Exploring the Durability of PURLs and Their Alternatives Working Group

A Working Group of the Depository Library Council

TO: Jen Kirk, Chair, Depository Library Council

FROM: Will Stringfellow, Chair, Exploring the Durability of PURLs and Their Alternatives Working Group

Will

SUBJECT: Exploring the Durability of PURLS and Their Alternatives Working Group Final Report and Recommendations

DATE: September 26, 2023

As Chair of the Exploring the Durability of PURLs and Their Alternatives Working Group, I am honored to transmit to you the final report and recommendations of the Working Group.

The Exploring the Durability of PURLs and Their Alternatives Working Group (Working Group) was charged with researching Persistent Uniform Resource Locators (PURLs), used by the Government Publishing Office (GPO) since 1998, and alternative types of persistent identifiers (PIDs). In addition to exploring various PIDs, the Working Group was charged with presenting their findings to Council along with any recommendations.

In its report the Working Group offers the following five recommendations:

1) The Depository Library Council accept the Principles for Persistent Identifiers and Persistent Identifier Systems for the Government Publishing Office contained in this report, and transmit them to GPO Director Halpern.
2) GPO enact the Principles for Persistent Identifiers and Persistent Identifier Systems through the following measures:
    a) GPO should seek to maintain stable systems for persistent identifiers and redirects in use within the CGP and other systems.
    b) GPO should develop and implement strategies to mitigate risk and improve management of the current system of PURLs, and any future system(s) of persistent identifiers and redirects, setting and assessing benchmarks going forward.
    c) To improve persistent access to the National Collection, GPO should increase the amount of content it manages through partnerships, contracts, interagency agreements, or ingestion into GovInfo, the FDLP Web Archive, or other mechanisms under local control.
    d) In order to increase the persistence and use of web content, particularly content that is not covered through interagency agreements, GPO should expand its FDLP Web Archive. This should include adding more websites as seed URLs to the collection so that more content is under GPO's direct control.
    e) GPO should explore technical solutions that will allow PID metadata to be visible and distinguished from the bibliographic metadata available in the Catalog of Government Publications.

   f) GPO should seek to work more closely with executive, legislative, judicial, and independent agencies to assure that their public information is collected, preserved, described, and made accessible for the National Collection.

3) GPO explore the potential opportunities of prospective PID systems for additional uses beyond the current implementation of GPO's PURL system, in order to improve services to FDLs.
4) GPO seek, as much as is possible within the Federal technology environment, to leverage interagency efficiencies in exploring technical solutions for needs related to PID system(s).
5) GPO offer training on PURLs that includes, but is not limited to, PURL referral reports, how to create a library's report profile, how the data can be used to promote Government resources, and features in GPO bibliographic records that relate to PURLs.

Recommendations #1 and #5 were previously transmitted to and accepted by Council at the 2023 Spring Meeting of the Depository Library Council. Council accepted and approved both recommendations on May 2, 2023 and formally recommended both to the Director of the Government Publishing Office on May 17, 2023 as Council recommendations. The Working Group now requests Council consider recommendations #2, #3, and #4 for transmittal to GPO.

The Working Group began addressing its charge in the spring of 2020. It investigated and researched PURLs and alternative PIDs, conducted focus group sessions with the depository library community, reviewed GPO's current implementation of PURLs, and offered presentations on PIDs. Because many PID schemas are similar, rather than recommend a particular one for GPO's consideration, the Working Group developed Principles to guide GPO's implementation and administration of PIDs for whichever schema best fits within GPO's technical infrastructure plans. All of this culminated in the Working Group's final report and recommendations.

I want to take a moment to thank all of the members of the Working Group, including current and former members of Council, members of the depository library community, and GPO staff for all of their amazing efforts which resulted in the final report and recommendations. Notably the Working Group began its work during the time the COVID-19 pandemic began, during which the members showed incredible dedication and commitment and without all of their efforts the final report and recommendations would not have been possible.

Attached is the Final Report of the PURL WG which includes recommendations. With this transmittal, the Exploring the Durability of PURLs and their Alternatives Working Group completes its charge.

CC: Exploring the Durability of PURLs and their Alternatives Working Group.

Final Report of the Depository Library Council's Working Group on Exploring the Durability of PURLs and Their Alternatives

September 2023

Depository Library Council

Advisory body to the Director of the U.S. Government Publishing Office

Exploring the Durability of PURLs and Their Alternatives Working Group Members

Depository Library Council

- Renée Bosman (June 1, 2020 – May 31, 2023)*
  University of North Carolina Chapel Hill
- Alicia Kubas (June 1, 2018 — May 31, 2021)*
  Library Services and Content Management, GPO (formerly University of Minnesota)
- Rick Mikulski (June 1, 2019 – May 31, 2023)*
  College of William & Mary
- Allen Moye (June 1, 2021 – May 31, 2024)
  DePaul College of Law
- Laura Sare, Secretary (June 1, 2019 – May 31, 2022)*
  Texas A&M University
- Robbie Sittel (June 1, 2018 — May 31, 2021)*
  Plano Public Library (formerly University of North Texas)
- Will Stringfellow, Chair (June 1, 2019 – May 31, 2022)*

Depository Library Community

- James R. Jacobs
  Stanford University
- Shari Laster
  Arizona State University

U.S. Government Publishing Office

- Alec Bradley
  Programs, Strategy and Technology
- Ashley Dahlen
  Library Services and Content Management
- Cynthia Etkin, Designated Federal Officer
  Office of the Superintendent of Documents

[*] *These members were on the Depository Library Council when the Working Group was established. As their Council term expired, they remained on the Working Group as members of the Depository Library Community.*

Executive Summary

The U.S. Government Publishing Office (GPO) implemented the use of Persistent Uniform Resource Locators (PURLs) in 1998 to provide Federal depository libraries (FDLs), and other users of Federal Government information, stable URL access to online digital content. In the more than twenty years since then, technological innovations have transformed the internet; libraries; and how digital publications are managed and accessed. Given the changed landscape, the Depository Library Council (DLC) established the *Exploring the Durability of PURLs and Their Alternatives Working Group* (WG) to investigate persistent identifier schemas, including PURLs, to determine if PURLs are still the best persistent identifier (PID) schema for GPO to provide permanent public access to born digital and digitally imaged Government information.

During the course of its work, the WG determined it was unable to recommend a specific PID system to GPO. Because many PID schemas are very similar and WG members lacked knowledge about their application within GPO's technical infrastructure, Principles were developed to guide GPO's implementation and administration of PIDs to ensure permanent public access to the U.S. Government's information.

PID and PID System Principles for the Government Publishing Office

PERSISTENT IDENTIFIER	PERSISTENT IDENTIFIER SYSTEMS
PIDs must be unique and provide direct long-term access to a specific digital document or object that is openly accessible.	PIDs must be part of a system that is stable, secure, and interoperable with other systems.
PIDs must be unchanging and never reused.	PIDs must have publicly accessible metadata.
PIDs must enable access to a digital resource under the control of a trusted entity or an established partner.	PIDs should work regardless of the users' access starting point and the access system or delivery service used.
PIDs must enable access to the specific object described in the metadata for the resource.	The PID system must resolve the identifier to the kernel metadata.

The WG incorporated GPO's vision of *America Informed* through free permanent public access to the U.S. Government's information into the development of these principles, and sought to consider the current state of Federal information technology infrastructure and cybersecurity. The four PID systems that most closely align with the Principles were evaluated: ARKs (Archival Resource Key), DOIs (Digital Object Identifier), Handles, and PURLs.

PIDs are considered persistent if the identifier is managed over time, does not change regardless of the location of the object it identifies, and continues to redirect the user even if the domain name or server location of the object changes. Implementing PIDs alone does not equate to the persistent access. For GPO to achieve permanent public access to digital objects within the *National Collection of U.S. Government Public Information*, whether they are hosted on GPO's digital infrastructure or by another institution or agency, both persistence of the identifier and of the object itself are needed. There must be a commitment to active administration of the PIDs and the PID system, this includes hosting content, continuously monitoring the location of digital objects not hosted by GPO and managing the PID system.

