UNITED STATES GOVERNMENT PRINTING OFFICE (GPO)

CONCEPT OF OPERATIONS
(CONOPS V2.0)

FOR THE

FUTURE DIGITAL SYSTEM (FDsys)

Final
May 16, 2005

FINAL

Document Change Control Sheet

Document Title: Concept of Operations (ConOps)

Date	Filename/version #	Author	Revision Description
10/01/04	Future Digital System.DC.COP.1.0.DOC	Future Digital System Phase 2 Team	Baseline Concept of Operations
05/16/05	FDsys ConOp v2.0	Future Digital System Phase 3 Team	Revised Concept of Operations Reflecting New Information from Preliminary Requirements

FINAL

PREFACE

The proposed Future Digital System will ingest, preserve, and provide access to the information produced by the U.S. Government--including information produced by all three branches of Government --and to the material currently in the custody of the Government Printing Office (GPO). The proposed system is envisioned as a comprehensive, systematic, and dynamic means for preserving any kind of content, independent of specific hardware and/or software. The system will enable GPO customers to obtain hard copy publications and to electronically access the content they want, and it will enable GPO to deliver that content in the formats its customer's desire. The system should automate many of the content lifecycle processes and make it easier to deliver the content in formats suited to the needs of GPO customers. The proposed system is required to be policy neutral so that it can support not only GPO's current operational policies but also changes that are expected to emerge.

This Concept of Operations document provides a conceptual overview of the proposed Future Digital System. It is a living document that will be coordinated in a collaborative manner with industry, public, and Government stakeholders to ensure the viability of the concepts it presents.

FINAL

TABLE OF CONTENTS

FINAL

TABLE OF CONTENTS

FINAL

TABLE OF CONTENTS

FINAL

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

FINAL

CONCEPT OF OPERATIONS (CONOPS)

1 SCOPE

This document describes the desired characteristics of the Future Digital System from the user's viewpoint. The sections below identify the proposed system, provide a document overview and describe the approach used to generate the document, and provide a brief overview of the system.

1.1 IDENTIFICATION

The proposed Future Digital System will encompass all of the associated equipment, facilities, material, software, hardware, policy and technical documentation, services, and personnel required for its operations and support at the United States Government Printing Office (GPO).

1.2 DOCUMENT OVERVIEW

The *Future Digital System ConOps* serves as a vehicle to communicate the high-level system characteristics of the envisioned system to the user, buyer, developer, and other stakeholders. The ideas expressed herein have resulted from a thorough analysis of the challenges involved in providing content, which have been addressed through the use of the OAIS model. The *Future Digital System ConOps* will be used to generate high level system specification that will be captured in a *Future Digital System Requirements Document (RD)* that is yet to be developed. The *Future Digital System ConOps* should be reviewed together with the *Future Digital System RD* upon its completion, as the *Future Digital System RD* may contain additional information that has not been presented herein.

This document contains the following sections;

- **Section 1**, Scope, describes the approach used to develop the *Future Digital System ConOps*.
- **Section 2**, References, lists the reference documentation that was used as a basis to create the document.
- **Section 3**, Current GPO Situation, describes the current GPO systems.
- **Section 4**, Justification For and Nature of Changes, discusses the justification for and the nature of changes based on the most current information available.
- **Section 5**, Concepts for the Proposed System, discusses proposed system concepts.
- **Section 6**, Operational Scenarios, describes various operational scenarios.
- **Section 7**, Summary of Impacts, summarizes operational, organizational, and other impacts that could be expected to occur during development.
- **Section 8**, Analysis of the Proposed System, analyzes the proposed system.
- **Section 9**, Notes, provides additional information, including an acronym list and glossary.

1.2.1 Approach. This document was developed using the concept analysis approach, which is a process of analyzing a problem domain and an operational environment for the purpose of specifying the characteristics of a proposed system from the user's perspective. This approach is useful for clarifying, resolving, and reconciling vague and conflicting needs, wants, and opinions that often conflict. It also minimizes the potential for designing a system in which each individual function meets its specifications but that fails to meet the users' needs as a whole, was minimized.

Information was obtained from analyses of the "As-Is," or legacy systems to identify the shortcomings of these systems. Through this analysis, the capabilities required of a "To-Be" system were captured and subsequently characterized in a number of collaborative concept papers, from which a vision for the future GPO system – the Future Digital System – emerged. The information gathered in this process enabled the refinement of user classes and the creation of operational scenarios that document the actions of the to-be system users.

In order to devise an integrated approach to development of the Future Digital System, it then became necessary to reexamine GPO's current business processes used to confront the challenges of collecting, gathering, and disseminating information, as well as its approach to preserving this content, including the lifecycle management of that content. This effort must continue into the future in order to realize the vision of GPO and to ensure the right of the citizenry to access Federal Government information.

1.2.2 IEEE Standard. The *Future Digital System ConOps* document was generated using guidance provided by the IEEE Std. 1362-1998, *IEEE Guide for Information Technology-System Definition-Concept of Operations (ConOps) Document*.

1.3 SYSTEM OVERVIEW

The Future Digital System will be composed of the necessary technology and business practices that will enable GPO to ingest, manage, preserve, and provide access to content that is disseminated in hard copy and to content that is electronically stored.

GPO believes that management of both electronic and non-electronic content should be an integrated process that provides maximum efficiency and value for users. GPO has taken a lifecycle management approach to this data that promotes more effective and efficient processes by sharing relevant data and that promotes seamless transition from one phase to another. The proposed system should support GPO's end-to-end lifecycle management processes, including processes for the creation of content and for the transfer, ingest, management, and access of all electronic and non-electronic content.

The Future Digital System will ingest, preserve, and provide access to the information produced by the U.S. Government--including information produced by all three branches of Government --and to the material currently in the custody of GPO and Federal depository libraries. The proposed system is envisioned as a comprehensive, systematic, and dynamic means for preserving any kind of content independently of specific hardware and/or software. When it becomes operational, The Future Digital System will enable GPO customers to obtain hard copy publications and to electronically access and retrieve the content they want, and it will enable GPO to deliver that content in the formats its customers desire. The system should automate many of the content lifecycle processes and make it easier to deliver the content in formats suited to the needs of GPO customers.

The Future Digital System tools and content must be accessible. Accessibility is making tools and content available and usable for all users including those with disabilities. The Future Digital System will follow established best practices and regulations for accessibility (e.g. Section 508, W3C, etc.)

2 REFERENCES

The standards, guidelines, and GPO and the Future Digital System documentation used to support the *Future Digital System ConOps* are described in the following sections.

2.1 STANDARDS AND GUIDELINES

- American Library Association, *et. al. Anglo-American cataloging rules, second edition.* Chicago: American Library Association, 1978.

- Digital Library Federation Benchmark Working Group 2001-2002. *Benchmark for faithful digital reproductions of monographs and serials version 1, December 2002.* Washington, D.C.: Digital Library Federation, 2002. Available at http://www.diglib.org/standards/bmarkfin.htm

- Digital Library Forum. *A framework of guidance for building good digital collections.* Washington, D.C.: Institute for Museum and Library Services, 2001. Available at http://www.imls.gov/pubs/forumframework.htm

- *IEEE guide for information technology-system definition-concept of operations (ConOps) document. IEEE Std. 1362-1998* New York: Institute of Electrical and Electronics Engineers, 1998.

- Interagency Committee on Government Information. *Recommended policies and guidelines for federal public websites*. Washington, D.C.: Office of Management and Budget, 2004. Available at http:www.cio.gov/documents/ICGI/ICGI-June9report.pdf

- Interagency Committee on Government Information. *Requirements for enabling the identification, categorization and consistent retrieval of government information*. Washington D. C.: Office of Management and Budget, 2004. Available at http:www.cio.gov/documents/ICGI/ICGI-June9report.pdf

- Koyanl, Sanjay J., Robert W. Bailey, Janice R. Nall, Susan Allison, et al. *Research-based web design & usability guidelines*. Washington, D.C. [?]: U.S. Department of Health and Human Services, 2003. Available at http://usability.gov/pdfs/guidelines.html

- United States Government Accountability Office. *Government Printing Office: Actions to strengthen and sustain GPO's transformation.* GAO-04-830 Washington: U.S. General Accounting Office, 2004

- W3C. *Web content accessibility guidelines 1.0* Cambridge, MA [?]: W3C, 1999. Available at http://www.w3.org/TR/WCAG10/

- *The large-scale archival storage of digital objects*, Jim Linden, et al, The British Library, Digital Preservation Coalition Technology Watch Series Report 04-03, February 2005, http://www.dpconline.org/docs/dpctw04-03.pdf

- *Strategic Vision for the 21st Century*, U.S. Government Printing Office, December 1, 2004, http://www.main.gpo.gov/pub_print/STRATEGICPLAN.html

2.2 GPO DOCUMENTATION

- *Baseline Requirements for Digital Reformatting and Delivery of Legacy Federal Documents Collections,* prepared by the Center for Research Libraries, 11/29/04
- *Contract terms, quality assurance through attributes (QATAP).* GPO Publication 310.1, 2002
- *Federal document repositories: decision framework by tangible repository type,* prepared by the Center for Research Libraries, September 18, 2004*
- *GPO agency procedural handbook.* GPO Publication 305.1, 1998
- *Guide to Federal printing and publishing: what every Federal publisher should know about the publishing process.*, [2002?]
- *The Guidelines: best practices for preparing & submitting electronic design & prepress files* GPO Publication 300.6, 2001
- *Managing the FDLP Electronic Collection, Second Edition* , June 18, 2004 *
- *The National Bibliography of U.S. Government Publications: Initial Planning Statement,* June 18, 2004*
- *National Collection of U.S. Government Publications* - Revised June 18, 2004*
- *Printing Procurement Regulations.* GPO Publication 305.3, 1999
- *Report from the Meeting of Experts on Digital Preservation,* March 12, 2004*
- *Report from the Meeting of Experts on Digital Preservation: Metadata Specifications*, June 14, 2004*
- *Style manual.* Washington: U.S. Government Printing Office, 2000
- *U.S. Government Online Bookstore Replacement Proposal, January 2003* [unpublished]
- *U.S. Government Printing Office PKI Business Plan, October 28, 2003.* [unpublished]

*Reports available from http://www.gpoaccess.gov/about/reports/index.html

2.3 Laws and Regulations

- "Access to Federal Electronic Information" Title 44 *U.S. Code,* Chapter 41, 2000 edition
- "Depository Library Program" Title 44 *U.S. Code*, Chapter 19, 2000 edition
- "Distribution and Sale of Public Documents" Title 44 *U.S. Code,* Chapter 17 2000 edition
- "Production and Procurement of Printing and Binding" Title 44 *U.S. Code,* Chapter 5, 2000 edition

- "Vocational Rehabilitation and Other Rehabilitation Services--Rights and Advocacy" Title 29 *U.S. Code* Chapter 16, Subchapter V, 2000 edition

2.4 SYSTEM DOCUMENTATION

Numerous working papers were generated from the working group composed of GPO and Governmental agency representatives. This pool of information is reflected in this *Future Digital System ConOps* document.

3 CURRENT GPO SITUATION

This section describes the background, objectives, and scope of the current situation at GPO and the systems in use; operational policies or constraints; modes of operation; user classes and other involved personnel; and the support environment.

3.1 BACKGROUND, OBJECTIVES, AND SCOPE OF THE CURRENT SITUATION

GPO systems currently in place have been developed primarily to support dissemination of printed publications. In general, GPO dedicates its in-house printing equipment to congressional printing and contracts out most printing for the executive branch. Contracting is accomplished through an acquisition program that relies on the commercial sector and passes the contractors' costs on to its Government customers. Pre-qualified businesses, small to large in size, compete for printing jobs that GPO printing experts oversee to ensure that the contractors meet all customer requirements for quality.

In addition, GPO provides a range of services to Government agencies, including, for example, CD-ROM development and production, archiving/storage, conversion of products to electronic format, Web hosting, and Web page design and development. However, GPO's capability to provide these products and services is not well known among its Government agency customers. As a result, the agencies are developing these products and services on their own, which further complicates the issue of "fugitive documents."

In the past, GPO's implementation of new products and services has been conducted in an ad-hoc manner, which has resulted in the development of disparate systems. Additionally, GPO has failed to update its technological abilities to keep pace with changes in the information dissemination environment, and as a result it must update its technology to address the needs of today's customers and information users and stay alert to future trends and changing needs.

3.1.1 Analysis of the Current Systems. GPO's current environment consists of legacy systems and manual operations. These are incapable of providing the breadth and depth of functionality that the proposed system will provide. Currently no single system or group of systems exists that will provide the capabilities envisioned for the proposed system.

3.1.2 Motivation for a New System. Many factors taken together comprise the motivation for developing a system that will adequately preserve content for as long as that content is needed while also providing the citizenry and Federal agencies access to it. These factors include the introduction of new technologies that support new and different types of content with enhanced formats; a lack of automated capability, which imposes narrow limits on access to that content; and the fact that numerous Federal agencies, without GPO's knowledge, independently contract with commercial entities to produce their content and in some cases host it on the World Wide Web. The only viable solution that will enable GPO to meet its vision and mission is to design, develop, and implement a new system.

Current printing industry trends show that the total volume of printed material has been declining for the past few years, which is primarily due to the use of electronic media options, and this trend is expected to continue. The move to electronic dissemination is the latest phase in the electronic publishing revolution that has transformed the printing industry in recent decades. This revolution was driven by the development of increasingly sophisticated electronic publishing ("desktop publishing") software, run on personal computers, that allows users to design documents, including both images and text; and by the parallel development of electronic laser printer/copier technology with capabilities that approach those of high-end presses. These tools allow users to produce documents that formerly would have required professional printing expertise and large printing systems.

These technologies have brought major economic and industrial changes to the printing industry. As electronic publishing software becomes increasingly sophisticated, user-friendly, and reliable, it approaches an ideal situation, at least from the perspective of the print customer, who can now develop software files that can be reproduced with little or no intervention by printing professionals. As the printing process is simplified, the customer can take responsibility for more of the work. Thus, the technologies diminish the value that printing organizations such as GPO add to the printing process, particularly for simpler printing jobs. Nonetheless, professional expertise remains critical for many aspects of printing, and for many print jobs it is still not possible to bypass the printing professional altogether.

