-Supplement" (2004-) by the German Historical Institute	http://bibpurl.oclc.org/web/13625	X		X		X	
19	"World War I: Trenches on the Web: An Internet history of the Great War" (1996) by Mike Iavarone	http://bibpurl.oclc.org/web/324	X		X		X	
20	"Apeiron: Studies in infinite nature" (1987, 1990s) by the Universidad Autonoma de Zacatecas	http://bibpurl.oclc.org/web/6131	X		X		X	

Error Messages for Failed PURL Links

Link	Error Message	Date
5	<p>Error sourcing resource resource ffcpl:/resources/inet96.html not found in module urn:org:purl:documentation ffcpl:/resources/inet96.html executing subrequest:</p> <p>[SOURCE jetty://purl.org/docs/inet96.html in urn:org:ten60:netkernel:tpt:http as com.ten60.netkernel.urii.aspect.IAspectBinaryStream]</p>	12/07/15
6	Problem loading the page	12/07/15 10/28/16
10	<p>Server Error in '/' Application.</p> <p><i>Runtime Error</i> Description: An exception occurred while processing your request. Additionally, another exception occurred while executing the custom error page for the first exception. The request has been terminated.</p> <p>A different error on the third test: <i>Page not found</i> We're sorry, but we were unable to locate the page you requested.</p>	12/07/15 10/28/16 03/04/18
12	<p>Error sourcing resource resource ffcpl:/resources/new_purl_summary.html not found in module urn:org:purl:documentation ffcpl:/resources/new_purl_summary.html executing subrequest:</p> <p>[SOURCE jetty://purl.org/docs/new_purl_summary.html in urn:org:ten60:netkernel:tpt:http as com.ten60.netkernel.urii.aspect.IAspectBinaryStream]</p>	12/14/15 10/28/16

Appendix B

R = resolved F = failed

DOI Link Testing			12/07/15		10/28/16		03/04/18	
#	Object Name	PID Link	R	F	R	F	R	F
1	"Legal scholarship and digital publishing: Has anything changed in the way we do legal research?" (2002) by E. Dana Neacsu	doi: 10.1300/J113v21n02_06	X		X		X	
2	"The Fedora Project: An open-source digital object repository management system" (2003) by Thornton Staples, Ross Wayland, and Sandra Payette	doi: 10.1045/april2003-staples	X		X		X	
3	"If we share data will anyone use them? Data sharing and reuse in the long tail of science and technology" (2013) by Jillian Wallis, Elizabeth Rolando, and Christine Borgman	doi: 10.1371/journal.pone.0067332	X		X		X	
4	"The data curation continuum: Managing data objects in institutional repositories" (2007) by Andrew Treloar, David Groenewegan, and Catherine Harboe-Lee	doi: 10.1045/september2007-treloar	X		X		X	
5	"Regulation of gene expression by reactive oxygen" (1999) by Timothy Dalton, Howard Shertzer, and Alvaro Puga	doi: 10.1146/annurev.pharmtox.39.1.67	X		X		X	
6	"Choosing a digital asset management system that's right for you" (2009) by Deborah Kaplan	doi: 10.1080/15332740902897360	X		X		X	
7	"Hidden in plain sight? Records for on-demand academic public lectures in OCLC WorldCat: A Survey" (2010) by George Wrenn	doi: 10.1080/01639374.2010.492465	X		X		X	

DOI Link Testing			12/07/15		10/28/16		03/04/18	
#	Object Name	PID Link	R	F	R	F	R	F
8	"Making data citable: DataCite" (2014) by Jan Brase	doi: 10.1007/978-3-319-00026-8	X		X		X	
9	"ICT in education development in Africa" (2010) by Chijioke J. Evoh	doi: 10.4018/978-1-61520-847-0-ch017		X		X	X	
10	"Spatial pattern and process in forest stands within the Virginia piedmont" (2005) by Daniel Druckenbrod, Herman Shugart, and Ian Davies	doi: 10.1658/1100-9233(2005)016[0037:SPAPIF]2.0.CO;2	X		X		X	
11	"Wheelchair controlled by hands gestures recognition: A natural user interface" (2009) by Arminda Guerra Lopes	doi: 10.4018/978-1-4666-8833-9.ch014	X		X		X	
12	"Variation in terrestrial habitat use by four pool-breeding amphibian species" (2005) by Jonathan V. Regosin, Bryan S. Windmiller, Rebecca Homan, and J. Michael Reed	doi: 10.2193/0022-541X(2005)69[1481:VITHUB]2.0.CO;2	X		X		X	
13	"Restitution of works of art pursuant to private and public international law" (2001) by Wojciech Kowalski	doi: 10.1163/ej.9789041116093.009-244	X		X		X	
14	"Farewell to an idea: Episodes from a History of Modernism" (2000) by Andrew Hemingway and Paul Jaskot	doi: 10.1163/156920600794750810	X		X		X	
15	"The universal survey museum" (2013) by Carol Duncan and Alan Wallach	doi: 10.1111/j.1467-8365.1980.tb00089.x	X		X		X	
16	"The Human Rights Documentation Initiative at the University of Texas Libraries" (2010) by Christian Kelleher, T-Kay Sangwand, Kevin Wood, & Yves Kamuronsi	http://tandfonline.com/doi/abs/10.1080/13614576.2010.528342	X		X		X	