Persistence is challenging and multifaceted. As with most technologies, the policy, administrative, and organizational issues surrounding persistent identifiers are the most critical aspects of implementing a PID system and providing persistence. The real issue for the end user is the persistent access to the objects themselves. This requires coordinated persistence of the objects and the PIDs that direct to the objects. Since the internet is a constantly changing space, no matter how stable the PID system, simply pointing to a digital object on a website does not guarantee persistent access.

Recommendations

The PURL WG makes the following high-level recommendations:

1. The Depository Library Council accept the Principles for Persistent Identifiers and Persistent Identifier Systems for the Government Publishing Office contained in this report, and transmit them to GPO Director Halpern.

2. GPO enact the Principles for Persistent Identifiers and Persistent Identifier Systems through the following measures:
    a. GPO should seek to maintain stable systems for persistent identifiers and redirects in use within the *Catalog of U.S. Government Publications* and other systems.

    b. GPO should develop and implement strategies to mitigate risk and improve management of the current system of PURLs, and any future system(s) of persistent identifiers and redirects, setting and assessing benchmarks going forward.

    c. To improve persistent access to the National Collection, GPO should increase the amount of content it manages through partnerships, contracts, interagency agreements, or ingestion into GovInfo, the FDLP Web Archive, or other mechanisms under local control.

   d. In order to increase the persistence and use of web content, particularly content that is not covered through interagency agreements, GPO should expand its FDLP Web Archive. This should include adding more websites as seed URLs to the collection so that more content is under GPO's direct control.
   e. GPO should explore technical solutions that will allow PID metadata to be visible and distinguished from the bibliographic metadata available in the *Catalog of U.S. Government Publications*.
   f. GPO should seek to work more closely with executive, legislative, judicial, and independent agencies to assure that their public information is collected, preserved, described, and made accessible for the National Collection.

3. GPO explore the potential opportunities of prospective PID systems for additional uses beyond the current implementation of GPO's PURL system, in order to improve services to FDLs.
4. GPO seek, as much as is possible within the Federal technology environment, to leverage interagency efficiencies in exploring technical solutions for needs related to PID system(s).
5. GPO offer training on PURLs that includes, but is not limited to, PURL referral reports, how to create a library's report profile, how the data can be used to promote Government resources, and features in GPO bibliographic records that relate to PURLs.

Table of Contents

I. Introduction

Since March 1998, the Government Publishing Office (GPO) has used Persistent Uniform Resource Locators (PURLs) to provide Federal depository libraries (FDLs), and other users of Federal Government information, stable URL access to Federal information.

Not surprisingly, during the more than twenty years since GPO's implementation of PURLs, libraries and the internet have transformed. With the advances in technology and the immeasurable proliferation of born-digital and digital-only Federal information, libraries have examined and revised their collection policies and plans, and they are cognizant of, and responding to, the changing information seeking behaviors of their users. This has resulted in an increased interest in serving library patrons with online content. In seeking to meet the changing needs of their end-users, depository library coordinators have shown an interest in their collections and services becoming more, mostly, or all-digital, depending on those needs.

Given the library landscape and the amount of online content to which GPO links,[1] the Depository Library Council (DLC) agreed on the importance of evaluating various persistent identifier (PID) schemas, including PURLs, to ensure GPO's implementation is robust enough to best serve depository libraries and users of online Federal Government information.

Therefore, in late fall of 2019, the Depository Library Council (DLC) established the *Exploring the Durability of PURLs and Their Alternatives Working Group* (WG). It is charged with investigating persistent identifier schemas, including PURLs; reporting its findings; and providing recommendations to the DLC for consideration. This report fulfills that charge.

II. About Persistent Identifiers

At a basic level PIDs are unique labels assigned to identify digital objects. PIDs are considered persistent if the identifier is managed over time, does not change regardless of the location of the object it identifies, and continues to redirect the user even if the domain name or server location of the object changes. The practical purpose behind a PID is to seamlessly provide reliable and long-term access to an identified resource. A PID creates a relationship between an identified resource and a mechanism to access and retrieve that resource. A PID is only serviceable if the identified resource continues to be available,

[1] As of mid-May 2023, GPO had created 329,671 PURLs **to direct users to Federal digital content.**

meaning a PID is only persistent if it is maintained over time, and if the system supporting it migrates forward with changes in technology.[2] PIDs enable increased access to resources globally, both by increasing findability and by providing a stand-in for the URL to get to the content. This also allows PIDs to mediate between data sources and value-added services such as discovery tools.[3]

It may be helpful to understand PIDs with an example. In the print world, a catalog record describes a book and where it resides in the library. For the most part, the catalog record describing the book does not change. However, if the book is moved to another location in the library, the location information in the record should be updated to direct a user to the new location. From the time that a digital object is assigned a PID, the object's descriptive metadata generally does not change, but when the digital object's 'location' has changed, the PID should be updated to change how it directs, or 'resolves', access for the user.

The PID itself is a unique string of numbers and/or letters with internal syntax that references a referring target or location and is associated with a resolution service that connects the PID to the digital object.[4] For a link associated with a PID to be actionable, the target URL needs to be correct and the resolution service needs to be available. Ideally, a PID should be assigned to a single version of a digital object. There is a difference between a PID that identifies a unique digital object, and a persistent link (or redirect) that links to dynamic content, such as directory content that continuously updates. To make a resource persistently accessible through its identifier over time, each version of an available resource should have a unique PID, which should remain available over time.

A PID **service** includes the technical hardware and software to create and maintain PIDs, and must operate within a governance framework.[5] Broadly speaking, PID systems must be able to generate, assign, and resolve PIDs based on mapping and/or a template or rules; batch-update resolver mapping tables; store PIDs in the object's source system; export PIDs from the source system to import into other systems; present PIDs as human- and machine-readable data; and retrieve objects and metadata based on the PID both by

[2] Juha Hakala, "Persistent Identifiers: An Overview," *Technology Watch Report (TWR): Standards in Metadata and Interoperability* (October 13, 2010), http://www.persid.org/downloads/PI-intro-2010-09-22.pdf.

[3] Ulrich Schwardmann, Martin Fenner, Maggie Hellström, Hylke Koers, Hervé L'Hours, Brian Matthews, Raphael Ritz, Mario Valle, Mark van de Sanden, and Themis Zamani, *PID Architecture for the EOSC* (December 2020), https://op.europa.eu/s/oTo3.

[4] Schwardmann et al., *PID Architecture*.

[5] Schwardmann et al., *PID Architecture*.

human and machine-actionable methods.[6] Internally, PID systems include an administrative registration function, a maintenance function, and a resolution process that often interfaces with a global resolution service.[7]

In the print world, a union catalog is a tool that lists all copies in all libraries of a particular book. A union catalog helps libraries locate copies of books they do not hold locally. One of the strengths of some PID systems is that, much like a union catalog, each PID can resolve to multiple copies of the same object at multiple locations or servers.[8] Journal publishers often run multiple mirror sites so that users can access a particular article from a geographically close server, thus speeding up the access process for the end user and balancing server load across a network of servers instead of relying on one central server. In this instance, the PID supports a "multiple resolution" system as opposed to "simple resolution." This functionality could potentially strengthen redundant access to content in a variety of settings.

There are additional optional features that can be built into PIDs. For example, some PIDs can handle resolution for a fragment or subordinate target such as a section of a PDF or an embedded image. The functionality of these PIDs depends both on the PID service for resolution and the local web server's framework for delivering the result.[9]

At a mature stage of implementation, PID systems should address the openness of PID record storage, universality of applications, and alternatives to access.[10] Any process of certification or recognition of trustworthiness for PID systems should address stability and performance, long term persistence, and security and support.[11] With appropriate infrastructure choices, PIDs can be an intrinsic solution to the FAIR principles (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable).[12] PID systems that register, store, and retrieve metadata, and that utilize a global resolver, can be designed to provide the PID metadata in standardized and machine-actionable formats, although associated infrastructure to fully

[6] Lukas Koster, "Persistent Identifiers for Heritage Objects," *The Code4Lib Journal,* no. 47 (February 17, 2020), https://journal.code4lib.org/articles/14978.

[7] Schwardmann et al., *PID Architecture*.

[8] See https://www.doi.org/doi_handbook/3_Resolution.html#3.3.

[9] Schwardmann et al., *PID Architecture*.

[10] Schwardmann et al., *PID Architecture*.

[11] Schwardmann et al., *PID Architecture*.