The advent of the Internet permits the instantaneous distribution of the electronic documents produced by the new publishing processes, breaking the link between printing and dissemination. With the increasing use of the Web, the electronic dissemination of information not only becomes practical, it also becomes more economical than dissemination on paper.

As a result, many organizations are changing from a print to an electronic focus. In the early stages of the electronic publishing revolution, organizations tended to prepare a document for printing and then convert the print layout to electronic form—in other words, focusing on printing rather than dissemination. Increasingly, however, organizations are changing their focus to providing information—not necessarily on paper.

Today an organization may employ computers to generate both plates used for printing as well as electronic files for dissemination. Tomorrow, the organization may create only an electronic representation of the information, which can be disseminated through various media, such as Web sites. A printed version would be produced only upon request.

As in private industry, printing and dissemination in the Federal Government are significantly affected by a changing technological environment which presents both financial and management challenges to GPO. Just as the volume of material provided to private firms for printing has decreased over the past few years, so has the volume of material that Federal agencies provide to GPO. In addition, Federal agencies are publishing more items directly to the Web—without creating paper documents at all—and they are able to print and disseminate information without using GPO services. Similarly, individuals are downloading documents from Government Web sites, such as *GPO Access*, rather than purchasing paper copies of Government documents, thus reducing document sales.

These changes in Federal printing and dissemination are also creating challenges for GPO's long-standing structure for centralized printing and dissemination. Agencies are currently required to notify GPO of published documents (if they used other printing sources), which allows GPO to review them to determine whether they should be disseminated to the depository libraries. If so, GPO can then add a rider to the agency's print contract to obtain the number of copies that it needs for dissemination. However, if agencies do not notify GPO of their intent to print, these documents become "fugitive documents" and may not be available to the public through the depository library program.

These current GPO technical systems and strategies for the transfer, ingest, preservation, and sustained access of content are inadequate and inefficient. GPO can expect to digitize volumes of legacy data and continue to receive content in a variety of formats that GPO must be able to ingest, manage (including preservation), disseminate, and provide access to in the future.

In order to meet GPO's strategic goals, the Future Digital System should be able to accomplish the following goals:

- Support GPO's content, content management, and content delivery processes and continuing improvements with the efficiency, quality, effectiveness, and timeliness required by those processes;
- Provide access to descriptions of all types of content preserved by GPO;
- Accept/ingest content in a variety of complex formats;
- Accommodate future digital formats;
- Ensure the authenticity of the content that GPO preserves;
- Provide access to the content; and
- Support flexible services for content that GPO will manage on behalf of other Federal agencies.

3.1.3 Modes of Operation of the Current System or Situation. Modes of operation for the current legacy systems are provided in Section 3.4, Modes of Operation.

3.1.4 Classes of Users. Classes of users for the current legacy systems are identified in Section 3.5, User Classes and Other Involved Personnel.

3.2 OPERATIONAL POLICIES AND CONSTRAINTS

There are no other constraints beyond those mentioned in previous sections and in Section 4, Justification for and Nature of Changes. Limitations of a number of the systems that comprise the current environment are discussed in the subsections of Section 3.3, Description of the Current Environment.

3.3 DESCRIPTION OF THE CURRENT ENVIRONMENT

The detailed Current-State documentation is based on a simple reference model, as depicted in **Figure 3-1, Simple Reference Model**. In addition to supporting the description of the current state of GPO, this model will also be used to develop the functional requirements for the Proposed System.

Figure 3-1. Simple Reference Model

GPO's mission is to make the content of the Government publicly available – forever. To do this, GPO assists Federal agencies with their printing and publishing needs and subsequently ingests, manages, preserves, and delivers this content in order to meet its mission. For GPO and its customers, the printing and publishing process begins with agency customers submitting orders or information to GPO.

Orders requiring GPO services are submitted to GPO's Customer Services area, specifically through GPO's Departmental Account Representative (DARD) or Regional Printing Procurement Office (RPPO).

Information-only content is submitted to, or harvested by, the Superintendent of Documents (SuDocs) through Information Dissemination acquisitions and discovery specialists. The specialists who make up these three groups are the primary points of contact between Federal agencies and GPO.

- DARD and RPPO staffs have expertise in GPO products, services, and applicable technologies. They can provide reports on work in progress or technical advice on such matters as commercial term contracts and electronic photocomposition. These specialists can determine the appropriate method of production (e.g., in-plant or procured printing).

- ID staff ensures that electronic-only information products are included in the various GPO dissemination programs. These programs include the FDLP and Sales Program. Federal agencies can also take advantage of GPO's experience and expertise in electronic information production and dissemination through *GPO Access*.

For more detailed information please refer to *GPO's Guide to Federal Printing and Publishing* and *GPO Agency Procedural Handbook*.

A description of each of the independent systems is provided in Section 3.3.1. These systems alone or grouped together are incapable of addressing or satisfying the mission needs of GPO. The information in the subsections below is provided for completeness; however, to reiterate, based on the conclusions of the analyses that have been performed on the current systems, they do not meet the operational needs that are required to achieve GPO mission goals and objectives. As such, information regarding the following components of the current systems has not been provided.

- Cost of system operations
- Operational risk factors
- Performance characteristics
- Quality attributes
- Provisions for safety, security, privacy, integrity, and continuity of operations in emergencies

3.3.1 Descriptions and Capabilities of Existing Systems. The following subsections provide high-level functional description of the existing situation. For a detailed explanation of the descriptions and capabilities of the existing system, please refer to Appendix A.

3.3.1.1 Content Ingest. In the context of GPO's function and authority, content is the informational matter received through a chain of responsibility and authority from Federal agencies that GPO processes to the agency's specifications into publications in various formats, with the intention of disseminating it to the public; or informational matter that is discovered and harvested from the Web for preservation of access. A content object is formed by combining content with Representation Information (RI) that maps or organizes the data into a useful, meaningful presentation. Thus a page of a Congressional bill is composed of data (in this case text) and the RI that arranges that text into columns with line numbers on the printed page or screen. In the case of a printed bill, the RI arranges the content meaningfully on the page and is fixed in time by the printing process, after which the RI is discarded. The bill presented online depends on its RI being preserved alongside the content to generate a meaningful rendering each time it is invoked.

GPO receives content from agencies in a variety of formats, intended for a variety of output products. This diversity influences the treatment content receives in the course of processing by GPO.

Content may be received by GPO in digital form, structured for hard- or soft-copy output, or in analog form from which digital files for printing are created by GPO staff. Digital inputs may range from structured files intended for producing hardcopy output or web presentations, to minimally structured ASCII text to

be loaded into a searchable database. GPO's practice has been to accommodate agency requirements and processes by not limiting input forms.

Content in the current environment is managed on two largely unconnected levels. In the short term, content management consists of the management of jobs in the production process, the organizing principle of which is GPO jacket, literally a jacket or folder onto or into which specifications and other information are recorded. Most of this metadata is acquired or created by GPO and is of limited value, with a short lifespan, ending with the delivery of the completed work or service.

In the longer term, management of content is focused on distribution of publications through several channels: sale, deposit of publications in libraries, and delivery of online publications through *GPO Access* alongside searching and locator services. Metadata supporting these longer-term activities is created by GPO and maintained and augmented over time.

3.3.1.2 Content Management. GPO's current Content Management System (CMS) consists of four independent areas (Customer Services, In Plant Production Services, Information Dissemination, and Information Technology). Once information is processed from the customer, it is subsequently stored and accessible from multiple areas within GPO.

Customer Services and In Plant Production Services work in tandem via a mainframe system to process customer orders and deliver products, as specified, in tangible or intangible formats. A limited portion of these orders are transferred to Information Technology (IT) or Information Dissemination (ID) for additional processing and archiving.

IT maintains GPO's servers and Web services (e.g., *GPO Access* and agency Web hosting) in a non-automated fashion.

GPO has investigated possible enterprise-wide CMSs, recognizing the need for an enterprise-wide solution instead of the current situation of multiple individual systems that do not integrate.

Information needs to be located and accessed by the public. Historically, GPO has created descriptive metadata files in the form of cataloging and bibliographic information. However, due to fugitive documents as well as limited financial and human resources, the comprehensive nature of this program has eroded into cataloging the documents in the FDLP. This descriptive information is integrated into reference and search tools such as The Monthly Catalog of Government Publications (MOCAT), a suite of tools called Locator Tools and Services, and the *Catalog of U.S. Government Publications* (CGP) application on *GPO Access*.

GPO gets information for reports from a database called the Procurement Information and Control System (PICS). Information is manually keyed into PICS from forms or transferred from other software.

Once the public locates a publication they assume that it is official and authentic. GPO has begun implementing a Public Key Infrastructure (PKI) in order to make sure documents are not changed without GPO knowing about it. Unfortunately, the public also expects to have the most recent version of a document, which might not be possible since GPO requires no formal verification for documents that are processed. The approval process is one way to establish the version of a document, but a better way might be to use a CMS. However, GPO does not currently have the financial resources to implement a CMS.

3.3.1.3 Content Delivery. Under legal authority of Title 44, Chapters 17, 19, and 41 of the United States Code, GPO's Office of Information Dissemination (Superintendent of Documents) administers various dissemination programs with the mission of providing permanent public access to official Federal Government information. These include the Federal Depository Library Program (FDLP), GPO Sales

Program, and *GPO Access* public Web site. The FDLP distributes electronic and tangible publications to a network of Federal Depository libraries across the country. Select publications are also available for sale to the public via GPO Sales Program, including the U.S. Government Online Bookstore. *GPO Access,* the primary vehicle for the dissemination of electronic publications via the FDLP, provides public access to full-text databases of official Federal publications at no fee.

Due to decreased demand and high expenses, the Sales Program has not been able to recover costs during the last several fiscal years. Additional analysis of inventory data has shown the Sales Program is only selling 35% of the inventory ordered. As a result, GPO is developing a plan to reduce costs by making a substantial portion of its products available through Print on Demand (POD). When a request is made for a printed copy of a document, an SF1 is manually generated and a jacket number is created for billing purposes. Production then locates the file on the repository and outputs to a digital press.

The ability to produce a publication on demand via digital printing hinges upon the availability of a press-optimized digital file for that publication. GPO's POD Committee (including representatives of ID, Customer Services, and Plant Operations), is actively obtaining press optimized (Portable Document Format) PDF files through various methods. If a press optimized PDF file does not exist for a given publication, hard copies of that publication are obtained and scanned. Due to budget constraints of the project (the concept being to spend as little as possible digitizing titles that do not sell in high-volumes in the first place), approval has not been granted to procure a formal CMS to manage these files. As an interim solution until a CMS can be implemented, the files are currently being stored in a content repository, which grants authorized users full access to the files, allowing them to create, change, move, and delete all files on the server.

Agency customers regularly request conventional printed hard copy jobs through GPO. These customers currently have two options: Printing Procurement and Production. If they do not specify which is preferred, GPO chooses the procurement method. In all cases the agency must complete and submit an order form. The majority of forms must be submitted manually, with the exceptions of the pilot program GPODirect (*http://gpodirect.gpo.gov*, formerly known as the OMB Compact) and the Web-enabled mainframe application PICSWEB (*http://govprint.access.gpo.gov*), which allows registered users to electronically submit Direct Deal term contract order forms (2511). No other online processes for submitting printing requirements are currently in operation.

Customers submitting order forms to Printing Procurement have multiple contracting vehicles available to them. The Simplified Purchase Agreement (SPA) and AgencyDirect both allow agencies to procure their own printing. If an agency chooses to go through GPO's procurement process, a specification is written and bids are solicited. The lowest responsible, responsive bidder as determined by the Contracting Officer within the guidelines of the Printing Procurement Regulation (PPR) is awarded the job. Term contracts may be written if the requirement will be repeated throughout a specific time period (usually a year). Direct Deal contracts are managed by the agency, with GPO handling only administrative duties. GPO-placed term contracts are available to all agencies, with all contracting and administration handled through GPO. Any work submitted through Printing Procurement will be completed by commercial printing vendors.

Many agency customers are also requesting from GPO, and GPO's affiliated contractors, digital files that they can place on line for viewing and/or download. Agencies requiring digital files may go through Production or Printing Procurement. Both will accept hard copy and digital input for digital file creation. These files may replace the printed documents. However, the vast majority are created as supplements to the printed pieces. The current system is very printing-centric. Digital layout files are usually created in professional design programs following guidelines for a printed end result (CMYK color, high resolution images, etc.). Once the printed product is completed the files are repurposed for screen usage. Few files are currently created for strictly digital output.

Production has the capability of creating digital files from both hard copy and digital input. If hard copy is provided, Production can scan into a digital format. Optical Character Recognition (OCR) scanning and subsequent search features are available upon request by the agency. While some hard copy is sent to production for scanning, the current system is inadequate to handle a large influx of agency documents. Personnel, training, and equipment issues must be addressed if this area of revenue is to grow.

Typesetting from hard copy into a digital format is also available. Once a digital file has been created from the scanned hard copy or the digital input, additional features such as bookmarks, links, and indexing may be added to certain digital file types (PDF, etc.). A choice of file formats including PDF, JPG, ASCII, etc., may be provided back to the customer, with PDF being the most commonly requested format. Fill-able PDF files for online use may be created from both hard copy and existing digital files. The agency may also choose from a variety of media, with CD-ROM being the most widely used.

When a digital file is requested in addition to the printed product, the file type and media to be returned to the customer are identified on the order form, along with any additional requirements. Once the printed publication is completed, the specified digital file should be created from the production files and should be an exact representation of the printed product (format, structure, etc.). The file will be supplied to the agency in the format and on the media requested on the order form. This process should be followed by GPO's in-house production facilities and outside contractors.

Many customers also request Web page capabilities in the form of digital files, typically HTML. Printing Procurement can request digital files such as HTML from outside vendors. Various areas of Production can also produce digital files and, in addition, can create and/or maintain Web sites for agency customers.

Agencies requiring digital file creation may submit an order form to Printing Procurement for bidding by contractors. The written specifications include information on the hard copy or file type and media submitted to the contractor, as well as the file type and media to be returned to the customer. Additional requirements for deliverables include searching capabilities, metadata creation, etc. Additional requirements for digital files with Web page capabilities (such as HTML) include coding, version compatibility, etc. Repurposed Deliverables Specification Language is available for inclusion into specifications. This language includes specific requirements for PDF and HTML deliverables. Once specifications are complete, bids are solicited and accepted in the same way they would be for a hard copy job described above.

