



Office of the Superintendent of Documents

**SUPERINTENDENT OF DOCUMENTS
PUBLIC POLICY STATEMENT
2016-5**

EFFECTIVE: 12/19/2016

SUBJECT: Harvesting Digital Federal Government Information Dissemination Products for GPO's Superintendent of Documents Programs

PURPOSE

To affirm that harvesting is an appropriate proactive approach for acquiring content for the public information programs administered by the Superintendent of Documents and, further, to provide guidance for determining information dissemination products that are within scope for manual and automated harvesting from Federal agency websites.

BACKGROUND

The Government Publishing Office (GPO) has been the Government's agent for providing public access to Government information for over a century. The mandates of Chapters 17, 19, and 41 of Title 44, United States Code established GPO's responsibility for providing public access and comprehensive indexing to tangible and digital Government publications. This is the authority for the Federal Depository Library Program (FDLP), the Cataloging & Indexing Program (C&I), and for GPO's System of Online Access, public information programs within the purview of the Superintendent of Documents.

GPO defines harvested content as digital content within scope of Superintendent of Documents public information programs that is gathered from Federal agency websites. Harvesting technologies are used by GPO to discover and capture publications and websites that otherwise have not been identified by GPO staff. GPO staff use harvesting technology to capture digital content from Federal agency websites that fall within the scope of the FDLP, C&I, or GPO's System of Online Access public information programs.

The Superintendent of Documents defines Federal Agency public websites as online information resources or content and data delivery services maintained in whole or in part by the departments and agencies of the Federal Government or that are operated by a contractor, or other organization on behalf of that agency.

POLICY

GPO will acquire digital Federal Government information dissemination products for inclusion in the FDLP, C&I Program, and GPO's System of Online Access through automated and manual harvesting of official Federal public websites.

Security and industry best practices for harvesting will be employed. When online content cannot be harvested from official websites, GPO will pursue content partnerships with the originating agency.

Any automated tools or services GPO uses for harvesting will be approved by GPO's Office of the Chief Information Officer.

SCOPE

This policy pertains to all Federal Government information dissemination products and services subject to the jurisdiction of the public information programs of the Superintendent of Documents. However, the following limitations apply:

- GPO harvests from official Federal Government websites, or from the website of an agency's partner(s) that contain official Government information dissemination products.
- GPO will seek to enter into a content partnership with an agency when GPO cannot harvest content from the agency's website. A partnership can establish a joint commitment between GPO and the agency to preserve permanent public access to content within scope of the FDLP, C&I, or GPO's System of Online Access. Content partnerships will be developed following the guidelines set forth in the Library Services and Content Management partnerships policy.

See related Superintendent of Documents Public Policy Statements:

- Scope of Government Information Products Included in the Cataloging and Indexing Program and Disseminated Through the Federal Depository Library Program ([SOD-PPS-2016-1](#)).
- Content Scope for GPO's System of Online Access ([SOD-PPS-2016-2](#)).

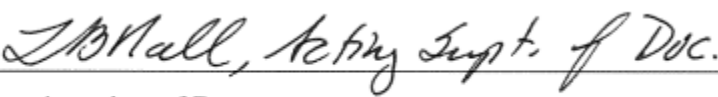
APPLICATION

This policy applies to all appropriate elements of Library Services and Content Management (LSCM). The Superintendent of Documents, through the LSCM Managing Director, must authorize any exceptions to this policy. Exceptions will be documented in writing to the LSCM Web Archiving Working Group.

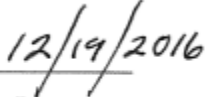
DEFINITIONS

Automated Harvesting	The use of technology tools and services that extract information from websites in order to preserve their original structure and functionality.
Harvest	The identification and replication of content resident on Web servers outside GPO's control.
Information Dissemination Products	Any book, paper, map, machine-readable material, audiovisual production, or other documentary material, regardless of physical form or characteristic, disseminated by an agency, or its partner, to the public. ¹
Manual Harvesting	The identification and downloading of digital publications, saving all associated publication files.
Permanent Public Access	Government information dissemination products within the scope of the public information and dissemination programs of the Superintendent of Documents which are required to remain available for continuous free public access (44 U.S.C. §1911).

APPROVED:



Superintendent of Documents



Date

¹ The Superintendent of Documents adopted the use of this definition, with minor edits, which is used by the Office of Management and Budget and Executive Branch agencies. It is found in OMB Circular A-130.