[12] Schwardmann et al., *PID Architecture*.

leverage this approach has yet to be developed.[13] PID systems also hold the promise for transparent and actionable communication about resources, which could extend to rights management, preservation metadata, and more.[14] While PIDs do not inherently perform integrity checks, it is possible to design PID systems that complement an intrinsic identifier -- such as Public Key Infrastructure (PKI) currently employed by GPO[15] — which adds to a guarantee of authenticity.[16]

Because GPO uses a particular variation of a PURL system, it is helpful to understand how a PURL system works, and how this differs from other PID systems. At a basic level, a PURL system associates the PURL with a target URL, and allows for the creation of PURLs and the maintenance of associated URLs. PURLs can be assigned to any discrete resource to provide reliable access over time, assuming the resource when accessed has sufficient metadata or in situ information to be usable.[17] While some PIDs are directed through a centralized registration service, PURLs are directed through a service that is integrated with the system. This means that it can be operated independently from a centralized service, which offers flexibility but also adds significant challenges to longevity.[18]

GPO's PURL implementation fits in a broad category of private PID systems, alongside other institutions that have a tool to generate identifiers associated with their own namespace, and a tool to redirect the uniform resource identifier, or URI, for the PID to the resource URL.[19] While private PIDs can be costly and less robust than centralized systems, they offer more flexibility for the managing institution in terms of implementation.[20] This

[13] Tobias Weigel, Beth Plale, Mark Parsons, Gabriel Zhou, Yu Luo, Ulrich Schwardmann, Robert Quick, Margareta Hellström, and Kei Kurakawa, "RDA Recommendation on PID Kernel Information," *Research Data Alliance* (November 19, 2019), https://www.rd-alliance.org/system/files/RDA%20Recommendation%20on%20PID%20Kernel%20Information_final.pdf.

[14] Hakala, "Persistent Identifiers."

[15] For more information, see: https://www.gpo.gov/how-to-work-with-us/agency/services-for-agencies/public-key-infrastructure.

[16] Schwardmann et al., *PID Architecture*.

[17] Keith Shafer, Stuart Weibel, Erik Jul, and Jon Fausey, "Introduction to Persistent Uniform Resource Locators," *Proceedings from INet 1996* (June 27, 1996), https://web.archive.org/web/20160103053338/http://www.isoc.org/inet96/proceedings/a4/a4_1.htm.

[18] Larry Stone, *Handle Project: Competitive Evaluation of PURLs* (March 22, 2000), http://web.mit.edu/handle/www/purl-eval.html.

[19] Koster, "Persistent Identifiers."

[20] Koster, "Persistent Identifiers."

is distinct from the centralized system of PURLs that are managed by the Internet Archive — which took over OCLC's PURL service in 2016.[21]

Persistent Identifiers and Persistent Access

PIDs and persistent access are both necessary for GPO to guarantee no-fee permanent public access to digital government information. In order to guarantee permanent access, PIDs must stay the same over time, referencing the same object even if the web address of the object changes.

If an agency website is updated, a PURL that directs to a PDF on that site will not provide persistent access unless it redirects to the new site. Conversely, if the agency completely removes the PDF document, the PURL isn't what is impeding access; persistent access has been lost because the object is missing. The same would be true of any system that redirects from one URL to another.

A persistent identifier is not synonymous with persistent access, but is one requirement for it. From the end-user's perspective, the URL associated with the persistent identifier must resolve to and retrieve the object, even if the object's location changes. A commitment to active administration of the persistent identifier system is required and is of the utmost importance to ensure persistent access. Active administration includes hosting content, continuously monitoring the location of digital objects not hosted by GPO and managing the PID system. This enables the user to access the resource in their web browser via the persistent identifier.

The ultimate goal for GPO is permanent access to digital objects within the *National Collection of U.S. Government Public Information*, whether they are hosted on GPO's digital infrastructure or by another institution or agency. Both persistence of the identifier(s) and of the object itself are needed to provide persistent access. GPO is strongly committed to providing permanent public access, which includes the active administration of the PID system and the PIDs, along with monitoring the locations of associated digital objects.

Persistent Identifier Schemas

PID systems pair a structured identifier with a resolver that provides a URL pathway to desired content in a digital repository or on the web. The identifier remains constant over

[21] For more on this, see: http://blog.archive.org/2016/09/27/persistent-url-service-purl-org-now-run-by-the-internet-archive/

time and is able to resolve access to the associated resource or object, with a location that may or may not be persistent. As discussed earlier, a PID alone does not provide permanent access. Permanent access is dependent upon a chosen system that pairs a persistent identifier to an actionable resolver that can locate the object, which is typically not persistent. Permanent access also requires on-going maintenance of the system.

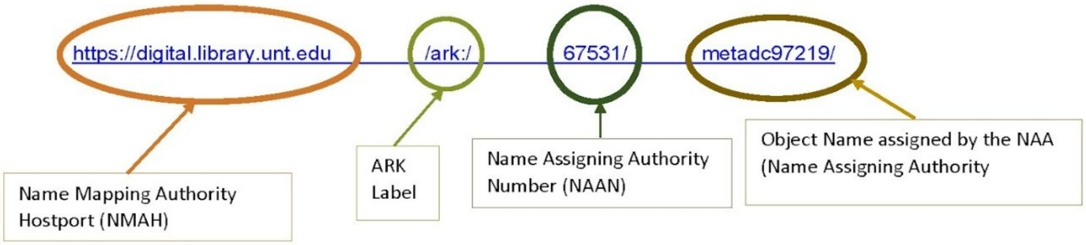
The Working Group members began by researching the characteristics and durability of persistent identifiers (PIDs). During the course of its work, the WG determined it was unable to recommend a specific PID system to GPO due to the lack of knowledge about various PIDs and PID systems and their application within GPO's technical infrastructure. The WG instead developed principles to guide GPO's implementation and administration of PIDs to ensure permanent public access to the U.S. Government's information. The group narrowed the evaluated number of PID systems to the four that most closely align with the WG's recommended principles. The four PID systems are described below.

PID Descriptions

Archival Resource Key (ARK)

Developed by the California Digital Library (CDL), an Archival Resource Key, or ARK, is a persistent, actionable naming scheme.[22] An ARK is a persistent identifier that links the user to three things — 1. the object, 2. its metadata, and 3. the commitment statement of the access provider. ARKs are managed and hosted on a locally controlled server, so there are no fees to assign or use ARKs.

The ARK system produces a special kind of URL that formalizes the roles of the Name Assigning Authority (NAA) and the Name Mapping Authority (NMA). ARKs act as web "redirects" and ordinary "get" requests from a secure web server.

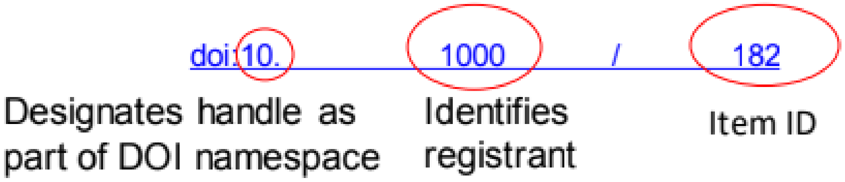

https://digital.library.unt.edu
/ark:/
67531/
metadc97219/
Name Mapping Authority Hostport (NMAH)
ARK Label
Name Assigning Authority Number (NAAN)
Object Name assigned by the NAA (Name Assigning Authority

[22] For more information, see: https://arks.org/about/.

Digital Object Identifier (DOI)

DOI is a framework for persistent identifiers that provides actionable, unique identification for digital resources.[23] DOI is a "digital identifier of an object" *not* an "identifier of a digital object." A DOI can be assigned to any physical, digital, or abstract entity.

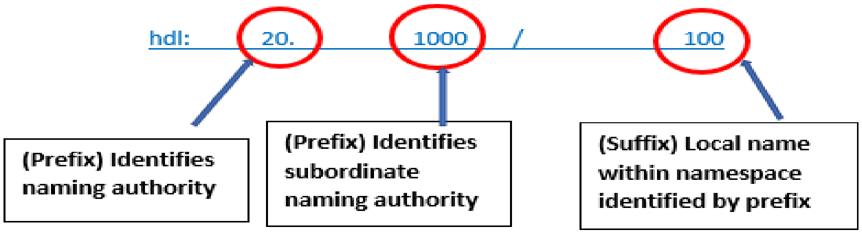
Each DOI is a series of numbers and punctuation that provide a unique identification for objects of any type. DOIs contain a prefix and a suffix separated by a slash (prefix/suffix) (for example, the doi 10.1000/182 has the prefix 10.1000 and the suffix 182). The prefix indicates the registered entity, and the suffix indicates the specific item within that entity's registered identifiers. Many DOIs are displayed as URLs.