In addition to customer agencies, GPO's ID section also utilizes digital files for soft copy display. Files submitted to ID (from both In Plant Production and outside contractors) are evaluated to determine if they fall within the scope of ID programs. When a file is determined to be within scope, it is further processed for display on *GPO Access*. Additional derivatives may be created for different purposes from the existing digital file. Examples of processing options include WAIS database indexing and optimizing for placement on a web server. Additional requirements such as OCR scanning, bookmarks, breaking large files into smaller more easily downloaded components, etc., are available to the ID department.

The primary digital media type currently used at GPO is the compact disk (CD). CDs may be formatted for both Macintosh and Windows computers. Virtually all computer systems are able to read CDs, making them the most widely accepted form of hard media available today. GPO accepts, outputs, duplicates, and replicates CDs.

GPO is working on expanding its in-house Digital Video Disk (DVD) capabilities. Some areas of Production currently have the ability to read and/or write DVDs while others do not. The section that handles CD replicating has DVD equipment in place and is conducing testing on DVD replication. Until testing is completed, a contract with an outside vendor is in place to handle DVD replication work from customers.

Many outdated types of media, such as floppy disks, ZIP drives, etc., are no longer typical at GPO. However, customers with older media may be able to submit their digital files to GPO's Production section for hard copy output and in addition have the files returned to them on a CD.

Jobs that do not go through GPO's in-house facilities are submitted to Printing Procurement. Specifications can be written for any type of digital media provided by the agency customer and will be procured from an outside contractor. While the most common form of media accepted by contractors is the CD, some contractors have the capability of utilizing other media.

3.4 MODES OF OPERATION

The current modes of operation provide limited automation of processing and preserving content requiring human intervention at many steps in the process. GPO's current systems environment is unable to meet the operational needs of GPO; therefore, this section has been tailored out.

3.5 USER CLASSES AND OTHER INVOLVED PERSONNEL

3.5.1 Organizational Structure. Reference GPO Notice 105-123, organizational structure.

3.5.2 Profiles of User Classes. User classes are characterized in the subsections 3.3.1.1, 3.3.1.2, and 3.3.1.3. For a detailed explanation of the descriptions, please refer to Appendix A.

3.5.3 Interactions Among User Classes. Interactions among user classes are characterized in the subsections 3.3.1.1, 3.3.1.2, and 3.3.1.3.

3.5.4 Other Involved Personnel. Other involved personnel are referenced in the subsections 3.3.1.1, 3.3.1.2, and 3.3.1.3.

3.6 SUPPORT ENVIRONMENT

The limited number of existing systems is individually supported by in-house Government support from developers of the systems. Each system has its own unique support environment. Current systems do not meet GPO needs and it would be cost prohibitive to upgrade or modify. For these reasons, the following information is omitted from this document:

- Identification of the support concepts; and
- Identification of the support environment for the current system, including:
  - Support agency or agencies,
  - Facilities,
  - Equipment,
  - Support software,
  - Repair or replacement criteria,
  - Maintenance levels and cycles,
  - Storage,
  - Distribution, and
  - Supply Levels.

4 JUSTIFICATION FOR AND NATURE OF CHANGES

The following subsections describe the proposed system in terms of justification for changes, the nature of the desired changes, priorities among changes, changes considered but not included, and assumptions and constraints associated with building the Future Digital System.

4.1 JUSTIFICATION FOR CHANGES

GPO systems have been developed primarily to support the dissemination of printed publications. The majority of the printing is currently performed for the United States Congress, while most of the executive branch printing is contracted out. Through the use of an acquisition program, contracts are passed onto the commercial sector, with costs being passed to Government customers. In addition to the dissemination of printed publications, GPO provides a wide range of services to agencies that range from CD-ROM development and production to Web hosting. One of the problems facing the agency is that these and other services are not well known, with the result that fugitive documents are created. Lastly, GPO has been unable to keep pace with current technologies.

Current printing industry trends show that the total volume of printed material has been declining for the past few years, and this trend is expected to continue. A major factor in this declining volume is the use of electronic media options. The move to electronic dissemination is the latest phase in the electronic publishing revolution that has transformed the printing industry in recent decades. This revolution was driven by the development of increasingly sophisticated electronic publishing ("desktop publishing") software, run on personal computers, that allows users to design documents including both images and text, and the parallel development of electronic laser printer/copier technology with capabilities that approach those of high-end presses. These tools allow users to produce documents that formerly would have required professional printing expertise and large printing systems.

These technologies have brought major economic and industrial changes to the printing industry. As electronic publishing software becomes increasingly sophisticated, user-friendly, and reliable, it approaches the ideal of the print customer being able to produce files that can be reproduced on the press with little or no intervention by printing professionals. As the printing process is simplified, the customer can take responsibility for more of the work. Thus, the technologies diminish the value that printing organizations such as GPO add to the printing process, particularly for simpler printing jobs. Nonetheless, professional expertise remains critical for many aspects of printing, and for many print jobs it is still not possible to bypass the printing professional altogether.

The advent of the Internet permits the instantaneous distribution of the electronic documents produced by the new publishing processes, breaking the link between printing and dissemination. With the increasing use of the Web, the electronic dissemination of information becomes not only practical but also more economical than dissemination on paper.

GPO will need to document user requirements for the future system. Two significant weaknesses repeatedly mentioned in the General Accounting Office (GAO) report to Congressional Addressees of June 2004 are "presentation of new products and services" and "responsiveness to customer needs," indicating that many current and future requirements are not being satisfied by GPO. Users are either doing without or fulfilling those needs elsewhere. Documentation of user wants and needs for both the present and the future is critical to ensure the success of the system.

4.2 JUSTIFICATION FOR AND NATURE OF CHANGES

The following subsections describe the proposed system in terms of justification for changes, the nature of the desired changes, priorities among changes, changes considered but not included, and assumptions and constraints associated with building the system.

4.2.1 Description of Desired Changes. As previously indicated, there is no existing single system or group of systems that adequately meets the needs and mission of GPO; therefore, there is no existing system on which to base the changes. For this reason, the following items are omitted from discussion:

- System process changes;
- Interface changes;
- Personnel changes;
- Environment changes;
- Operational changes;
- Support changes; and
- Any other changes not previously identified.

4.2.2 Proposed System Attributes. From an overall system perspective, the system should possess the following attributes.

- *Infrastructure independence*: An architecture that allows preservation of content independent of any specific hardware and software that was used to produce them;
- *Modularity*: Ability to use plug-in components that can be replaced with minimal impact to remaining components as workload and technology change;
- *Scalability*: Capable of accommodating growth and managing differing sizes of repositories and ever increasing volumes of content;
- *Extensibility*: Be able to handle additional kinds of content over time, not limited to specific types that exist today;
- *Comprehensiveness*: Provide support for content management lifecycle processes for all types of records; and
- *Flexibility*: Enable GPO to tailor content-based services to suit its customers' needs and enable GPO to implement progressive improvements in its business process over time.

4.2.3 Proposed System Capabilities. To meet strategic objectives, GPO must integrate its solution for preservation and long-term access to content with the lifecycle management of that content throughout the Federal Government. To meet the challenges of today and the future, the system should be able to:

- Accept the transfer of content in a wide variety of formats as they were created or stored by their creators and the flexibility to easily adapt to future file formats;
- Ingest, preserve, and provide access to that content;
- Store content in a manner that is independent of any particular hardware and software component over long periods of time;
- Scale in order to store and preserve content based on the predicted digitizing of existing hard copy publications and the discovery and harvest of in scope Federal content from Web sites;
- Provide access to the content in electronic form for all users based on established user rights and privileges, thus ensuring that the system users are able to access all of the content that they are entitled to see;
- Provide access to the content in a manner that is consistent with current technology and the changing expectations of its diverse user communities;

- Adapt to changing technology in order to continue to provide access to and delivery of content desired by the user community; and
- Identify the essential characteristics of the content that is being preserved for the purposes of authentication and certification.

The proposed GPO system should provide the following capabilities in support of GPO content management lifecycle processes.

- Provide end-to-end automated work processes that streamline the content management lifecycle processes for all content;
- Manage the creation, review, and approval of content;
- Support the transfer process of all content (electronic and non-electronic) to GPO, FDLP, and other repositories;
- Support the Preservation Services;
- Ensure that content contained as part of service orders/requests, sales contracts, and/or other agreements that identify content that is to be transferred to GPO, specify the terms and conditions of such transfers that conform to GPO and other Federal standards and requirements as required;
- Support end-to-end tracking of all content during the process of transfer, maintenance in FDLP, processing, preservation, and continuing use;
- Accept transfers of content, check that the content conforms to terms and conditions of the service order specified transfer, and store them in the system;
- Ensure that the content transferred to GPO remains free from corruption and is accessible as GPO undergoes changes in IT;
- Support the description of content held by GPO so that it is clearly identified, discoverable, and retrievable;
- An automated tool must exist for any internal and external user to inform GPO of publications they become aware of in the future;
- Dispose of certain content (e.g., content out of scope for permanent preservation, or in-process work files) as stipulated by the service order or other agreement;
- Enforce restrictions on access and release of content;
- Provide access to electronic content;
- Output authentic and certified copies of the content;
- Output copies of the content as specified by customers;
- Monitor system performance;
- Maintain system security; and
- Provide audit trails of system activity.

4.2.4 Proposed System Interfaces. The proposed GPO system will be Government-wide and will operate within the context of the Federal Enterprise Architecture. The system will be capable of interfacing with other applications throughout the Federal Government for transfer of content to GPO, for retrieval of content by their creators, and for content management lifecycle processes in which GPO interacts with other entities in all three branches of the Government. The volume and diversity of content that is ingested and disseminated, and the potential for expected heavy use of the system, will have considerable impact on GPO computing environment.

Interfaces to other GPO systems as well as to other Government agency systems will be accommodated by the system. Specific interfaces are yet to be determined and will be described in a future ISSD when identified.

4.3 PRIORITIES AMONG CHANGES

The initial and final operating capabilities of the proposed system have yet to be determined; however, the system must have the capability to accept digital content from Government agencies, must be able to accept digitized existing hard copy content, must be able to locate and harvest content that has been placed on Web sites, manage and preserve the content, and provide access to that content. As the vision for the system comes into focus, an incremental implementation approach will be utilized. Using an incremental approach, the functionality required by each increment to be satisfied by the proposed system can be determined and prioritized based on perceived GPO need.

4.4 CHANGES CONSIDERED BUT NOT INCLUDED

There are no changes to the proposed system that were considered in the proposed list of attributes, capabilities, or system interfaces identified in Section 4.2.2 through and including Section 4.2.4. Changes to the items provided in those sections will not be known until the completion of Phase 3 of the Future Digital System project, the systems analysis and design phase; however, discussions in preparation of this *the Future Digital System ConOps* led GPO to conclude that system will not handle billing operations, instead the system will interface with GPO financial system.

4.5 ASSUMPTIONS

The following form the assumptions as currently known for the Future Digital System.

1. GPO will evolve into a Content Originator (a publisher) in addition to its historic role as a Service Provider.
    a. GPO will repurpose and repackage content to create new versions for its dissemination and access programs.

2. Deposited or born digital content will be the primary method of ingest. Conversion of documents is expected to be a transitional activity. Furthermore, Harvesting will continue to be available as required.

3. Agency customers will have more direct control over buying and publishing information products and services. For example, direct electronic ordering for any good or service.

4. GPO will collect and preserve in scope content of the Federal Government for public access.
    a. GPO will amass a comprehensive collection of Government publications from all three branches of Government.
    b. GPO will provide access to that content in a variety of forms and formats.
    c. GPO will preserve that content for access over time.
    d. GPO will receive that content in a variety of forms and formats.

5. The information that flows through GPO can be systematically evaluated and determinations about inclusion can be made by the system. If process and scope are properly identified and defined, a system can be built to make the inclusion/no inclusion determination.

6. Repurposing content for specific user groups (e.g., small business) or markets is a potential new revenue opportunity for GPO. Content granularity will be a key component of repurposing content.

7. Capturing content is essential to the success of the system.

a. Missing content (known as fugitive documents for hard copy output) can be minimized, if GPO provides effective toolsets that aid in capturing content.
b. Capturing all content will give GPO the richest collection possible.

8. The system will process all GPO work, including work that has no preservation or access value provided it meets agency business needs.

9. System concept should not be constrained by statutory limitations.

10. The volume of print will continue to decline.

11. Digital media will still be required as an output vehicle for delivering content.

12. The system will address usability and accessibility requirements and best practices.

13. Storage Management will adhere to best practices for data management.

14. The system must be flexible, extensible, and adaptable.

15. Tools and processes may include human interactions into the foreseeable future.

16. User requirements regarding specific needs for content ingest are not identified. GPO must be able to interface with various agency level Content Management Systems (CMSs), manual processing of orders, etc.

17. GPO must create best practices for Content Ingest into the system and update GPO Style Manual.

18. Customer confidence with GPO administering and managing the agencies content. The system must demonstrate leadership in accepting and managing content to build customer confidence.

4.6 ADVERSE EFFECTS

The risks of not proceeding with development include the following:

- GPO will not be able to effectively achieve its mission and meet user needs if it does not build the system;
- Content will be lost without an effective system for ensuring both the preservation of and access to that content; and
- GPO will not be positioned to provide adequate guidance, assistance, or services to customers to manage their content.

5 CONCEPTS FOR THE PROPOSED SYSTEM

The following subsections describe the concepts of the proposed system with respect to its background, objectives, and scope; describe applicable operational policies and constraints; describe the proposed system; describe user classes and other involved personnel; and describe the support environment.

5.1 BACKGROUND, OBJECTIVES, AND SCOPE OF THE NEW SYSTEM

A high level system overview has been provided in Section 1.3, System Overview. Goals and motivation for the new system are discussed in Section 3.1.2, Motivation for a New System.

- GPO has adopted the use of the OAIS reference model for an archival system that is dedicated to preserving and maintaining access to digital information. See **Figure 5-1, Reference Model**.

