The Quest for Persistence

An overview of the challenges associated with ensuring digital content remains accessible

Allen Moye, DePaul University-College of Law
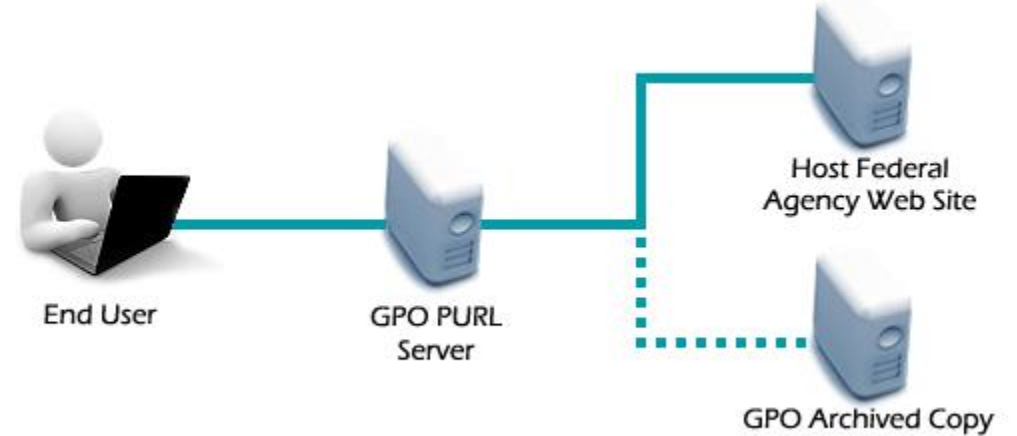
Laura Sare, Texas A&M University

Will Stringfellow, Vanderbilt University

Introduction

Since March 1998, GPO has used Persistent Uniform Resource Locators (PURLs) to provide Federal depository libraries (FDLs) and other parties stable URL access to online Federal information. Due to changes in library collection policies, advances in technology, and an increase in the amount of born-digital and digital-only Federal information, along with an increased interest in some depositories becoming mostly or all digital depositories, Council created a working group to investigate the durability of PURLs and their alternatives and to present findings and recommendations to Council for consideration.

Working Group Members

Renée Bosman, DLC Member

Allen Moye, DLC Member

Rick Mikulski, DLC Member

Laura Sare, DLC Member

Will Stringfellow, DLC Member

Alicia Kubas, Community Member*

James Jacobs, Community Member

Shari Laster, Community Member

Robbie Sittel, Community Member*

Alec Bradley, GPO

Ashley Dahlen, GPO

Cynthia Etkin, GPO Designated Federal Official

The Road so Far

- Conducted Focus Group Sessions.

- Review of the State of PURLs at GPO.

- Presentations at DLC meetings.

- Drafting PIDs Principles Document.

- Drafting Final Report to Council.

- Drafting Recommendations.

Persistent Identifiers vs. Persistent Access

- 2 Elements for long-term access to digital objects.
- Persistent Identifiers (PID)
  - A label/identifier for a single digital object.
  - Identifies and disambiguates objects.
  - Actionable.
  - Have underlying metadata for the PID, separate from the descriptive metadata of the object identified.
- Persistent Access
  - Refers to "object" being identified by a PID.
  - This is the actual .pdf, digital file, etc.

  - Must be maintained.

U.S. GOVERNMENT PUBLISHING OFFICE   GPO

What is a Principle?

- A comprehensive and fundamental law doctrine or assumption.
- A rule or belief that guides a person's actions.

- A rule or code of conduct.

Seven Principles of the Constitution

Popular Sovereignty

Republicanism

Federalism

Separation of Powers

Checks and Balances

Limited Government

Individual Rights

Principles

Five broad principles for "PURLS"

- Stability
- Validity
- Transparency
- Extensibility

- Interoperability

Principles (PIDs)

Persistent Identifiers "PIDs"

- PIDs must be unique and provide direct long-term access to a specific digital document or object that is openly accessible.
- PIDs must be unchanging and never reused.
- PIDs must enable access to a digital resource under the control of a trusted entity or an established partner.
- PIDs must enable access to the specific object described in the metadata for the resource.

Principles (PID systems)

- *Persistent Identifier Systems*

- Must be part of a system that is stable, secure, and interoperable with other systems.
- Should work regardless of the users' access starting point and the access system or delivery service used.
- Must have publicly accessible metadata.

Preliminary Findings

- The Working Group cannot recommend a specific PID schema or system; however, WG offers Recommended Principles for PIDs and PID systems.
- Persistent Identifiers and Persistent Access are two different elements both of which are necessary for long-term access.
- Major issue of persistent access is digital objects outside the control of GPO.
- PIDs may not be effective for every digital publication.

- Some PIDs have underlying metadata for the PID, which is not descriptive of the object the PID is labelling.

Preliminary Recommendations

1. The Depository Library Council should accept the [Recommended Principles for Persistent Identifiers and Persistent Identifier Systems](#) for the Government Publishing Office and transmit them to GPO Director Halpern.

2. GPO should enact the Principles for Persistent Identifiers and Persistent Identifier Systems through the following measures:

   1. Seek to maintain stable systems for identifiers in use within the CGP and other systems.
   2. Increase the percentage of content it manages that is under persistent control through ingestion into **govinfo**, web archiving, and interagency agreements.
   3. Analyze persistence of existing content and develop a strategy to increase management of both non-persistent identifiers and migrating content to persistent identifiers.
   4. Seek to add more URLs to the [FDLP Archive-It collection](#) in order to heighten the persistence of web content, particularly content that is not covered through interagency agreements.
   5. Explore technical solutions that will allow the metadata for identifier systems to be exposed and incorporated along with bibliographic metadata.
   6. Seek to work more closely with executive agencies to assure that their public information is collected, preserved, and described for the National Collection.

Preliminary Recommendations (cont.)

3. GPO should explore the broader implications on PIDs within a digital collection development system and workflow including digital deposit, unreported documents, digital preservation, and collaborative preservation systems.

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

1. Explore adopting systems already in use in the Federal technology ecosystem.
2. Seek to strengthen interagency technical cooperative programs such as CENDI.
3. Engage with GAO to determine technical requirements for future systems.

Questions???