Handles

The Handle system is a noncommercial decentralized identifier resolution system, established in 1995 and operated by the Corporation for National Research Initiatives (CNRI).[24] The Handle system includes a central registry to resolve URLs to the location of the digital object in a distributed environment. It is the basis for the DOI system. Initiatives can use commercial Handle licenses to operate their local systems, or their content management systems can operate a local Handle system.[25]

Handles consist of a prefix (for example, 20.1000/100), which indicates the naming authority, and a suffix, which indicates the local name for the digital resource.

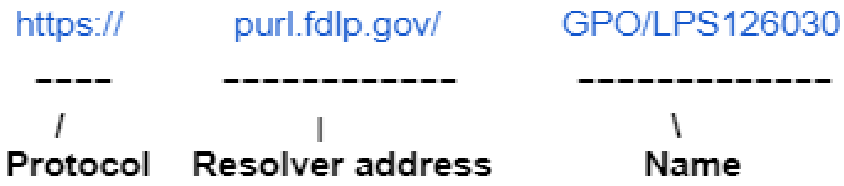

hdl:
20.
1000
/
100
(Prefix) Identifies naming authority
(Prefix) Identifies subordinate naming authority
(Suffix) Local name within namespace identified by prefix

[23] For more information, see: https://www.doi.org/the-identifier/resources/handbook/
[24] For more information, see: https://www.handle.net/
[25] For more information, see: https://project-thor.readme.io/docs/project-glossary

Persistent Uniform Resource Locator (PURL)

A Persistent Uniform Resource Locator or PURL is functionally a URL. Rather than pointing directly to the location of an Internet resource, a PURL points to an intermediate resolution service, which allows underlying web addresses of resources to change over time without affecting the PURL or its systems. PURLs are intended to provide long-term access and therefore do provide some sense of permanence but are not considered a long-term, archival mode for accessing resources.

PURL systems can be locally administered, or they can be part of the central purl.org service, which is currently administered by the Internet Archive.[26]



When comparing PURLs with a PID system such as Handles or DOIs, some of the differences are relevant to addressing the National Collection's digital preservation and access issues. For example, PURLs serve only as a redirect and thus are dependent on continued existence of the host domain that provides the persistent content.[27] A PURL server can also only resolve to its own host domain, meaning if GPO's PURL server fails all hosted PURLs also fail.[28]

All PURLs within a namespace must resolve using the same server.[29] PURLs can only resolve to one destination, not a cascading series of destinations to be used, in turn, depending on each destination's availability.[30] Other PIDs such as ARKs, Handles, and DOIs, can allow for distributed resolutions to alternate sources of the same object, different forms for the same content, and/or associated metadata.[31] For example, a journal

[26] https://blog.archive.org/2016/09/27/persistent-url-service-purl-org-now-run-by-the-internet-archive/

[27] Larry Stone, *Handle Project: Competitive Evaluation of PURLs* (March 22, 2000), http://web.mit.edu/handle/www/purl-eval.html

[28] In August 2009, GPO's PURL server suffered a significant hardware failure. While back-ups of critical files ensured no data loss, it took several weeks to reconfigure the software on a new server and rebuild the PURL resolution database. As a result of the outage of this critical system, GPO took steps to guarantee high availability and redundancy of the PURL application. For more information see: https://www.fdlp.gov/project/purl-enhancement-and-stabilization.

[29] Stone, *Handle Project*.

[30] Koster, "Persistent Identifiers."

[31] See: https://www.doi.org/doi_handbook/3_Resolution.html.

issue may have copies residing on multiple servers and the DOI can be configured to resolve to a specific copy based on network availability, server load, or other problems. PURLs are unable to provide this service. If digital deposit of National Collection content is implemented across the network of Federal depository libraries, PURLs would be insufficient as a solution to assign and resolve across a distributed system.

PID System Comparison Table

The table below compares some of the features and technical differences among the PID systems.[32] Additional detailed technical information for each PID system is available in the footnoted resources following the table.

	ARK[33]	DOI[34]	Handle[35]	PURL[36]
System administration by local institution	Y	N	Y	Y
Registry can be local or global entity	Y	Y	Y	Y
Generate PIDs locally	Y	N	Y	Y
Assign PIDs locally	Y	N	Y	Y
Resolved locally or through global entity	G/L	G	G/L	G/L
Scalability	Y	Y	Y	Y
Usage statistics	Y	Y	Y	Y
Global uniqueness	Y	Y	Y	Y
Interoperable/integrate with other PIDs	Y	Y	Y	Y
Long-term use	Y	Y	Y	Y
System has kernel metadata	Y	Y	Y	Y
PID metadata open and publicly accessible	Y	Y	Y	Y
PID metadata is separate from object metadata	Y	Y	Y	Y
PID metadata retrievable from local registry	Y	Y	Y	Y
PID metadata interoperable w/global registries	Y	Y	Y	Y
Resolves to multiple instances of digital object	Y	Y	Y	N
Fail safe process	Y	Y	Y	Y

[32] For more information on PID comparisons, see: Koster, "Persistent Identifiers."
[33] Information about ARKs: https://arks.org/
[34] Information about DOIs: https://www.doi.org/the-identifier/resources/handbook/
[35] Information about Handles: http://www.handle.net
[36] Information about PURLs: https://purl.archive.org/help. Note that the data for PURL reflect the centralized PURL.org service, not GPO's private PURL system.

III. Principles for Persistent Identifiers and Persistent Identifier Systems for the Government Publishing Office

The WG identified a need for principles to guide the determination of the best available solution for persistent access to Federal Government information in the current digital landscape.[37] Subsequently, the following principles for PIDs and PID systems were created for GPO. The WG incorporated GPO's vision of *America Informed* through free permanent public access to the U.S. Government's information into the development of these principles, and sought to consider the current state of Federal information technology infrastructure and cybersecurity.

PERSISTENT IDENTIFIER PRINCIPLES

PIDs must be unique and provide direct long-term access to a specific digital document or object that is openly accessible.

Each PID is associated with a unique and unduplicated URL that provides direct access. There is a one-to-one correlation between the PID and the digital object it identifies. PIDs should only be assigned to resources that will be preserved for the long term, which is to be understood over several hardware and software generations. GPO should employ PIDs only for digital objects that are openly accessible and can be viewed and downloaded without passing through a paywall.

PIDs must be unchanging and never reused.

The PID for the content must remain the same and should persist over time as long as the resource is publicly accessible. PIDs must never be reused for a different document or object.

PIDs must enable access to a digital resource under the control of a trusted entity or an established partner.

A core component of permanence and reliable performance is control of the digital content. GPO should employ PIDs only for digital objects that are under its control or under the control of an official partner, with a signed agreement requiring the transfer of content if the official partner is unable to maintain it in its current system.

[37] These Principles were first shared with the depository library community during the April 2022 Depository Library Council meeting. In June 2022, the International Standards Organization (ISO) issue Technical Specification 22943, Information and Documentation — Principles of Identification, which establishes core characteristics and expectations for PIDs and explains why they are important in information management.

PIDs must enable access to the specific object described in the metadata for the resource.

PIDs must resolve to the version of the resource indicated in the descriptive metadata that is validated or approved by GPO or an official partner. Resources must be disambiguated within the system so they are distinct and distinguishable from other resources. PIDs must be usable for different types of digital content and scalable for increasing amounts of content.

PERSISTENT IDENTIFIER SYSTEM PRINCIPLES

PIDs must be part of a system that is stable, secure, and interoperable with other systems.

The PID system(s) must provide reliable and stable access, with appropriate redundancy against outages and security issues. The system must connect and exchange information with tools, systems, and technologies used by GPO and others. PIDs must be able to be migrated with future system updates.

The PID system must resolve the identifier to the kernel metadata.