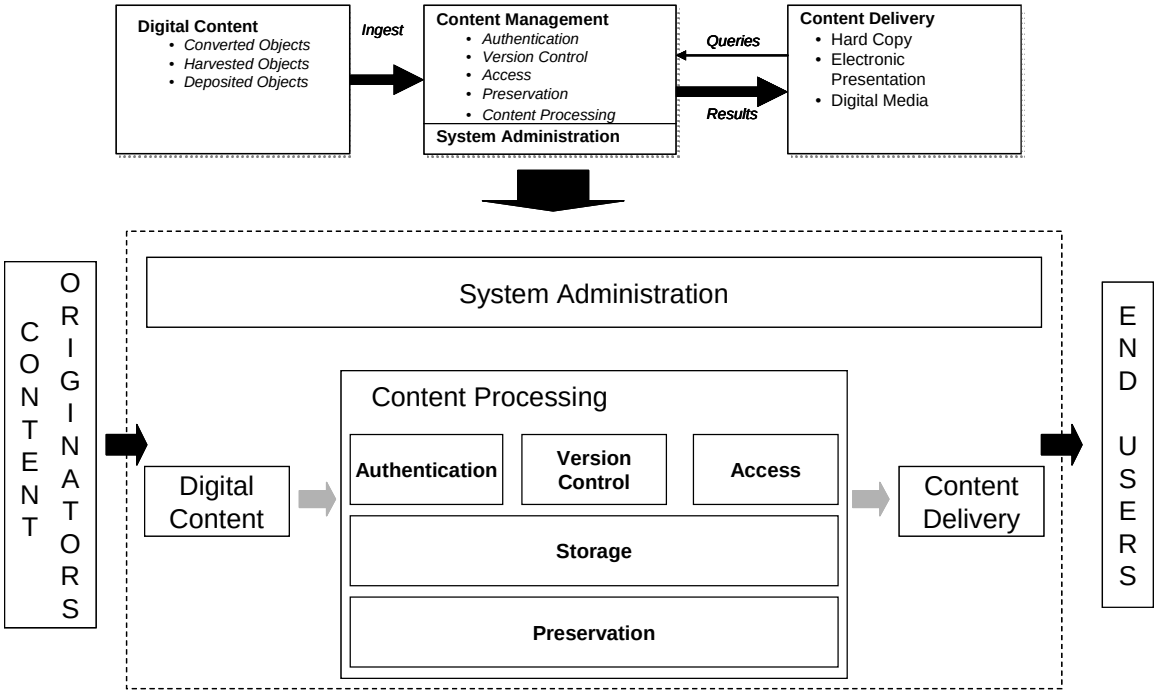


Figure 5-1. Reference Model

The Open Archival Information System (OAIS) reference model was developed by Consultative Committee on Space Data Systems (CCSDS) with broad input from other communities. It was issued by International Standards Organization (ISO) in 2003 as standard ISO 14721:2003: *Space data and information transfer systems -- Open archival information system -- Reference model*. OAIS is a domain neutral reference model with characteristics broadly applicable to the management of any information over time. The OAIS model has been adapted and used in other research collaborations and provides the scalability, extensibility, and interoperability required for a system of this magnitude. This model does not prescribe an implementation. Using the OAIS as a reference model begins the process of defining what is necessary to achieve GPO's strategic objectives for improving the content management lifecycle processes for the management of content of all types and the preservation of thereof. It is recommended that the system be an integrated system that provides OAIS foundation services (see Figure 5-1) such as ingestion of content, storage of that content in the form of electronic records for as long as needed, content

management, and the ability to provide access to the content from anywhere on demand. **Figure 5-2, OAIS Reference Model**, shows the OAIS Reference Model.

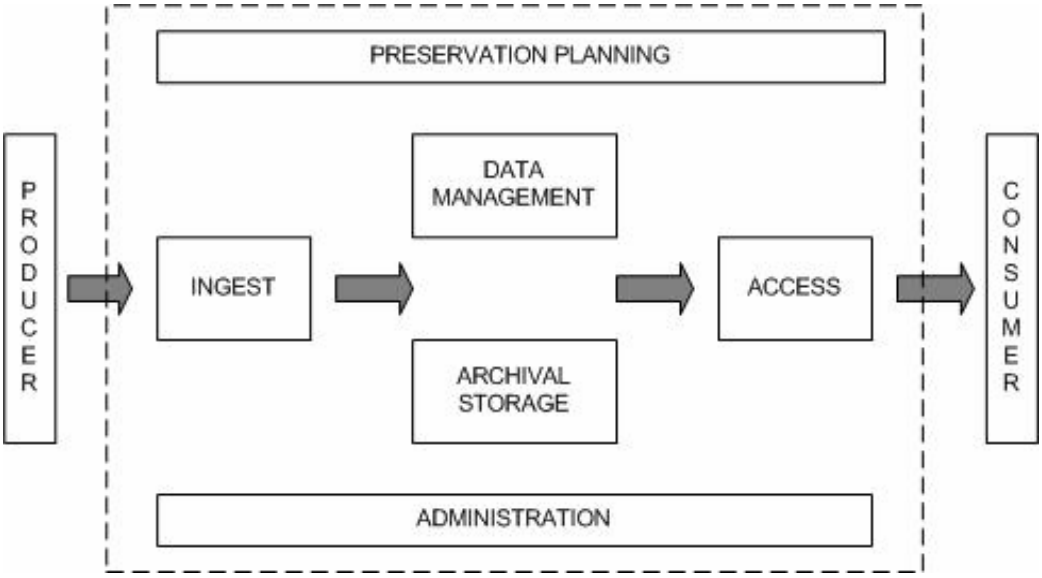


Figure 5-2. OAIS Reference Model

To meet System User expectations, the architecture for the Future Digital System should be able to:

- Ingest digital content from Federal publications.
- Manage content via an enterprise class CMS.
- Support an effective content delivery capability.

The Information managed by the system has two characteristics divided into categories: Content Package Information (CPI) and Business Process Information (BPI). Business process information is administrative, non-content specific information that is used within the business process and package description (PD) to support access aids and data mining. Content package information directly relates to the content and is ultimately used in the dissemination and preservation of the content to the End Users.

These information categories are further defined in a manner that is consistent with the OAIS and the definitions developed for GPO Future System (shown in **Figure 5-3, Expanded Categories of Information**).

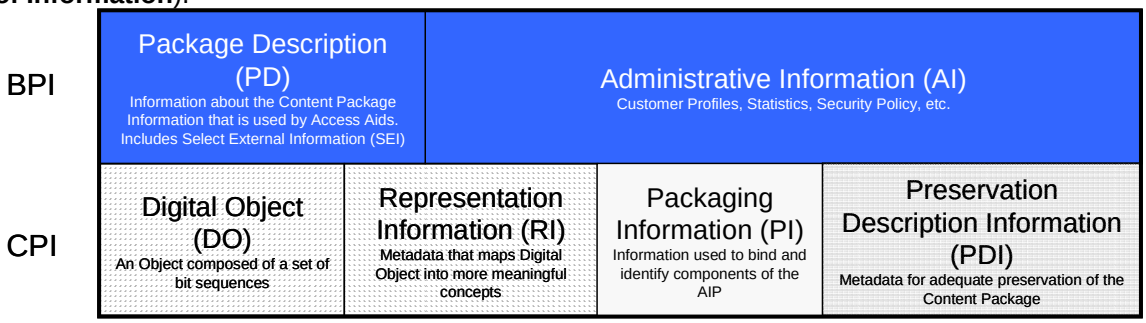


Figure 5-3. Expanded Categories of Information

The RI will include original metadata and any other information, objects, or applications that are required to render the publication at a specific level of accuracy. The RI should also include enough metadata for certification, version control, access, and preservation to take place. For example, a Harvested Content Package may contain an electronic journal, information about how the journal is structured and formatted, and metadata that contains information about the author, publication date, and version. **Figure 5-4** illustrates **Content Package Information**.

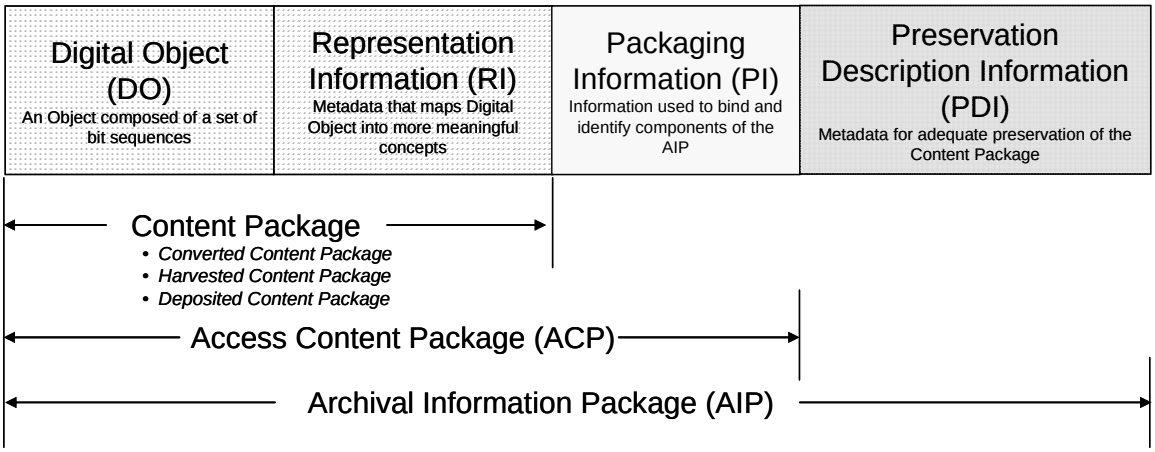


Figure 5-4. Content Package Information

Scope Definitions:

Content within scope of the Future Digital System –

> Content within the scope of the system is any information product originated by a U.S. Federal Government entity or its legal agent.

Content within scope of GPO dissemination programs –

> Content that is in scope for GPO's dissemination and access programs is defined as being authentic and/or as official publications, i.e., final published versions.

In the current state, GPO dissemination and access programs, and printing and binding authority are codified in statute. See Section 2, References. It is possible that legal or statutory mandates could change the scope of publications that are considered within scope. GPO's definition of a publication that falls within scope must therefore be flexible and allow for modifications and changes to be made to it as external elements change in the future.

5.2 OPERATIONAL POLICIES AND CONSTRAINTS

The proposed system is required to be policy neutral so that it can support not only GPO's current operational policies but also changes that are expected to emerge.

Constraints as currently known that may impact system architecture or major system components are:

- The system design and implementation is flexible and adaptable to changes in hardware, software, communication technology, processes, policy, personnel, locations, etc.;
- The system is responsive to policy, but is not policy-constrained;

- The system design is a balance of Commercial Off The Shelf and developed software;
- The system implementation is within the specified budget and timeframe;
- The system will support legacy systems during the transition to the Future Digital System;
- The system will interface with back-end infrastructure systems at GPO (e.g. Oracle); and
- The system will adhere to GPO privacy, security and accessibility policies.

5.3 DESCRIPTION OF PROPOSED SYSTEM

In order to describe the changes to current methods of operation at GPO, it was necessary to review the current functional environment and model its major components to individual capabilities that will be required of the system. Future Digital System components are described in Section 5.3, Description of Proposed System.

5.3.1 Digital Standards. The following subsections describe the components of the system that make up digital standards.

5.3.1.1 Information Packages. Information packages stored and managed in the Future Digital System are entities composed of target digital objects (content) and fundamental representation information, to which descriptive, technical, and administrative metadata are added to enable and facilitate lifecycle processes such as preservation, search, retrieval, and dissemination.

Four types of packages are created and employed by the system:

- Submission Information Package (SIP), which assembles and presents target content and metadata for Ingest
- Access Content Package (ACP), which contains target content and metadata relevant for Access (e.g. search, retrieval) functions
- Archival Information Package (AIP), which contains a fully faithful copy of the target object and rich metadata to facilitate preservation processes
- Dissemination Information Package (DIP), which contains a faithful copy of the target object and metadata to facilitate delivery of the information to users

Operational Environment and Characteristics. Information packages in the Future Digital System will provide the basis for content processing functions including ingest, version control, validation, preservation, and access, as well as for content delivery functions. In order for the system to effectively preserve information and provide permanent access, it must ensure that official content created by Government agencies, and essential information about that content, is managed in an environment that facilitates processing of and access to fully functional information, as well as an array of preservation processes.

Capabilities, Functions, and Features of the Future Digital System. The system will recognize related content objects and metadata as information packages. In order for the system to preserve content and assure access, information packages are created to be responsive to the particular demands of system functions. The Submission Information Package, the Access Content Package, and the Archival Information Package are created in Ingest. The Dissemination Information Package is created in Content Processing. In all cases, the packages must be composed of the fullest possible manifestation of the target object, free of the barriers of proprietary formats, and associated with metadata appropriate to the particular function.

Archival objects (that is, objects in the Archival Information Package) will typically capture the highest quality possible within the constraints of the production of the object or acquisition

methodology. Objects derived for use or access are created within the constraints of the preserved object to satisfy specific user demands, scenarios, or agency customer needs.

Standards and techniques governing the acquisition and processing of preservation level objects and high quality derivatives will derive from standards and best practices in the scholarly and information access communities, as well as statutory and regulatory requirements for permanent public access.

The information packages employed in the system will be iterated across the three sources of digital content: deposited, harvested, and converted: as illustrated below in **Figure 5-5, Content Packages.**

Content Packages	**Submission Information Package (SIP)**	**Access Content Package (ACP)**	**Archival Information Package (AIP)**	**Dissemination Information Package (DIP)**
Deposited Object	Digital object as submitted by CO	Validated Digital object	Faithful copy of the validated Digital object	Faithful copy of digital object rendered for dissemination
Deposited Metadata	Representation information, stylesheets, DTDs GPO submission level elements	SIP metadata + Descriptive metadata, packaging metadata, for access, content transformation, content management, derivation, etc	SIP Metadata + PDI for deposited content	ACP metadata + additional descriptive and packaging metadata for dissemination

Converted Object	Single page, 1:1, uncompressed TIFF, no cropping, deskewing, interpolation or other modification	Files created from converted object and/or other derivatives intended for access (e.g. text, JPEG)	Single page, 1:1, uncompressed TIFF, no cropping, deskewing, interpolation or other modification	Object retrieved from Access Content Storage
Converted Metadata	Full technical information from the conversion following NISO Z39.87-2002; GPO submission level	GPO submission level + metadata for access, content transformation, content management, derivation, etc	SIP + PDI for converted content	ACP metadata + packaging metadata for dissemination
Harvested Object	Digital object as harvested	Digital object as harvested and file(s) as converted for access	Digital object as harvested and file(s) rendered for optimal preservation	Faithful copy of file(s) in the ACP, rendered for dissemination
Harvested Metadata	Representation information and documentation of harvest & transformation(s); GPO submission level	SIP metadata + Descriptive metadata, packaging metadata, for access, content transformation, content management, derivation, etc	ACP metadata + PDI for harvested content	ACP metadata + additional descriptive packaging metadata for dissemination

Figure 5-5. Content Packages

5.3.1.2 Metadata. Actions or processes in the Future Digital System require and/or create information about target content in order to be carried out. This information is recorded, stored, and subsequently used as metadata. Metadata is a structured representation of information that facilitates interpretation, management, and location by describing essential attributes and significant properties of information. Fundamentally, metadata describes how, when, and by whom a particular set of data was collected, what the data is, where it resides, and how it is formatted.