DOI Link Testing			12/07/15		10/28/16		03/04/18	
#	Object Name	PID Link	R	F	R	F	R	F
17	"Agile ethics for massified research & visualization" (2012) by Fabian Neuhaus & Timothy Websmoor	http://dx.doi.org/10.1080/1369118X.2011.616519	X		X		X	
18	"Personal Internet archives & ethics" (2014) by Strine Lomborg	http://rea.sagepub.com/lookup/doi/10.1177/1747016112459450	X		X		X	
19	"Twitter use by the U.S. Congress" (2006) by Jennifer Holbeck, Justin M. Grimes, & Anthony Rogers	http://doi.wiley.com/10.1002/asi.21344	X		X		X	
20	"Managing sameness" (2008) by Karen Coyle	http://dx.doi.org/10.1016/j.acalib.2008.07.012	X		X		X	

Error Message for Failed DOI Link

Link	Error Message	Dates
9	<p>DOI Not Found</p> <p>10.4018/978-1-61520-847-0-ch017</p> <p>This DOI cannot be found in the DOI System. Possible reasons are:</p> <ul style="list-style-type: none"> • The DOI is incorrect in your source. Search for the item by name, title, or other metadata using a search engine. • The DOI was copied incorrectly. Check to see that the string includes all the characters before and after the slash and no sentence punctuation marks. • The DOI has not been activated yet. Please try again later and report the problem if the error continues. 	<p>12/07/2015</p> <p>10/28/2016</p>

Appendix C

R = resolved F = failed

ARK Link Testing			12/07/2015		10/28/2016		03/04/2018	
#	Object Title	PID Link	R	F	R	F	R	F
1	Photograph entitled "Crowd gathered around bomb explosion site. Corner of Steuart and Market Streets, San Francisco. During Preparedness Day Parade, July 22, 1916"	http://www.oac.cdlib.org/ark:/13030/tf2h4nb6vh/?brand=oac4	X		X		X	
2	"Campus architects and engineers," the University of California Los Angeles (1946-1990)	http://www.oac.cdlib.org/findaid/ark:/13030/kt0q2nd098/	X		X		X	
3	"Sketchbook 8" by Edgar Degas, (1870-1913)	http://gallica.bnf.fr/ark:/12148/btv1b84386844	X		X		X	
4	"Repeating pattern paper" by Jacquemart et Bénard (1799)	http://gallica.bnf.fr/ark:/12148/btv1b69003605.r=	X		X		X	
5	Photograph "Cutting the bicentennial cake, " Oxnard Public Library (1976)	http://www.oac.cdlib.org/ark:/13030/kt7s200803/?brand=oac4	X		X		X	
6	"The ARK persistent identifier scheme" (2009) by John Kunze and Richard P.C. Rogers	http://ark.cdlib.org/arkspec.pdf		X		X	X	
7	"Proceedings of the 3rd ECDL Workshop on Web Archives" (2003)	http://www.cdlib.org/inside/diglib/ark/arkcdl.pdf	X		X		X	
8	"Nice Opaque Identifier Generatar Commands (NOID)" (2006) by John Kunze and Michael A. Russell	http://www.cdlib.org/inside/diglib/ark/noid.pdf	X		X		X	
9	"Tom Bills Sculpture" (2013) by Tom Bills & David Roth	http://escholarship.org/uc/item/18w649qj	X		X		X	

ARK Link Testing			12/07/2015		10/28/2016		03/04/2018	
#	Object Title	PID Link	R	F	R	F	R	F
10	"Persistence of Vision" by Gina Werfel (2011)	http://escholarship.org/uc/item/8xn781rt	X		X		X	
11	"Luigi Boccherini: Dictionary of Persons, Places & Terms" (2010) by Jaime Tortella	http://escholarship.org/uc/item/1p00w6bd	X		X		X	
12	"Using Archival Resource Keys (ARKs) for Persistent Identification" (2008) by Mark Edward Philips	http://digital.library.unt.edu/ark:/67531/metadc28359/	X		X		X	
13	Glenn Supercomputer at the Ohio Supercomputer Center of the Ohio State Government	http://osc.edu/ark:/19495/hpclph70	X		X		X	
14	Aurelio Bulosan Papers, 1949-1979 collection at Archives West (Orbis Cascade Alliance)	http://nwda.orbiscascade.org/ark:/80444/xv48920	X		X		X	
15	Basalt Head, Egyptian, Late Period at the National Museum of Natural History	http://n2t.net/ark:/65665/31011d93c-466b-4a9e-9576-379bbca463b2	X		X		X	
16	Sioux Bear Claw Necklace at the National Museum of Natural History	http://n2t.net/ark:/65665/389153373-9b91-4fd2-94c9-23910d9c5f2f	X		X		X	
17	African Limba Robe at the National Museum of Natural History	http://n2t.net/ark:/65665/3e62c64cc-071f-4ea6-9978-e2e9145affe2	X		X		X	
18	"Green Marketing: A study of consumer perceptions and preferences in India" (2013) by Mayank Bhatia & Amit Jain	http://escholarship.org/uc/item/5mc39217	X		X		X	
19	"The study of second language acquisition" (1994) by Rod Ellis	http://escholarship.org/uc/item/6wq540t3	X		X		X	
20	"Identifying and misidentifying the Brown Recluse spider" (1998) by Rick Vetter	http://escholarship.org/uc/item/6sj701ns	X		X		X	

Error Message for Failed ARK Link

Link #	Error Message	Dates
6	Not Found The requested URL /arkspec.pdf was not found on this server. Additionally, a 404 Not Found error was encountered while trying to use an ErrorDocument to handle the request.	12/07/2015 10/28/2016