A review of the kernel metadata will enable validation that the PID resolves to the intended digital object.

PIDs should work regardless of the users' access starting point and the access system or delivery service used.

URLs that resolve through the PID system(s) should work reliably from any point of origin, allowing local systems to pass traffic through additional controls (such as a proxy server) as needed.

PIDs must have publicly accessible metadata.

PIDs must have metadata that is separate from the object metadata. PID metadata should be publicly accessible, retrievable, and interoperable with global registries.

IV. GPO's Implementation of PURLs

Prior to GPO's selection and implementation of PURLs in 1998, pilot testing was conducted using ARKs, DOIs, Handles, and PURLs. The decision to implement PURLs was based on the test findings:

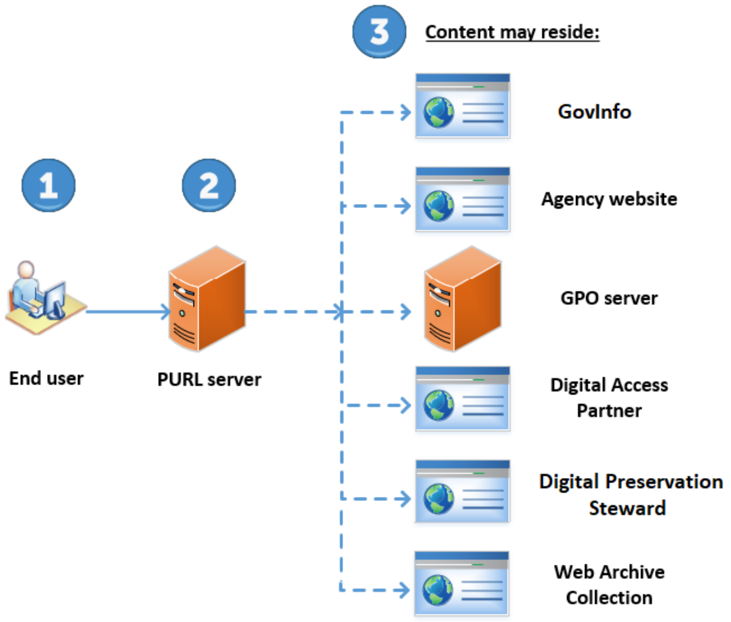
- PURL system was more intuitive for GPO staff to use;
- Setup and maintenance were much easier than with the other PID systems; and
- Costs for managing and maintaining the system could be absorbed into GPO's infrastructure.

In March 2007, GPO considered the possibility of implementing Handles. Ultimately, it was decided that PURLs would continue to be the persistent identifier of choice. The two biggest factors in this decision were cost and the need for additional staff.

GPO contracts out the PURL system administration, which includes security, backup, and report generation. However, the Library Services and Content Management (LSCM) catalogers create the PURLs and update the bibliographic records as necessary. LSCM's Office of Archival Management (OAM) staff maintain the PURLs to ensure persistent access to the digital Government information products.

As of mid-May 2023, there were 329,671 PURLs directing users to over 1,707 domains. Of these, 74% link to content that is under direct control of GPO. There are 117,673 PURLs directed to content in GovInfo alone.

Illustration of How a PURL Works



(1) An end user clicks on a PURL

(2) The GPO PURL server redirects the end user to ...

(3) Where the digital content resides

Creating PURLs

Superintendent of Documents policy dictates that persistent identifier technology be used and maintained to provide permanent links between bibliographic records and online Government information dissemination products. Specifically, PIDs will direct users to an archived copy of the online Government information dissemination product housed on a GPO server, partner server, or published to an agency's website.

Digital content is the primary format cataloged by GPO. PURLs are created during the cataloging process. Catalogers review the content to identify the best location to direct the PURL. They consider what works best at the time the PURL is created as well as for the future and they may opt to:

- Direct PURLs to a stable website, such as govinfo.gov or fraser.stlouisfed.org.
- Direct PURLs to an agency website.
- Harvest the individual file(s) for storage on a local server using software or a manual harvesting process and direct PURLs to the locally hosted content.
- Capture files or a website for the FDLP Web Archive and direct PURLs to the archived content.
- Direct PURLs to a Digital Access Partner website.
- Direct PURLs to a Digital Preservation Steward Partner website.
- Direct links to another trusted entity's web archive capture.

PURL links are found in GPO catalog records, which can be used by libraries in their local catalogs. PURLs are also used elsewhere, including LibGuides, research guides, course guides, and social media.

Verifying PURLs

There are two PURL verification processes that take place on a regular basis by the OAM. When errors are found, they find alternate locations to which PURLs are directed.

1. On the last day of each month a report is run on the PURL database and returns PURLs whose links resulted in 404 and 410 error codes that indicate content retrieval errors. On the first day of the subsequent month, OAM staff work through the report until all the errors are resolved.

2. On the 1st Friday of every month, after security patches are installed on PURL and other appropriate GPO servers, a system check is performed to ensure the hardware and software continue to operate and resolve or redirect as expected.

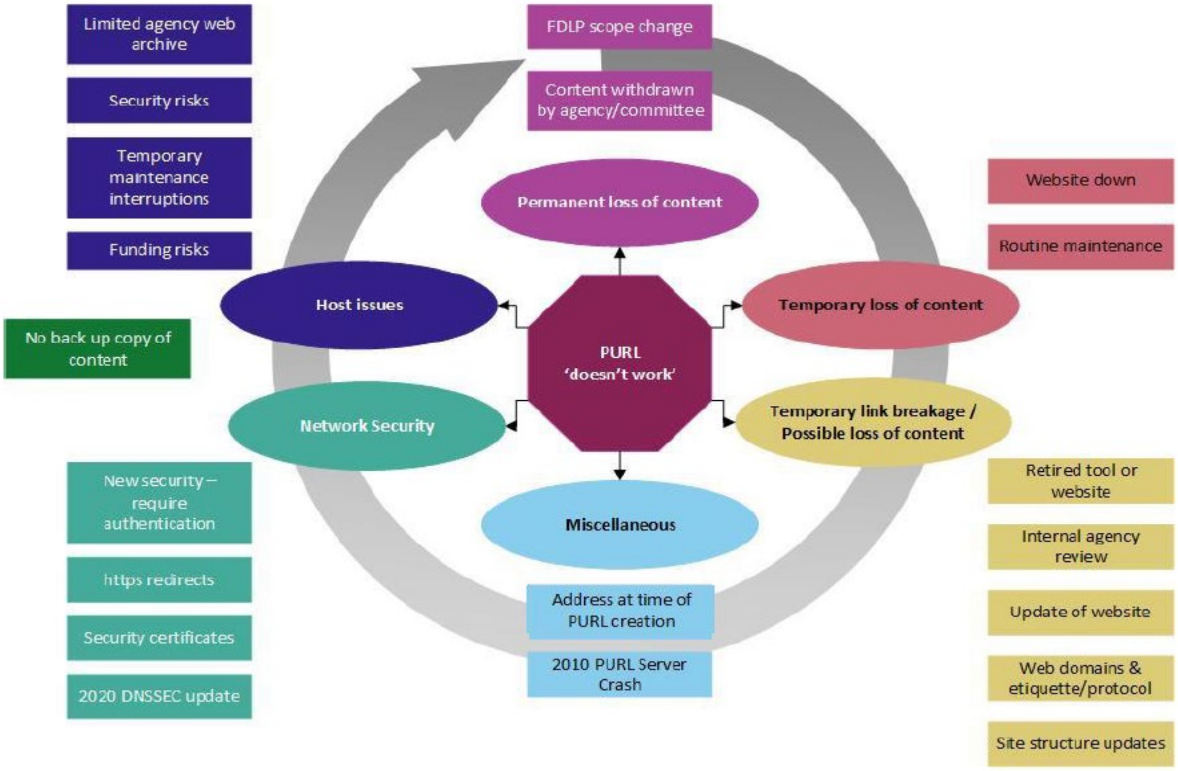
Between 2011 and 2023 (as of May 31, 2023), 320,051 PURLs were updated an average of 1.7 times. Maintaining PURLs at this level illustrates GPO's commitment to actively managing PURLs, ensuring persistent access to Federal Government information so end users will obtain their desired content.