Operational Environment and Characteristics. Metadata creates a systematic approach to expressing information derived or discerned from the content itself or from processes associated with the content. It encompasses static properties (i.e., those related to the specific instance or version of the content being processed, queried, or preserved) as well as the temporal aspects of the lifecycle of the object, a continuum extending from creation through system ingest, preservation, content management, access, and use. Metadata is generally classified in several broad categories, according to its function:

- Descriptive - such as bibliographic information describing, classifying, and characterizing the identity and context of the content.
- Technical - describing file format, computer environment, functionality, etc., in which the content was created or acquired and the attributes of the technical environment necessary to render the content meaningfully.
- Structural - describing interrelationships and hierarchies of files and content.
- Administrative - describing rights, ownership, conditions of use, business rules, etc.
- Preservation - information necessary to maintain viability (the bit stream is intact and readable), renderability (translation of the bit stream into a form useable by humans), and understandability (the rendered content can be interpreted and understood by the intended user). Preservation metadata draws heavily on the other four categories.

Metadata in the Future Digital System must record essential properties and attributes which can be mapped to the major elements in the Future Digital System metadata model, which is broadly adapted from the OAIS metadata model.

Major System Components and High Level Interconnection. **Figure 5-6, Future Digital System Metadata Model**, illustrates the major system components and high level interconnection.

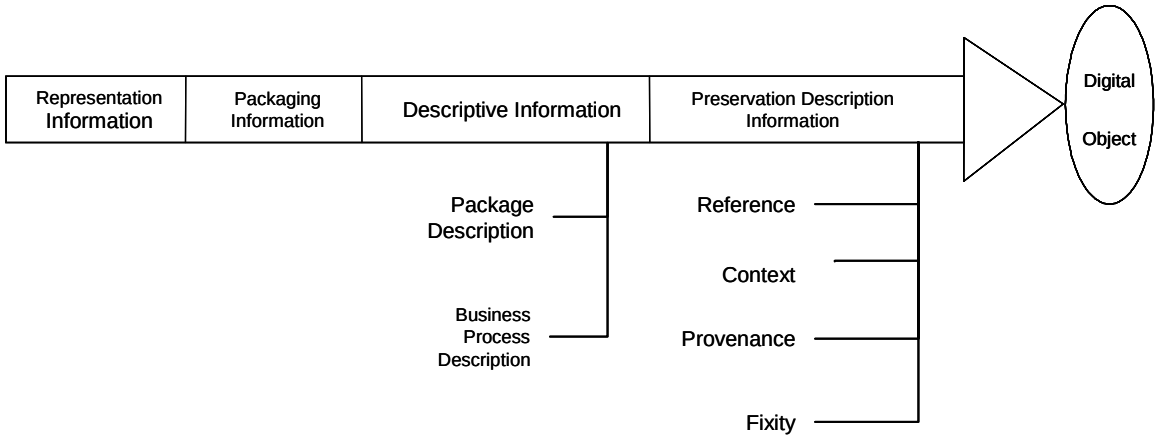


Figure 5-6. Future Digital System Metadata Model

Metadata describes, characterizes, or supports processes related to digital objects that comprise content within the system. Metadata will also describe and support business processes within the system that are not related directly to specific content (e.g., system operation and traffic, user profiles, administrative data, or ordering information).

Throughout the lifecycle of digital objects within the system, metadata is captured, collected, or discerned and recorded either in support of particular processes or to document their outcome. These actions relate to the OAIS model, in which an information package is formed by adding successive layers of metadata to the content information.

The metadata required for these information packages can be categorized into the four major areas above:

- RI allows content to be rendered and understood. Without it, the digital object is merely an undifferentiated bit stream. Although it is not necessarily expressed as metadata for use in rendering, it must be in order for information to be passed to other functions in the system.
- Preservation Description Information (PDI) is necessary to maintain viability (the bit stream is intact and readable), renderability (translation of the bit stream into a form useable by humans or machines), and understandability (the rendered content can be interpreted and understood by the intended user). It is composed of descriptive and technical elements which record identity, fixity, provenance, context, technical environment, lifecycle events, and preservation actions.
- Packaging Information is used to bind and identify the components of the information package into a unit.
- Descriptive Information facilitates access to and management of content. It is composed of elements which record identity, responsibility, version, fixity, structure and technical attributes, administrative data, reference information, and lifecycle events,

In general, metadata will express information that is essential to the processes and products of the Future Digital System. That information will be accumulated by capture from processes, by discernment or assignment by processes or Service Specialists, or by acquisition from an external source. For example, a single digital object, the product of a conversion made by GPO from a printed book, will have metadata captured from the scanning process with which it will be ingested into the system, and this product will subsequently acquire bibliographic information (which is descriptive and structural metadata) as part of cataloging; PDI as a result of the preparation of the Archival Information Package (AIP); and information that will assist in its being available for ordering. Metadata will accumulate according to needs within the system and will be retained for varying periods also depending on applicable system needs. Metadata will drive processes and will record processes for the purpose of facilitating subsequent processes and actions.

Interfaces to Systems and Procedures. In order for metadata to be usefully employed, it must be organized into meaningful, agreed-upon structures that can be interpreted by the system. No single overarching schema exists to express all the data elements necessary for a highly complex system of interconnected processes such as the Future Digital System. It will be necessary to employ multiple schema and ensure that inputs can be shared and utilized throughout.

Ideally, a metadata scheme will define specific data elements and provide a structured presentation of qualifiers and standards, with consistent forms of entry. It is common for particular communities of users or custodians of information to define standards for metadata. Three existing examples are illustrative:

1. Bibliographic information used by libraries for the creation of catalogs and indexes is created according to rigid, internationally vetted standards and is expressed in a scheme known as the

MARC (Machine Readable Cataloging) Standard, which is nearly universally used in the library community that comprises a significant user class for GPO.

2. The METS (Metadata Encoding and Transmission Standard) schema has been developed by the Library of Congress and provides a standard framework for encoding all types of metadata and for expressing complex relationships between the types. Although it is a comparatively recent development, it has gained wide acceptance.

3. In 2003-04, GPO has participated in an international effort to define a core scheme for preservation metadata, led by the Research Libraries Group and Online Computer Library Center (OCLC), Inc. (both broad-based membership organizations that offer various services and utilities to member libraries, most of which are standards-based). The PREMIS working group is scheduled to release a detailed scheme with accompanying implementations scenarios in early 2005.

Capabilities, Functions and Features of the Future Digital System. The system must effectively recognize and interact with a wide variety of established practices and standards for metadata in multiple schema that are in common use for differing functions. Implementation of the system must define methodology for selecting and integrating appropriate schema based on requirements of use and users. As a policy-neutral environment, the system will have the ability to incorporate selected schema for particular functions, and to adapt to modifications in those schema.

The system should effectively recognize and interact with established practices and standards for metadata in multiple schema and provide a capability for mapping or so-called "crosswalks" between schema in use for differing functions.

5.3.1.3 Style Tools. The ability to acquire relevant content is a critical component of the system. The system must allow for the seamless creation, submission, validation, management, and access to the content, and it must provide a simple, easy-to-use method to encourage Content Originators to provide content.

Operational Environment and Characteristics. The system will contain a Style Toolset that allows customers to prepare content prior to ingest by GPO. This toolset must allow for ease of content generation and creation, including auto content tracking, auto layout (if required at this phase), collaboration, and management toolsets.

GPO will create, maintain, and establish a set of best practices for the creation and submission of Content Packages. This guideline will also be used to validate SIP material. This set of best practices will be loosely modeled on the traditional GPO Style Manual and GPO publication 300.6; however, both must be significantly updated to reflect modern information lifecycle creation, collaboration, and management. In addition, the system must contain multiple dynamic toolsets that allow agency customers to provide content and for GPO to ingest that content.

Major System Components and High Level Interconnection.

Content Creation:

- Accepts and validates raw or structured information.
- Content is digitally identified.

Creative Services:

- The system will compose submitted content into appropriate layouts based on user requirements.

- Templates may also be provided to format content based on customized layouts.
- System must allow for a "best view," i.e., creates the best possible layout for the content.

Content Management – Originator's View:

- Provenance is documented.
- Edits and checks are performed in real time as content is managed.
- Content Originator will be allowed to authorize access (i.e., selected users at proof stage, controlled items limit user access, etc.).

Content Validation:

- Content and structure are validated against GPO business rules (best practices and the *Style Manual*).

Interfaces to Systems and Procedures. The Future Digital System's Style Tools will work independently or in concert with existing publishing processes to enable workflow. Style Tools may include collaboration, auto-composition, internal agency chain of responsibility, and version control. Style Tools will be optimized to provide content to GPO which conforms to best practices.

The Future Digital System will allow for a variety of customer interfaces. Examples include:

1. Acceptance of common desktop publishing native files (e.g., Quark XPress, Adobe InDesign, Adobe Illustrator, Adobe Photoshop, etc.);
2. A custom software application that exists on a Content Originator's desktop and performs similar layout functions; and,
3. An interface for audio content that would include speech recognition and conversion.

Interfaces will be extended to the entire content originating sphere of influence and include collaboration tools, approval tools, and other features that would aid customers in creating content. These interfaces must allow customers to easily submit, create, and manage content and to provide the necessary descriptive information for GPO's business and dissemination needs. These interfaces must not hinder agency work and must serve as an incentive.

Capabilities, Functions and Features of the Future Digital System. Style Tools and Ingest capabilities for the system will promote use among the Content Originators of the Federal Government. This requirement encompasses both short-term needs (e.g., accepting non-structured files such as professional publishing files and/or legacy material) and long-term needs such as creating a structured, flexible and extensible standard for content submission. The goal of Style Tools is to move GPO upstream in the content origination process. All Style Tools must conform to GPO's best practices guide for content submission. The Style Tools must allow for:

1. Ease of use, as this is a critical factor if GPO is to fulfill its mission. The interface for Style Tools must allow a novice user to begin using the system easily and quickly with little or no training.
2. Advanced user interface.

5.3.2 Content Originator Ordering Content Originator Ordering provides a system interface for Content Originators. Content Originators may submit content, order and re-order content, and specify Content Delivery and other service options through Content Originator Ordering.

Operational Environment and Characteristics Content Originator Ordering should allow Content Originators and GPO Users to discover the cost of content delivery, choose delivery options, request delivery, and discover payment/billing status for delivery of content when applicable. Content

Originator Ordering shall provide the capability to create, acquire, and store metadata elements specific to ordering functions, preservation needs, version, and job specifications (e.g., SF1, 952, 2511, 3868).

Major System Components and High Level Interconnection. Content Originator Ordering will interface with Style Tools and Ingest. Content Originator Ordering will pass content to Ingest, notify Content Evaluators when orders are placed, and integrate with GPO's financial systems. Context specific help and support will be accessible through Content Originator Ordering.

Capabilities, Functions and Features of the Future Digital System.

- Provides Content Originators with a direct interface to the system.
- Allows Content Originators to order and reorder content.
- Captures all agency processing requirements (e.g., the system must capture all relevant metadata the Content Originator supplies).
- Provides the cost of content delivery based on Content Originator order information.
- Allows GPO Service Specialists to augment Content Originator orders and job specifications (e.g., riders).
- Allows Content Originators to specify options (e.g., output media, quantities, specifications, other preferences) for delivery of content (hard copy, electronic presentation, and digital media).

5.3.3 Digital Content. Digital content is born digital content, converted legacy documents, and harvested material. The focus of the Future Digital System will be on born digital content. The system will also support converted legacy content and harvested material. However, as the amount of deposited content submitted to GPO increases, there will be a gradual decrease in the need for legacy conversion and harvesting.

Digital Content also includes all content currently residing within GPO (e.g., GPO Access).

5.3.3.1 Deposited Content. Deposited content is defined as content intentionally submitted to GPO by Content Originators. The SIP for deposited content will include the digital object received from the Content Originator as well as corresponding customer processing requirements and additional metadata. GPO must establish best practices, templates, and standards for capturing deposited content, including metadata to capture all the customers' requirements. The system must accept all common formats used in the industry. Automatic composition tools may be applied to the digital content upon receipt to create a compliant format for the system.

Operational Environment and Characteristics. The Future Digital System will require the establishment of standards/best practices for orders placed by Content Originators. Moreover, the system must be able to accept content that may be furnished in various proprietary formats. For example, Content Originators may interface the system via customer agency level CMSs. Additional Content Originators will rely on GPO to process orders from manually furnished paperwork.

The system must capture all agency processing information (e.g., billing information, jacket number, agency reference information, etc.). Required metadata will include technical contact information (e.g., GPO 952) that relates to the agency customer requirements relative to rendering formats, quantity, and size (for paper versions), etc.

The system must be able to accept deposited content that is furnished in a wide variety of formats and media, with GPO being able to convert the content into a SIP that is compliant with the best practices referenced above.

GPO will require the Content Originator to provide information about dissemination and preservation needs for the Deposited Content Package. Metadata supporting long-term preservation and access will be created by GPO. The system will be required to maintain and augment this information over time.

Major System Components and High Level Interconnection. Deposited content must be rendered into Submission Information Packages for ingest into the system. The system must be able to interface with various CMSs that are utilized by Content Originators.

Interfaces to Systems and Procedures. Content Originators will use an interface that includes a validation process to input requirements into the Future Digital System. The validation process will contain GPO best practices and templates for content structure and required metadata that will assist the Content Evaluators. Metadata must also be transferred with the content.

Capabilities, Functions, and Features of the Future Digital System.