Challenges of Link Stability — Why PURLS Might Not Resolve

The internet, networks, and digital formats have evolved over the years. The way content is presented on websites has evolved as well. Agencies routinely refresh, update, or shut down websites. This continually evolving landscape presents challenges for link stability and the permanence of digital content.

Users may conflate PURL (or any other persistent identifier) link persistence with persistent access to and permanence of the digital content they seek. From their perspective, they either access the content they need, or they do not. And when the information they seek is not retrieved, they may assume that the PURL is broken. If a PURL does not resolve, the digital object has moved or persistent access is interrupted because of any number of reasons, as shown in the illustration below, *Why PURLs May Not Resolve to the Digital Object.* This underscores the importance of the commitment to managing PIDs and ensuring persistent access.

Why PURLs May Not Resolve to the Digital Object


Limited agency web archive
Security risks
Temporary maintenance interruptions
Funding risks
No back up copy of content
FDLP scope change
Content withdrawn by agency/committee
Permanent loss of content
Website down
Routine maintenance
Host issues
PURL 'doesn't work'
Temporary loss of content
Network Security
Temporary link breakage / Possible loss of content
New security – require authentication
https redirects
Security certificates
2020 DNSSEC update
Miscellaneous
Address at time of PURL creation
2010 PURL Server Crash
Retired tool or website
Internal agency review
Update of website
Web domains & etiquette/protocol
Site structure updates

It is important to note that the illustrated scenarios of not resolving to the digital object are not unique to PURLs. Any PID schema would not resolve given similar circumstances. With all the possible variables that can interrupt or deny persistent access, serious consideration should be given to reviewing policies, procedures, and processes related to PURLs and developing and implementing a risk mitigation plan

2020 DNSSEC Update

A critical network security update in 2020 impacted both Federal agencies hosting web content and end users seeking content. This critical update involved the adoption of the Domain Name System Security Extensions (DNSSEC), by institutional Information Technology departments and Internet Service Providers. Internet Service or Network Providers that adopted the DNSSEC will not resolve links to websites that are not configured for DNSSEC.

Three retired GPO domains were affected by this security update. PURLs using the prefix of those non-compliant domains would be unable to process the redirect to content if the user's network had been updated.

To address this, GPO updated PURL links to new domains configured for DNSSEC. This necessitated updating PURLs in the *Catalog of U.S. Government Publications* (CGP) bibliographic records. Unlike the PURL updates that had been performed since 1998, this particular update also required libraries to modify their local bibliographic records to reflect the change in the PURL links. Of the more than 265,000 PURLs at that time, 65% directed to the retired GPO domains. For libraries that did not update their records, those PURLs would not work. To help libraries update their catalog records, GPO temporarily provided access to a zip file of the 170,501 changed records for download on the CGP on GitHub.

Statistical Reports —PURL Administration and Tracking FDL Usage

GPO created a PURL reporting tool that collects administrative data and reports monthly statistics, including the number of PURLs created or modified, the top 50 resources by target, and PURL referrals by host or referrer domain. This tool is also available to Federal depository libraries to track usage of online Government publications within their institutions' IP ranges.[38]

Federal depository libraries are advised to use PURLs in their online catalogs, subject guides, and web pages to ensure their links point to official, active, Federal web-based

[38] For more information see *Persistent Uniform Resource Locator (PURL): Explanation, Purpose, and Tracking Usage at Your Library*.

resources and publications. To gain a better understanding of patron use of linked resources, libraries are encouraged to set up an institutional profile for tailored usage reports . A library establishes search strings (called patterns) and identifies IP addresses, which the tool searches for in the PURL server's log files to generate a report. Extracting web traffic from the log files, and adding select catalog metadata, provides the most accurate and meaningful statistics of PURL usage in libraries.

The following elements are included in the library's PURL referral report:

TIMESTAMP	When the PURL was clicked on (reported in Eastern Time)
HOST	Name of the end-user's network host domain or the network's external IP address used to click on the PURL
PURL	Link the end-user clicked on
TARGET URL	Where the end-user is redirected
SUDOC NUMBER, TITLE, AUTHOR, YEAR	Information derived from the PURL's catalog record found in the Catalog of U.S. Government Publications
PATTERN NOTE	Any note the library saved for the pattern
REFERRER DOMAIN	URL of website or tool the end-user clicked on the PURL

V. Considering the Needs of Federal Depository Libraries

To understand how depository library coordinators and other stakeholders use PURLs and to clarify their persistent identifier needs and expectations, the PURL Working Group conducted a series of focus groups. In January 2021, an FDLP News Alert solicited volunteers to participate in the focus groups, which resulted in twenty-four volunteers. There were sixteen participants from academic (general) libraries; three from state libraries, two from special libraries; and one from law school libraries, community college libraries, and Federal agency libraries. Three focus groups were established with no more than eight volunteers in each of the groups. The same questions were asked of all three focus groups, and volunteers were provided the questions in advance. One volunteer had an unexpected conflict with the focus group time, but did submit written responses to the questions. Below is a summary of their responses.

1. **Do you experience problems with PURLs? If so, what types of problems do you have?**

- All participants mentioned having problems with PURLs not resolving.
- PURLs leading to dead end pages/broken links.
- Broken links are particularly problematic for older PURLs.
- At least one person from each of the focus groups mentioned reporting broken links through askGPO. They attempted to find a valid URL to a document.
    - Some use the URL in the cataloging record to search the Wayback Machine.
    - Others search Google or agency websites for the publications.
- Those who reported broken links agreed that they are fixed quickly.
- One person said they wished GPO used a link checker.
- One person mentioned using MARCIVE for a retrospective project to address broken links.

Other problems mentioned:

- A few mentioned duplicate cataloging records as being problematic; they are difficult to manage when records are updated.
- The original URL is not always in the MARC record.
- Maintaining technology over time can be problematic.
- PURLs resolve to different types of objects, e.g., PDFs, serials, links.

2. **How visible are PURLs in your cataloging records and in the public catalog? Is having the original URL necessary or does it cause confusion?**

The visibility of the PURL in catalog records or in the OPAC depends upon what system is used or how record views are locally configured, e.g., MARC, standard, or short views — some of which can be defined by the library. Some of these responses may be unique to the library:

- PURLs are visible in Alma.
- TRAIL records do not contain the URL.
- Records loaded into Primo do not show the URL or PURL, rather the link says “US Gov Docs”.
- It depends on when the records were loaded; newer records display the PURL and notes contain the original URL that doesn’t link.

Including the original URL in the cataloging record is viewed as very important to librarians, and not so important to users.

- Librarians use the URL to find the publication in the Internet Archive or by searching the Wayback Machine.

- General agreement among participants that average users would not use the Internet Archive or other tools to track down publications from broken PURLs.
- There was a reminder that users will encounter records from commercial vendors and from repositories such as HathiTrust that may use a PID that is not a PURL.
- Including the original URL in the record may cause confusion for users, if it's viewable.

The single versus multiple records for various formats of a publication was discussed. One focus group member reported that their library uses the single record approach by adding the PURL information from the online record to their tangible record.

One person asked, "Can GPO make sure everything is included in the Wayback Machine?" They also admitted they didn't think this was possible, but instead wishful thinking.

3. **Do you actively incorporate digital content into non-ILS applications? (ex. LibGuides, course guides, etc.)**

   - Participants in all three focus groups indicated they incorporate links to government digital content into non-ILS applications.
   - Some include PURLs in LibGuides to link to the content. This was qualified by a couple people, "to better ensure the resource is there."
     - One participant said she uses PURLs in LibGuides specifically to refer users to the official government source, rather than to resources in commercial databases.
   - Others include links in their LibGuides to catalog records, which contain PURLs.
   - One person shared that she adds government-published serials that are openly available online to the library's A-Z journal listing and links to them through PURLs.
   - One person provides links to faculty members to include in their course reserves.

4. **Are there considerations from the coordinator/librarian perspective versus the user perspective in regard to the usefulness of PURLs? If so, what are your needs as a coordinator or librarian?**

    - There was agreement that from the user perspective it is of the utmost importance that the PURL resolves to the digital publication; if it does not, users are not likely to search any farther for that particular publication.
    - From the librarian perspective, the more information in the bibliographic record the better to aid in finding the publication if the PURL is broken.
    - One person pointed out that answers to this question may have more to do with how a particular ILS/catalog presents PURLs.
    - Having the URL in the bibliographic record was viewed as valuable for teaching how to evaluate web resources; the source of the publication or website often can be determined by URLs.