- Agencies to directly interface with GPO.
- Capability to utilize GPO furnished templates, automatic composition tools, and expert services, including pre-flight of digital content (confirming that requirements are complete).
- Ingest process must capture all agency processing requirements (e.g., the system must capture all relevant metadata the Content Originator supplies).
- The system must accept automated input requirements from customer agency CMSs in addition to manually furnished ordering requirements.
- Content Packages may be pushed or pulled from GPO servers, agency servers, storage devices, etc.
- Electronically accept, store, and transfer content in a wide variety of formats as they were created, with the flexibility to easily adapt to future formats.
- Identify the essential characteristics and structure of the content being processed along with RI.
- The system must have the capability to automatically make created or deposited content Section 508 compliant.

The future system must be able to handle additional types of digital content (e.g., audio, video, etc.) and not be limited to the formats that exist today.

5.3.3.2 Harvested Content. Harvested content is content within the scope of dissemination programs that is gathered from Federal agency Web sites. Discovery, assessment, and harvesting tools will be used to create a SIP.

The harvester will consist of discovery, assessment, and harvesting tools. The discovery tools will locate electronic content from Federal agency Web sites and provide information to the assessment tool. The assessment tool determines if the discovered content is within the scope of GPO dissemination programs and whether other versions of the content already exist in the system and establishes appropriate relationships between versions. The harvesting tool gathers the content and available metadata.

The system shall accept content and metadata delivered by the harvesting function into WIP Storage, from which the system will create a SIP. The system will have the ability to discern whether harvested content or metadata needs to be ingested, and will allow for removal of duplicate harvested content.

Operational Environment and Characteristics. The harvester will consist of tools that will be used to locate and discover publications that are in scope, many of which may be directly published to the Web by agencies.

The result of locating, assessment, and harvesting functions is the creation of a SIP. This includes the actual content that was located and any associated metadata.

Major System Components and High Level Interconnection and Functionality. The future process of locating publications within scope will be based on the definition of a publication that falls within the scope of GPO's dissemination programs. This definition will consist of the criteria that will describe the content, dissemination medium, file formats, structure, etc. External forces such as technology advancements may change the means and formats by which Federal agencies disseminate information.

Discovery Tools: The tool must discover content based on rules and instructions derived from criteria which define a publication within scope.

- The flexibility and ability to define and revise discovery rules and instructions.
- The ability to recognize all file types that may reside on Content Originator Web sites (e.g., PDF, HTML, audio, video, proprietary word processing software, dynamic content, rich media, XML, etc).
- The ability to discover and interpret metadata associated with a given electronic publication.
- Must discover, identify, and characterize information on Federal agency Web sites, including:
    - Deep Web information.
    - Query-based databases.
    - Agency Content Management Systems.
    - Dynamically generated Web pages.
    - Content contained on FTP servers.
    - Content contained behind proxy servers.
    - Content contained behind firewalls.

Assessment Tools: Tools must exist to assess discovered content prior to harvest. Service Specialists may assist in determining if the harvested content is in scope. The specific requirements for assessment tools include:

- The ability to filter out information or resources that do not fall within scope.
- The ability to discern whether a particular publication has already been cataloged by GPO, both in print and electronic formats.
- If a publication has been cataloged in another format or version, the assessment tools must be able to associate the harvested content with the existing content.
- Provide quality control functions to test accuracy/precision of rule application and to cycle results into rule creation/refinement.

Harvesting Tools: The harvesting tool must be able to systematically capture Digital Objects (DO) and Representation Information (RI). Upon capture, the harvested content will be passed to the assessment tools, where it will be evaluated against a set of rules that define a publication that is within scope.

The harvesting tool must also have the ability to derive harvesting rules from the discovery tools. The rules will include the location and structure of the publication along with any other information that is pertinent to the act of harvesting. The harvester will then use the rules to capture the DO and RI and the harvested content will be transferred to the assessment tool, where it will be packaged for Ingest into the CMS.

The following lists capabilities, functions, and features the harvesting tool must possess in the future state:

- The harvesting tool must be able to harvest, store, and add metadata to multiple records simultaneously. The metadata should include when and where the harvested content was captured. It should also include the date and time of capture.
- The harvesting cycle, or how often harvesting takes place, should be configurable.
- The harvesting tool will be able to physically capture a copy of the DO and RI, including original metadata and any other information, objects, or applications that are required to render the content at a specific level of accuracy.
- The harvesting tool will be able to capture all file types.
- The harvesting tool will have the ability to capture and interpret all metadata associated with a given electronic publication.

5.3.3.3 Converted Content. Converted content is digital content created from a tangible product. GPO will continue to work with various user communities on digitizing a comprehensive collection of legacy materials. This digital collection will be made available in the public domain for permanent public access through GPO's dissemination programs.

Operational Environment and Characteristics. The tangible resources held in Federal depository libraries are the legacy materials currently being considered for digitization. Much of this material is in need of preservation and curatorial care.

While scanning may be one method of conversion, there may be other means by which content is converted (e.g., manual text encoding).

Other forms of digitization exist currently and could possibly evolve in the future. There may be instances in which a successful conversion and/or OCR for a given tangible legacy document becomes improbable or impossible due its physical condition and/or characteristics. In these cases, it may be most practical to manually recreate these documents (e.g. using manual text encoding).

Other formats exist within the legacy collection that are not text based (not in "book" format). These file types include analog audio and video. Specifications will be developed on a case-by-case basis for the creation of these files.

Conversion may require a mechanism by which the work performed by the various conversion partners can be tracked and managed in order to avoid duplication of efforts. This effort may necessitate the creation of and/or interface with a registry of conversion projects.

Major System Components and High Level Interconnection. The major entities providing a converted object to GPO are:

- GPO staff; and
- External Service Providers including contractors, library partners and Federal agencies.

Conversion must be performed by the various entities described above at a level of quality that is adequate to support preservation as well as future iterations of derivative products through which GPO will provide public access. As a result, converted files must be compatible with standards and best practices (e.g., resolution, compression, and searchable text creation) put forth by independent organizations, libraries, other Federal agencies, or a combination of these entities, and adopted by GPO.

Capabilities, Functions and Features of the Future Digital System. Conversion projects will produce digital facsimiles and necessary metadata for the system.

5.3.4 Ingest. Ingest is the function that accepts the Submission Information Package (SIP) and prepares the SIP for Content Processing.

Operational Environment and Characteristics GPO must establish best practices and standards for content received during the Ingest process, including metadata to capture all the customers' requirements. The system must accept common formats used in the industry. The system must transform and validate SIPs according to GPO established best practices and standards.

Major System Components and High Level Interconnection. Ingest accepts SIPs created from deposited, harvested, and converted content and content created using GPO Style Tools. Ingest transforms SIPs into AIPs and ACPs and transfers the Content Packages to Content Processing. Decisions regarding scope, authentication, versioning, and persistent naming are made at Ingest.

Interfaces to Systems and Procedures Content Originators and GPO Users will have the capability to submit content to the system through Ingest.

Capabilities, Functions and Features of the Future Digital System.

- Accepts conforming SIPs.
- Rejects non-conforming SIPs and notifies GPO Content Evaluator that content was not ingested.
- Makes scope determinations on SIP (e.g., GPO, FDLP, National Bibliography, Sales Program).
- Validates that SIP metadata meets minimum submission level requirements.
- Authenticates SIP as official and/or authentic as applicable.
- Accepts and transfers integrity marks associated with the SIP.
- Accepts and transfers version information associated with the SIP.
- Performs Version Control by detecting if ingested content is already located within the system.
- Accepts or assigns a Unique ID.
- Can accept or assign a Persistent Name.
- Creates metadata.
- Transforms SIP into AIP and passes the AIP to storage.
- Transforms SIP into ACP and passes the ACP to storage.

5.3.5 Content Processing/System Administration. Content is processed in a centralized fashion within the system. Rules for Authentication, Version Control, Access and Preservation are executed within Content Processing. Centralizing this processing functionality supports a structured Storage Management System (SMS) and streamlines the required processing.

Content Processing accepts, processes, manages and delivers content. Access Content Packages (ACPs) and Archival Content Packages (AIPs) are key content elements that serve to develop Dissemination Information Packages (DIPs) for Content Delivery as well as the basis for the package that is supplied to the Preservation Component for archiving.

Content Processing includes the rule-processing elements for Authentication, Version Control, and Access and Preservation.

System Administration is the link to the Content Originators, the End Users, GPO Business Management functions, and GPO Operations Manager, and it provides the necessary support for processing the content to meet the Content Originator and End User needs as well as the business process needs.

As shown in **Figure 5-7, System Administration/Content Processing Component**, the proposed system can be drawn to illustrate the overarching responsibility of System Administration and Content Processing. System Administration ultimately coordinates the Ingest of Digital Content and the Content Delivery.

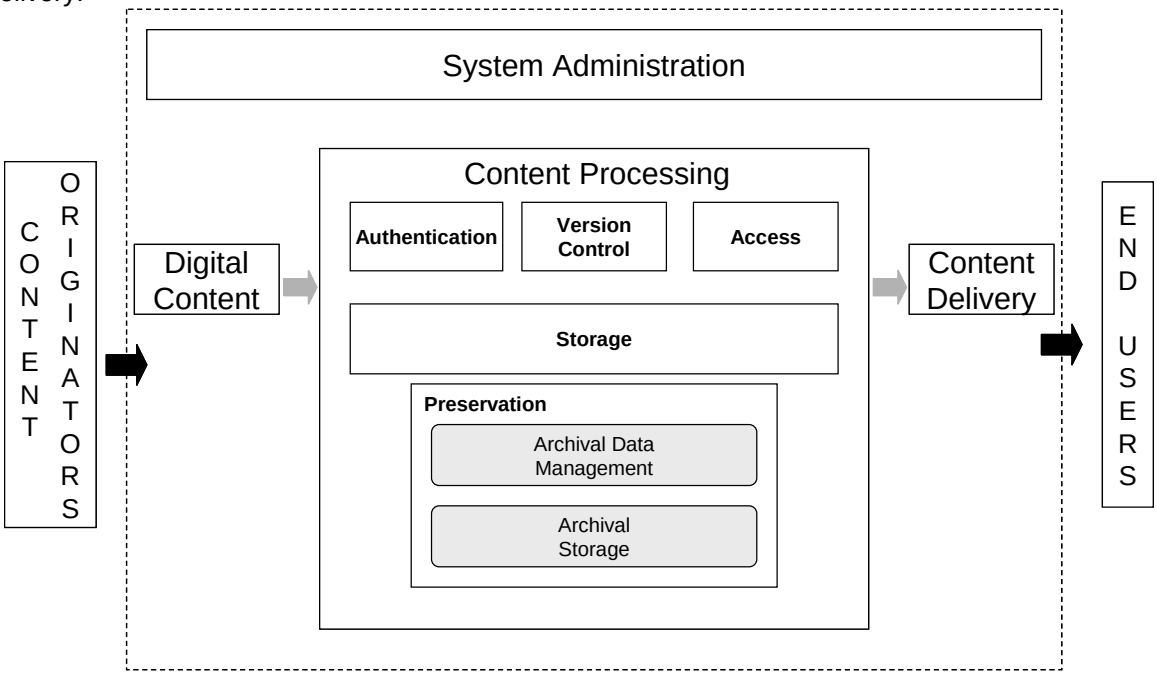


Figure 5-7. System Administration/Content Processing Component

5.3.5.1 Workflow. Workflow Controller or Enterprise Service Bus (ESB), working with System Administration, provides instructions to the other functional elements in the system to process the content in accordance with the Content Originator and End User requirements and in accordance with GPO process and business requirements.

BPI, the Administrative Information that is non-content specific, information that is used within the business process, is managed by the System Administration component (reference Figure 5-7).

Operational Environment and Characteristics. The System Administration Component works across the system, coordinating the functional interactions between the system elements to deliver on the Content Originator and End User requirements.

Since the system is an information management system, a majority of processing will take place in a centralized Content Processing Component. This component *processes* the Content Packages in accordance with the instructions provided by the System Administration Component.

Major System Components and High Level Interconnection. **Figure 5-8 System Administration/Content Processing in the Future Digital System** provides an overview of where System Administration and Content Processing fit within the overall system.

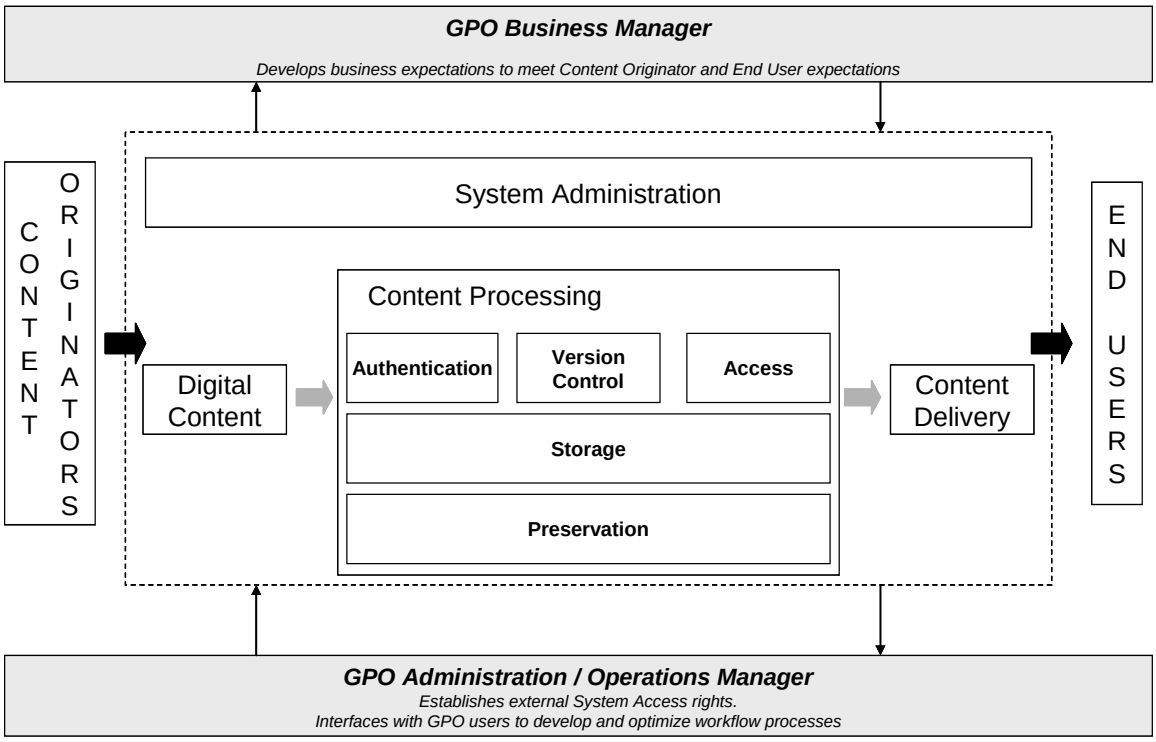


Figure 5-8. System Administration/Content Processing in the Future Digital System

Interfaces to Systems and Procedures.

System Administration:

System Administration is the link to the Content Originators, End Users, GPO Business Management functions, and GPO Operations Manager and provides the necessary support for processing the content to meet the Content Originator and End User needs as well as the business process needs.

In addition, this is the Workflow Controller for the system. In this role, System Administration provides instructions to the other functional components in the system to process the content.

BPI, the Administrative Information that is non-content specific and used within the business process, is managed by the System Administration Component.

Content Processing:

Figure 5-9, Content Processing, illustrates the Content Processing Component.

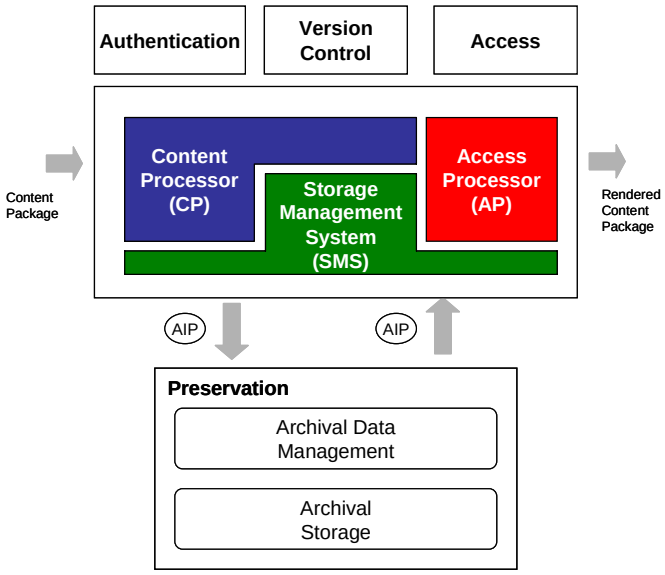


Figure 5-9. Content Processing

Content is processed in a centralized fashion within the system. The system accepts, processes, manages, and delivers Content, AIPs, and the key content elements that serve as the "master" packages for Content Delivery as well as the basis for the package that is supplied to Preservation for archiving.

Content Processing includes the rule-processing elements for Authentication, Version Control, and Access as well as the Content Processing and Preservation.

Authentication –

Authentication will be called upon to analyze the Content Packages and assign the appropriate Authentication attributes, consistent with the requirements defined in the Authentication section.

Version Control –

Version Control will be called upon to analyze the Content Packages and assign the appropriate Version, consistent with the requirements defined in the Version section, which includes Version Triggers and Chain of Custody.

Access –

Access will be called upon to locate and analyze the ACPs and develop Catalog and Index entries for the Content Packages as required. Access also delivers the Access functions, consistent with the requirements defined in the Access sections, which include Search, Cataloging and Reference Tools, Request, User Support, and Interface. Data Mining provides access to BPI.

Preservation –

Preservation will be called upon to execute the digital preservation processes according to the rules and policies established and maintained within the System Administration Component and consistent with the National Collection.

Access Processor (AP) –

- *Receives* requests for Content Delivery via Access
- *Provides* Request to SMS for retrieval of ACP to be used to create a DIP in the CP

Content Processor (CP) –

- *Receives* and *Processes* Content Packages
- *Creates* AIP by collecting the PDI by:
- *Creates* ACPs consistent for Access
- *Creates* DIPs to meet Access requests

Storage Management System (SMS) –

- *Receives* AIP from the CP
- *Receives* ACPs from CP
- *Manages* Pre-Ingest WIP Stores
- *Manages* Content Processing Storage (temporary storage)

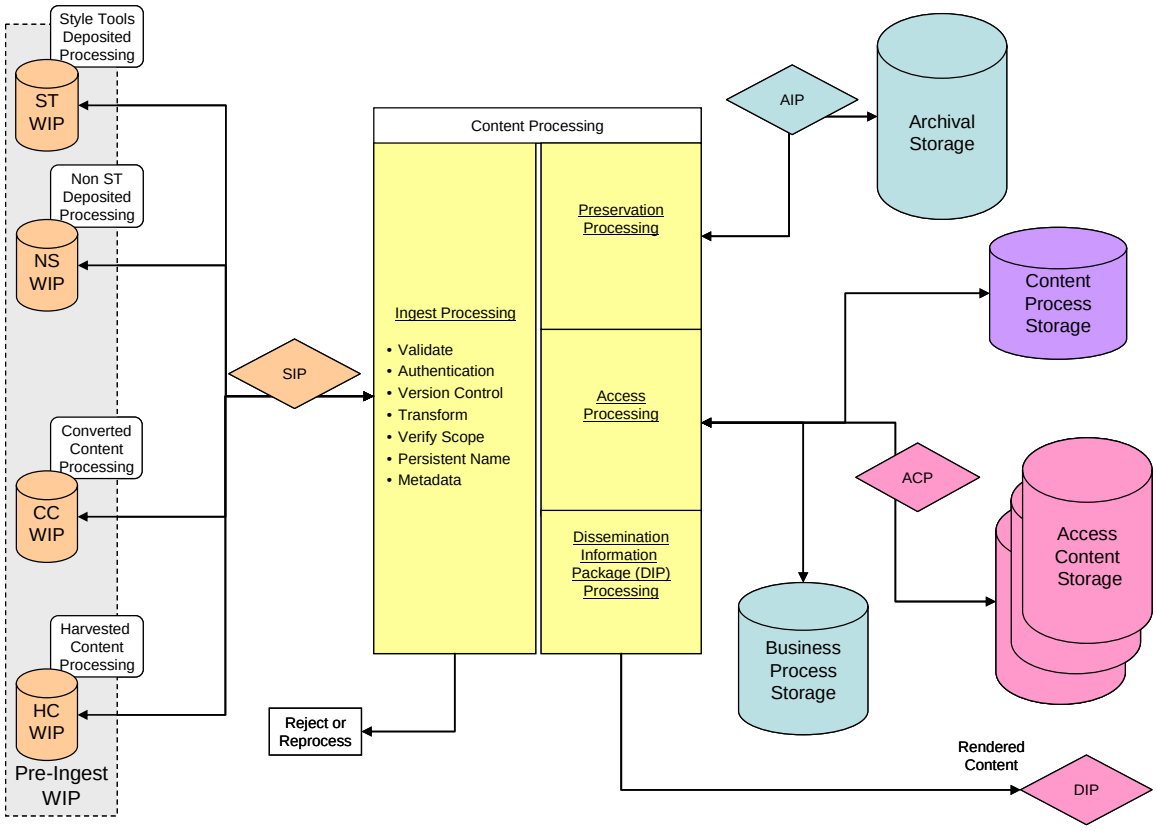


Figure 5-10. Content Processing and Storage Management

5.3.5.2 Unique Identifier. Unique Identifiers are character strings that uniquely identify all content within the system throughout the content lifecycle. Content managed by the system will be assigned an identifier that exists only once and thus is linked indefinitely to the corresponding content. The uniqueness of the assigned identifier ensures that the identifier will refer to only one object.

Operational Environment and Characteristics. The system will create and assign Unique Identifiers to content as defined by GPO business rules.

Digital Objects: A Unique Identifier will be assigned to all digital objects upon receipt into the system.

Content Packages: A Unique Identifier will be assigned to Content Packages (ACP, AIP)

Job Orders: A Unique Identifier will be assigned to Jobs that are created during the CO Ordering Process.

Interfaces to Systems and Procedures. Style Tools will assign Unique ID's to digital objects, which will be passed to Ingest. The system will assign Unique Identifiers to content not created using Style Tools at Ingest. All assigned unique identifiers will be recorded and used in metadata. Once assigned, a Unique Identifier cannot be reused within the system.

Capabilities, Functions and Features of the Future Digital System.

- Ability to create and assign an alphanumeric identifier (ANI) for each unique digital object, job, and content package within the system
- Ability to create and assign a unique ID to a related or continuous piece of content in context
- Ability to record Unique Identifiers in metadata

5.3.5.3 Persistent Name. In order for the digital content managed by the Future Digital System to be easily found and shared by a wide range of users with different needs and using different systems, there must be a simple way of reliably and unambiguously identifying each resource independent of its location.

Operational Environment and Characteristics. Persistent Naming allows for an interoperable schema of identifiers that uniquely identify content, support permanent access to that content, and support access to information about the content.

Major System Components and High Level Interconnection. A resolution system will locate and provide access to content and metadata associated with assigned persistent names. The system must be able to associate persistent names to existing legacy GPO naming schemes (i.e. PURLs).

Interfaces to Systems and Procedures. The system will assign persistent names to content packages at Ingest. All assigned persistent names will be recorded and used in metadata. Once assigned, a persistent name cannot be reused within the system. Procedures will be necessary for updating the location information associated with the identifier if the content has been moved or removed from the system.

Capabilities, Functions and Features of the Future Digital System.

- Ability to create and assign a persistent name at Ingest.
- Scalability in terms of persistent name assignment and resolvability.
- Support for resolution of a single persistent name to multiple distributed locations and content versions.
- Capability to log persistent name transactions

5.3.5.4 Authentication. The Authentication function within the system will verify that the digital object is authentic and/or official. It will also have the capability to certify the content as authentic or official.

Authentication and certification are an integral part of the data and content management portions of the system. These functions affect all parts of the system, from ingest of a SIP through delivery of an access product via dissemination techniques. This section also addresses granularity issues as they relate to the certification process.

Operational Environment and Characteristics. According to the definition, a document could be considered authentic but not official, but a document could not be considered official if not authentic. GPO verifies that the information disseminated through GPO is official and/or authentic and, when requested, certifies this status to End Users.

The system will utilize technology in order to ensure these relationships. An example would be the implementation of Public Key Infrastructure (PKI). Implementing digital signatures would allow customers to determine that the files are unchanged since they were "authenticated" by GPO and help establish a clear chain of custody for electronic documents. GPO's PKI implementation would provide security for and safeguard official Federal Government publications.

Major System Components and High Level Interconnection.

- The Future Digital System must have the capability to verify the authenticity of the content, determine its status as an official document, and certify this to End Users.
- The original harvested or converted content ingested by GPO will be verified as authentic but not official. To make the harvested or converted content official, there must be action from GPO and the Content Originator.
- The system must have the capability to verify the authenticity of the content, collaborate with the Content Originator to establish its status as an official document, and certify this to End Users.
- The system must be able to certify to the End User that content disseminated by GPO is authentic and/or official.
- When encryption standards change, GPO must have the capability to re-certify content at the higher, industry standard encryption levels in order to keep the digital signatures valid and secure.
- The system must provide an automated validating mechanism/process by which the End User would be able to view digital certification information.
- The system must support the digital certification of files in all formats, static and dynamic.
- Digital content ingested by GPO will be authenticated by GPO and/or the publishing agency in its complete state. The system must also provide a means by which granular units of content could be digitally certified and/or authenticated.

5.3.5.5 Version Control. A version is a unique manifestation of a publication. Changes beyond an agreed upon threshold or tolerance constitute a new version, and that threshold is a version trigger. The activity of inspecting content for changes and activating the trigger is called version detection. Version triggers and information about content versions should be expressed in metadata.

Version Control within the system will be a process of evaluating digital content throughout its entire life cycle. The chain of responsibility will be reflected in the metadata.

The purpose of Version Control is to analyze content packages and assign appropriate version designation to all content packages in the system.

Operational Environment and Characteristics. GPO must determine, either from information provided by the Content Originator or discerned from the content itself, whether the content is a unique manifestation. GPO will be responsible for managing and preserving all versions throughout their lifecycles within the system.

GPO must develop policies and procedures to define versions and version triggers for all content packages, including all types of content. Best practices will be applied at all stages of the system process from ingest to delivery (for example: International Cataloging Rules).

Major System Components and High Level Interconnection. Version Control will be performed as content is processed.

Content Submission: Ideally, version information is included in the metadata by the Content Originator. This includes reference to other content versions, including hard copy versions.

Content Processing: Apply version control rules to assess the content package and determine content version within the system. Version control tools should be in place to inspect new content packages and monitor existing content packages within the system.

Access: This function will provide information about versions to the user, including version designations, relationships between versions, and chain of responsibility information.

Request and Content Delivery: The selected version must be delivered to the user with appropriate version information attached to indicate that the correct version was delivered.

Administrative Function: the system will allow authorized users to view version information and have responsibility for managing and preserving all versions and application of best practices.

Preservation: The system will preserve all versions of content packages using hierarchal storage management

Capabilities, Functions and Features of the Future Digital System. The capabilities, functions, and features of the proposed system are delineated below.

- Perform Version Control by detecting if ingested content packages are identical to existing content packages within the system;
- Assign and record version identifiers in metadata.
- Apply rules for version triggers.
- Manage all versions of content within the system.

5.3.5.6 Access.

Access will be the primary interface between End Users and the system. Access has been divided into the following functional areas:

- **Search** – Performing queries on content and metadata so that content may ultimately be retrieved from storage and delivered to users.
- **Request** – Requesting delivery of content and metadata.
- **Cataloging** – Adding metadata to content in the form of standard bibliographic records.
- **Reference Tools** – Compiling lists that point to content and resources.
- **Interface** – Creating user and system interfaces for all functional areas, as needed.
- **Support** – Supporting access to content and services.

All user classes will be able to interact with the system through Access. While Search will primarily provide access to content packages, data mining will provide access to BPI. Access will provide information to users on content package authentication and version control. The system may provide multiple levels of user access and store user preferences to support Access. Access must provide open and interoperable access to content packages.

5.3.5.6.1 Search. Search executes queries on content packages and select external information. The process of searching involves creating a query and refining that query until satisfactory results are returned. Query refinement may include combining input queries with stored queries and preferences. The enterprise Search tools must handle user searches of metadata and content both simultaneously and separately across multiple storage levels and internal and external devices. Search tools should produce a highly relevant, usable, and detailed results list that includes the location and description of meaningful information.