5. **Is your library using a schema besides PURLs? One example would be the use of Archival Resource Keys in an institutional repository or elsewhere in the library.**

    Other PID schemas being used in depositories are Handles, DOIs, and ARKs.

    - One person mentioned they are replacing PURLs with ARKs in their repository.
    - One commented that DOIs are used more with scholarly resources.
    - ARKs are used in their library because of their capability to link to different versions.
    - Handles, DOIs, and ARKs are used more often for e-resources.
    - Handles are not being used as often as they once were at one library.
    - Patrons understand some of these other schemas fairly well.

6. **Do you want/is it important to have usage statistics of users from your catalog/website clicking on links to government documents (similar to the current PURL reporting tool)? How does your library use these statistics?**

    There was agreement that usage statistics and PURL referral reports are important to:

    - Direct cataloging priorities;
    - Show administrators and stakeholders usage of the depository collection; and
    - Register use from campus that did not start in the catalog or discovery layer, e.g., from Google, GovInfo, or WorldCat.

Several limitations of the PURL referral reports were mentioned:

- In a consortial catalog, usage by institution or individual catalog is not available;
- In a shared catalog within an institution for which the IP range is the same, usage statistics by library are unavailable;
- Usage by academic departments or in residence halls is not available; and
- A library's catalog is not necessarily the central point of access.

Desired features of the usage report and the reporting tool were expressed:

- Ability to determine usage by academic department;
- The more information that can be provided, the better;
- Ability to compare print and digital usage;
- Identify usage of individual titles, e.g., most accessed title(s);
- Ability to get reports in a time series other than monthly;
- COUNTER compliance for PURL reports, allowing comparisons with commercial platforms;
- Pursue a partnership with Internet Archive since the Wayback Machine is so important; and
- A few people didn't know about, or wanted to know more about the PURL referral reports.

Overall the participants in the focus groups found PURLs highly useful and were happy with their performance, though they had some complaints and suggestions. PURLs are convenient in that participants don't have to update their catalog records every time a URL changes, especially when the links go to agency web pages. Participants also liked that PURLs were easily placed outside of library catalogs in LibGuides and other interfaces. PURL statistics were seen as valuable by those who were aware of them.

The most common complaint about PURLs was when they were broken.[39] Trying to access the content some other way is often challenging. Many participants would like more information in the bibliographic records, so that if the PURL is broken, they can try to find the information another way. Some went a step further and talked about the usefulness of including the original URL in the bibliographic record, and that if the document was in the

[39] Most of these occur when the agency changes its top-level domain, reorganizes its website, or removes content which causes the disappearance of digital materials. In some cases, libraries are using outdated records. In 2021 libraries were advised of a one-time update to PURL links. Libraries that have not updated their records may experience difficulty resolving to content.

Internet Archive, also including a link to that copy as well as a backup in case the PURL did not redirect as expected.

Several suggestions for improvements of PURLs were provided. The main one was that records for tangible information include a PURL to the electronic version in a MARC 856 field, so users could see both formats. Other suggestions were to also include links to documents in the Internet Archive as well, and that PURL statistics reports be more granular so that libraries could determine usage by building on campus. It was also suggested that the GPO report on the statistics so that libraries could publicize the collections more

There was great interest in usage statistics and the PURL referral reports. While there were suggestions for improving the reports, there were also focus group participants who wanted to learn more about them and still others who were unaware of the reports. This underscores the need for more educational opportunities for depository staff to learn about the PURL referral reports, how to create a library's report profile, and how the data can be used to promote Government resources. It is also apparent that a session on the features of GPO's bibliographic records for digital content would be helpful as well.

It is evident from the responses to questions and discussion during the focus groups that the most important aspect to participants is that the persistent identifier resolves to the intended digital object. This statement is true, whether the PID is a PURL or any other schema.

VI. PURL Working Group Conclusions and Recommendations

Conclusions

Over the period of its charge, the PURL working group diligently explored the issues surrounding the durability of Persistent Identifiers (PIDs). Initially, the group began exploring various PID systems, then compared those systems to GPO's current implementation of PURLs, while also examining the needs of Federal depository libraries as they relate to persistent access in various discovery environments. It was first assumed that through the work of this group, a specific PID system would be recommended for GPO to implement.

All of the PID systems explored manage persistent access to digital objects in roughly similar ways. Additionally, it became apparent that there are interconnected technical, policy, administrative, and organizational issues surrounding PIDs and persistence and that a straight-forward recommendation on the adoption of a specific PID system would not be viable.

Persistence is challenging and multifaceted. Persistence involves both persistence of the object itself as well as the ongoing organizational management of the PID and PID system. As with most technologies, the policy, administrative, and organizational issues surrounding persistent identifiers are the most critical aspects of implementing a PID system and providing persistence. The real issue for the end user is the persistent access to the objects themselves. This requires coordinated persistence of the objects and the PIDs that direct to the objects. Since the internet is a constantly changing space, no matter how stable the PID system, simply pointing to a digital object on a website does not guarantee persistent access.

Recommendations

Given these conclusions, the Working Group offers the following recommendations for Council's consideration:

1. The Depository Library Council accept the Recommended Principles for Persistent Identifiers and Persistent Identifier Systems for the Government Publishing Office contained in this report, and transmit them to GPO Director Halpern.
2. GPO enact the Principles for Persistent Identifiers and Persistent Identifier Systems through the following measures:
    a. GPO should seek to maintain stable systems for persistent identifiers and redirects in use within the CGP and other systems.
        i. GPO should evaluate options for PID systems, and systems for redirects to content that may change over time, that can be migrated forward as technology evolves. This evaluation should consider how potential systems will interoperate with existing library tools used by depositories.
        ii. GPO should conduct and disseminate an analysis of benefits and drawbacks prior to proceeding with any technical changes that would require libraries to update bibliographic records in their catalogs.
    b. GPO should develop and implement strategies to mitigate risk and improve management of the current system of PURLs, and any future system(s) of persistent identifiers and redirects, setting and assessing benchmarks going forward.

   i. GPO should conduct an analysis to determine the risk to the persistence of content with links managed in the existing PURL system, factoring in the current location of the digital object and whether or not a backup is in a trusted location. This analysis should inform the development and implementation of a risk management plan, and should be repeated on a regular basis.
   ii. GPO should regularly produce a report to the community that includes data about its persistent identifiers, including their creation, update, and decommission, and note relative levels of staffing resources directed toward these activities. This report should also provide progress on achieving goals toward risk mitigation.
   iii. GPO should consider appropriate strategies for serials and integrating resources that reflect the distinction between persistent identification of the title and a PID for versioned content.
   iv. GPO should assess the needs of user communities for additional features including redirects to currently updated resources and PIDs that enable granular access for digital objects.
   v. The provenance and subsequent custody of a resource should be possible to determine from PID metadata, which should be made available from the PID system.

c. To improve persistent access to the National Collection, GPO should increase the amount of content it manages through partnerships, contracts, interagency agreements, or ingestion into GovInfo, the FDLP Web Archive, or other mechanisms under local control.
   i. GPO should strive to have content under its control as much as possible. When local control of content is not possible, GPO should strive to establish official partnerships with agencies and FDLP Libraries that insure the transfer of material in case access cannot be maintained by the partner. When these partnerships are not in place, if access is lost, GPO should strive to replace lost content from an official source and when feasible take local control of the content.
   ii. GPO should improve its transparency about current official agreements and informal arrangements that affect persistence of access to content not under GPO’s direct control, limitations on access it provides, and how it is seeking to remedy them. Including this information in the PID metadata would allow other entities to make decisions when archiving content.

        iii. For content that is not under its direct control, GPO should work with trusted digital content partners to develop methods to transparently disclose and manage related persistent identifiers.

   d. In order to increase the persistence and use of web content, particularly content that is not covered through interagency agreements, GPO should expand its FDLP Web Archive. This should include adding more websites as seed URLs to the collection so that more content is under GPO's direct control.
        i. GPO should expand quality control activities for the content captured, and seek to identify and create redirects to other Federal agency web archive collections.
        ii. GPO should investigate the potential for improving persistence of serials and integrating resources through web archiving and associated PIDs.
        iii. GPO should explore options for extraction of publications from web archiving files (WARCs) to make them more discoverable and accessible.
        iv. GPO should enhance training opportunities related to web archive collections for library staff who facilitate use of the content.

   e. GPO should explore technical solutions that will allow PID metadata to be visible and distinguished from the bibliographic metadata available in the Catalog of Government Publications.

   f. GPO should seek to work more closely with executive, legislative, judicial, and independent agencies to assure that their public information is collected, preserved, described, and made accessible for the National Collection.

3. GPO explore the potential opportunities of prospective PID systems for additional uses beyond the current implementation of GPO's PURL system, in order to improve services to FDLs.

   a. GPO should continue to make statistics on persistent identifier and redirect usage available to depository libraries.

   b. GPO should investigate current needs and future opportunities for access to serials and continuing resources that provide for persistence of access and ease of access to the most up-to-date version or content.

   c. GPO should investigate the technical feasibility of adding multiple resolution capabilities to its persistent identifier infrastructure in order to strengthen access to distributed content in a digital FDLP.

   d. GPO should consider adopting workflows amenable to distributed PID creation, allowing trusted partners to create and maintain PIDs with appropriate measures in place to maintain the integrity of the system. This includes workflows for digital deposit, unreported documents, digital preservation partnerships, and collaborative preservation systems.

4. GPO seek, as much as is possible within the Federal technology environment, to leverage interagency efficiencies in exploring technical solutions for needs related to PID system(s).

5. GPO offer training on PURLs that includes, but is not limited to, PURL referral reports, how to create a library's report profile, how the data can be used to promote Government resources, and features in GPO bibliographic records that relate to PURLs.

VII. Appendix I: Acronyms and Glossary

856 Field
The MARC bibliographic record field for electronic location and access information.

ARK: Archival Resource Key
A type of persistent identifier. ARKs are generated and maintained by the organization that creates them and can link to an object or individual pages or parts within an object. Example: https://digital.library.unt.edu/ark:/67531/metacrs8169

COUNTER: Counting Online Usage of Networked Electronic Resources
The international standard used by librarians, publishers, and other content providers for reporting usage statistics for electronic resources in a standardized way. See: https://www.projectcounter.org/

CGP: *Catalog of U.S. Government Publications*
GPO's online catalog of historical and current U.S. Federal Government publications. See: https://catalog.gpo.gov/.

Digital Access Partner
Through an official signed agreement with GPO these partners make digital resources within scope of the FDLP publicly accessible at no fee. GPO directs users to these resources via bibliographic records in the *Catalog of U.S. Government Publications* (CGP) and PURLs.

Digital Deposit
Encompasses the practices, services, and workflows for the collaborative acquisition of born-digital and digitized Federal Government information for the National Collection of U.S. Government Public Information, including deposit mechanisms: (1) from GPO to depository libraries; (2) from Federal agencies to GPO; or (3) shared by libraries with their communities and deposited with GPO.

DLC: Depository Library Council
The advisory committee to the Director of GPO. See: https://www.fdlp.gov/about/depository-library-council

DOI: Digital Object Identifier
A type of persistent identifier. Each DOI is a series of numbers and punctuation that provide a unique identification for objects of any type. Example: http://dx.doi.org/10.4403/jlis.it-5494

FAIR Principles
Principles for research data management and stewardship: Findable, Accessible, Interoperable, Reusable. See: https://www.go-fair.org/fair-principles/.

FDL: Federal Depository Library

FDLP Web Archive
The collection of selected U.S. Government web sites that LSCM harvested to provide permanent public access to Federal agency web content. Access to the harvested sites is made available through links in the *Catalog of U. S. Government Publications*, or directly from https://archive-it.org/home/FDLPwebarchive.

GPO: Government Publishing Office
See: https://www.gpo.gov.

Handle
A type of persistent identifier. The Handle system includes a central registry to resolve URLs to the location of the digital object in a distributed environment. Example: hdl:20.1000/100

Kernel Metadata
Also known as PID Kernel Information, this is the metadata within a PID system that is associated with and describes each persistent identifier.

MARC: Machine-Readable Catalog
Various record types that can be read and interpreted by a computer. The following are MARC formats within the library and information professional context: Authority Records, Bibliographic Records, Classification Records, and Holdings Records. See: https://www.loc.gov/marc/marcdocz.html.

National Collection of U.S. Government Public Information
A geographically dispersed collection of the corpus of Federal Government public information that is accessible to the public at no cost. See: https://www.fdlp.gov/about-the-fdlp/the-national-collection.

PID: Persistent Identifier
Long-lasting references to a document, file, web page, or other object over the internet. These render traditional identifiers, e.g. ISBN or URL, resolvable and retrievable and must stay the same over time, even if the name or address of the object changes.

Private PID System
PID systems that are administered locally by an institution and do not utilize a global PID system or infrastructure. Institutions that administer private PID systems both create their own PIDs and also provide the mechanism for resolving and directing their private PIDs to the identified objects.

PURL: Persistent Uniform Resource Locator
A type of persistent identifier that directs users to an address, even as that address changes over time. PURLs are not assigned by an external registration agency, and the body that assigns a PURL (such as GPO) is also responsible for maintenance to ensure it always redirects to the correct location, providing persistent access.

Resolver
A resolution service or system that converts a domain name to an IP address as part of a network.

Trusted Entity
Federal or subnational agency, public or academic institution, non-profit, or other type of organization that can capably enter into and execute an agreement such as a Memorandum of Understanding or Memorandum of Agreement with GPO.

TDR: Trustworthy Digital Repository
A trustworthy digital repository has a mission to provide reliable, long-term access to digital resources to its Designated Community, now, and into the future. To fulfill this mission, the administrators of a TDR are committed to the continuous monitoring of risks to its systems and responsibilities, ongoing strategic action and technology implementation to meet the needs of its Designated Community, and regularly ensure the transparency of its preservation and assessment activities to the public. A TDR has gone through a process or audit to certify that it meets certain criteria. GPO's GovInfo is TDR-certified under ISO16363. See: http://www.iso16363.org.

URI: Uniform Resource Identifier
A globally unique string of characters used to refer to a specific resource. URIs follow predefined syntax rules and are a superset of URNs and URLs. Examples: ISBN, web domain, IP address.

URL: Uniform Resource Locator
A unique identifier that directs users to content residing on a network via a web address.

VIII. Appendix II: Works Cited

Hakala, Juha. “Persistent Identifiers: An Overview.” *Technology Watch Report (TWR): Standards in Metadata and Interoperability* (October 13, 2010). http://www.persid.org/downloads/PI-intro-2010-09-22.pdf.

International Organization for Standardization, *Information and documentation — Principles of identification ISO/TS 22943: 2022-06*. ISO/TS 22943:2022(en), Information and documentation — Principles of identification.

Koster, Lukas. “Persistent Identifiers for Heritage Objects.” *The Code4Lib Journal*, no. 47 (February 17, 2020). https://journal.code4lib.org/articles/14978.

Schwardmann, Ulrich, Martin Fenner, Maggie Hellström, Hylke Koers, Hervé L'Hours, Brian Matthews, Raphael Ritz, Mario Valle, Mark van de Sanden, and Themis Zamani. *PID Architecture for the EOSC*. (December 2020). https://op.europa.eu/s/oTo3.

Shafer, Keith, Stuart Weibel, Erik Jul, and Jon Fausey. “Introduction to Persistent Uniform Resource Locators.” Proceedings from INet 1996. (June 27, 1996). https://web.archive.org/web/20160103053338/http://www.isoc.org/inet96/proceedings/a4/a4_1.htm.

Stone, Larry. *Handle Project: Competitive Evaluation of PURLs*. (March 22, 2000). http://web.mit.edu/handle/www/purl-eval.html.

Weigel, Tobias, Beth Plale, Mark Parsons, Gabriel Zhou, Yu Luo, Ulrich Schwardmann, Robert Quick, Margareta Hellström, and Kei Kurakawa. “RDA Recommendation on PID Kernel Information.” *Research Data Alliance* (November 19, 2019). https://www.rd-alliance.org/system/files/RDA%20Recommendation%20on%20PID%20Kernel%20Information_final.pdf.