Operational Environment and Characteristics. Search should return what the user expects to receive and provide ways to access related information. The Search tools should be designed to meet the needs of all User Classes who will be searching the system. Users should have the ability to perform a search on content as well as the metadata that describes that content. Users may require the ability to rank and filter items in a results-set to further meet their needs.

It will be necessary for Search and associated functions to conform to international standards that promote interoperability among networked systems. Search should meet or exceed industry standards for search and retrieval technology.

Major System Components and High Level Interconnection. Search must include accessible and customizable graphical user interfaces that allow all users to submit and refine queries and view and export results sets. Search must conform to international standards for search interoperability.

Capabilities, Functions and Features of the Future Digital System.

- **Performance:** Search must perform at or beyond industry standards for search and retrieval technologies.
- **Standards Compatibility**: Search must support the library standard for catalog records, MARC, the Book Industry Study Group standard for Online Information Exchange (ONIX), and other relevant standards.
- **Interoperability**: Search function should conform to search interoperability standards. For example, ISO 23950 (formerly NISO Z39.50) compatibility provides for interoperable search across locators for information and collections of information. Conforming to interoperability standards will promote interoperability with Internet search engines and federated search utilities. etc.
- **Reporting**: Search must be able to interact with other elements in the system, including the System Administration and Data Mining functions, to log all searches for use in reporting usage statistics. Please refer to the Data Mining section for more information.
- **Search Multiple Formats and Levels of Granularity**: the system must have the ability to search multiple media, file formats (e.g., audio formats, video formats, PDF, ASCII, HTML, XML, and other future formats) and levels of granularity.

5.3.5.6.2 Request. Request will allow End Users to request delivery of content and services available from the system. Request will allow users to discover the cost, if any, associated with the timely delivery

of content and services available from the system. Content and services may include GPO sales publications, publications within the FDLP.

Operational Environment and Characteristics. Request should accommodate all expected user requirements (e.g., request delivery of free content, allow users to discover the cost of content delivery, choose delivery options, submit payment for delivery). Furthermore, Request should allow Users to order and submit payment for services delivered by GPO or its agents. Variable fee schedules will be applied based on User Class. These services may include but will not be limited to training, creative, consultation, and legal services. Examples include training requests via the Institute for Federal Printing and Electronic Publishing and training requests from the Federal depository libraries. Based on User Class, some users will not be charged for GPO services.

Major System Components and High Level Interconnections. Request tools must work in conjunction with Search, Reference Tools, Cataloging, and Retrieval tools for Content Delivery. Request will provide the cost of delivering content that has been located by Search, request delivery of content, and facilitate the acceptance of payment. The Request, with the corresponding location and description of the content, will then be sent to the retrieval tools within Content Delivery, and the Retrieval tools will request content from Content Processing. Request must also integrate with other business processes such as inventory control, order tracking, financial, etc.

Interfaces to Systems and Procedures. End User requests for content and services will be sent from Request to Retrieval within Content Delivery.

Capabilities, Functions, and Features of the Future Digital System.

- **Interoperable**. Request will have the ability to interact with GPO's IT infrastructure for a variety of services, including financial, inventory, etc. In addition, the Request function must be designed so that it is accessible to third party systems.
- **Fee and no-fee based Requests.** The system must have the capability to process no-fee based and fee based content delivery requests.
- **Security of Customer Transactions**. Request must ensure that customer transactions can be conducted in a secure environment.
- **Request Content and Services**. Users should have the ability to request both content and services.
- **Stored User Preferences and Request History**. Request must enable customers to store and access user preferences and order histories in a secure environment. The user preferences may contain request status, delivery preferences, preferred payment methods, request tracking, and other customer information. Customers who have purchased products from GPO must be able to review details of all current and prior requests.
- **Unique Identifier**. Request will have the ability to generate a Unique Identifier for each request.

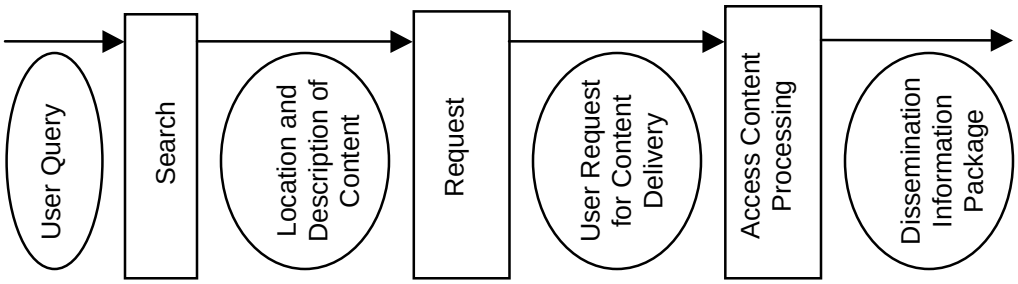


Figure 5-11. Search and Request

5.3.5.6.3 Cataloging. These tools create descriptive metadata that conform to accepted standards, and support access and delivery of standard bibliographic records.

What GPO historically called the "Cataloging and Indexing Program" is now referred to as the National Bibliography (NB) Program. Content in the scope for the National Bibliography are the final published versions of authentic U.S. Government publications.

Cataloging is comprised of the processes involved in describing content to identify or characterize it, providing "entry points" (terms) peculiar to the information or document, e.g., author, title, subject, and format information, by which the information can be located and retrieved. The immediate product of Cataloging is bibliographic records, which are then compiled into catalogs. Indexing is the process of compiling a set of identifiers that characterize a document or other piece of information by analyzing the content of the item and expressing it in the terms of a particular system of indexing. In GPO context, cataloging and indexing is the statutory term for the processes that produce the *Catalog of U.S. Government Publications* and its indexes and various sales products, such as the *Subject Bibliographies*. In the Future Digital System context, Cataloging is the process or results of applying bibliographic control to final published versions.

Operational Environment and Characteristics. GPO has a legal mandate under 44 U.S.C. 1710-11 to prepare and publish a "comprehensive index of public documents," including "every document issued or published…not confidential in character." GPO's library customers expect that this mandate will be fulfilled through the creation of descriptive (access) metadata, i.e., cataloging or bibliographic records, that conform to accepted national library standards and practices.

Major System Components and High Level Interconnection. The Cataloging activities take place within Access, in the Content Processing layer of the system. GPO Service Specialists create descriptive metadata that the system is capable of expressing in library or other standard structures.

Not all Future Digital System content will be cataloged. For example, an agency print order for envelopes will result in metadata in the system, but the envelopes will not meet the scope criteria to qualify for cataloging.

Interoperability with the Integrated Library System (ILS) and the Online Computer Library Center, Inc. (OCLC) system for creating bibliographic metadata and storage capabilities for structured data is required. Bibliographic metadata includes links to content maintained on various sites managed by or completely external to GPO, e.g. the OCLC Digital Archive, on various agency and library sites, etc.

The Cataloging process will use applicable metadata elements, including metadata created by the Content Originator that is harvested along with the digital object to which it is related. GPO will also acquire bibliographic metadata from external Content Originators and Service Providers (e.g., library and agency partners, OCLC).

GPO provides metadata records to various users (e.g., individual libraries, value-added resellers, the Library of Congress, etc.) in a variety of standard formats (e.g., MARC or ONIX).

Interfaces to Systems and Procedures. The Cataloging activities take place within Access, in the Content Processing layer. Cataloging will use applicable metadata elements that are acquired with SIPs. Ingest will review SIPs for required metadata elements.

Capabilities, Functions, and Features of the Future Digital System. Cataloging will draw upon metadata acquired and stored by Future Digital System, and will transform or organize that data into meaningful structures (e.g. MARC, Dublin Core, ONIX).

5.3.5.6.4 Reference Tools. Reference Tools present information locators to End Users, and add value to the End User experience by assisting in the discovery and access process. Reference Tools will contain metadata, references to metadata, and references to content. Reference Tools will assist users in locating content packages and select external information.

Major System Components and High Level Interconnection. Reference Tools will include lists and resources that assist users in locating and accessing content. Reference Tools will have the ability to create, acquire and store metadata (e.g., MARC), references to metadata (e.g. Subject Bibliographies), and references to content (e.g., Federal Agency Internet Sites, Browse Topics, etc.).

Capabilities, Functions and Features of the Future Digital System.

- Reference Tools should be applications or lists that assist users in locating and accessing segments of related content and/or metadata within or outside the system.
- The interface to the tools will be published such that external tools will be able to function with the Reference Tools. For example, in order to enhance the visibility of Government publications, the Reference Tools should be interoperable with other industry standard third party reference tools, such as Internet search engines and publication distributors.
- Reference Tools will interact with Search tools.
- Reference Tools will create additional value for the user experience.
- Reference Tools may be available in a variety of formats, including electronic and tangible formats.
- Reference Tools will be created based on internal and external user needs.
- Reference Tools will have the ability to manage and interpret metadata.

5.3.5.6.5 Interface. Interface will provide users with direct access to content, metadata, and services available from the system. Interface will meet the needs of all user classes. The system should provide graphical user interfaces and system interfaces where deemed appropriate.

Operational Environment and Characteristics. GPO will develop interfaces for all internal and external user classes that will allow users to perform authorized functions. Users will have the ability to "opt-in" to acquire the capability to customize default interfaces in order to create an environment better suited to their needs and preferences. In accordance with GPO's mission, a default End User interface will be provided to allow citizens to access official Federal Government information without requiring them to submit any form of identification.

Major System Components and High-Level Interconnections. Interface will be comprised of graphical user interface and system interfaces. Graphical user interfaces will consist of a workbench, or set of user tools related to performing a system function. Changes to a user's workbench will be saved and made available to "opt-in" users in subsequent sessions.

Interfaces to Systems and Procedures. Interfaces will be made available for all system functions.

Capabilities, Functions, and Features of the Future Digital System.

- The system should provide an interface for all functional elements.

- The system should provide for the creation of system interfaces that promote interoperability among networked systems (e.g., APIs)
- The system should provide for the creation of graphical user interfaces.

5.3.5.6.6 User Support. User Support will assist in delivering services to Users (e.g., interactive user assistance, real-time alert services, training requests, and relational databases that track user queries and preferences). These services may or may not be delivered in conjunction with Content Delivery.

Operational Environment and Characteristics. User Support enhances GPO's ability to provide service to users, such as answering questions, training, etc. User Support receives and processes requests to be executed within and outside the system.

Major System Components and High-Level Interconnections. Major system components will include relational databases, customer relationship management tools, and other tools that are visible through Access.

Capabilities, Functions, and Features of the Future Digital System.

- Ability for authorized Users to input, store, and manage user preferences and queries.
- Support a helpdesk and knowledge bases.
- Facilitate training requests.
- Enable alert services (e.g., RSS feeds, blogs, listservs).
- Support interactive information exchange (e.g., chat, discussion groups, web conferencing).

5.3.5.6.7 Data Mining. Tools and processes for the extraction, analysis, and presentation of BPI to enhance internal and external business efficiencies.

Operational Environment and Characteristics. GPO will provide intuitive data mining capabilities using BPI, including selected access to external data repositories. The system will be able to capture the use history of various dissemination tools (e.g., access and downloads from Web sites and databases, the path users took through the site, etc.), subject to privacy and legal restrictions. The ability to track purchases (e.g., through GPO Sales Program, capturing all data from orders received via the system Ordering Tools, etc.) will also be required.

GPO must provide users with tools to notify them about the latest data posted on Web sites that is relevant. The system will allow for various data mining capabilities.

Major System Components and High Level Interconnection. The major system components of the Data Mining function include an interface that allows users to submit and refine queries and view and export results sets.

Interfaces to Systems and Procedures. Data Mining is a sub-component of Access within the Content Management Component, with interfaces to the System Administration Component and the Content Processing Component.

User queries will be submitted against the system's BPI and external data repositories using various automated data mining tools.

Capabilities, Functions and Features of the Future Digital System.

- Ability to extract BPI in multiple formats from the entire collection.

- Ability to normalize data based on administrator defined parameters (e.g. identify missing values or metadata, data formats, types and discrepancies, and anomalies.).
- Ability to perform analyses on BPI (e.g. cross tabulations, categorization, clusterization, regression analysis, data patterns and relationships).
- Ability to present BPI according to user preferences and GPO business rules (e.g. views based on access levels, exporting of results, linking of results to data).
- Ability to data mine BPI within the system at multiple levels of aggregation and granularity (e.g., contract Service Provider performance history, customer agency billing information, ordering habits, preferences of customers and users, etc).

5.3.5.7 Preservation Services. The Future Digital System will provide digital archiving and Preservation Services for in-scope content. Preservation Services will enable comprehensive, timely, permanent public access to the final published, official version(s) of U.S. Government publications in digital formats. Preservation of *tangible* versions of in-scope content is beyond the scope of the Future Digital System, but information about or access to such content may be provided by Future Digital System metadata.

Operational Environment and Characteristics. Users expect GPO to provide permanent public access to official U.S. Government publications. The mandates of 44 U.S.C. Chapters 19 and 41 establish GPO's responsibility for providing permanent public access to tangible and digital U.S. Government publications.

Future Digital System will manage preservation processes for archived content. Preservation copies of digital publications, AIPs, with associated metadata, will be maintained in Future Digital System Archival Storage.

GPO will provide online public access and other services derived from the digital preservation masters in Archival Storage. Access Content Packages, sometimes referred to as “derivatives,” of the stored digital publications will be available for no-fee online use by the public as well as for print-on-demand and other dissemination services.

GPO’s digital archiving and preservation objectives are achieved by developing and/or following best practices that comply with an adequate, coherent, and widely understood framework for reliable, accountable, and manageable digital archives.

Several levels of public access to authentic and official publications are provided. Tangible products are accessible in Federal depository libraries. Digital content is accessible at or through Federal depository libraries, from the content originating agencies’ Internet sites, from Future Digital System Access Content Storage, or from GPO permanent public access partners. Only when all of these avenues of access are exhausted do the AIPs stored in Archival Storage come into use, and in those circumstances Access Content Packages are produced to support public access. There is no direct public access or use of content in Archival Storage.

Archival Storage will consist of collections of AIPs with identical content located at multiple sites. This redundancy ensures guaranteed preservation in the event of a disaster or significant discontinuation of service at a single site.

Under 44 U.S.C. Chapter 31, Records Management by Federal Agencies, and Chapter 33, Disposal of Records, agencies are responsible for transferring Federal records created during their conduct of public business to the National Archives and Records Administration (NARA). Content defined as Federal records (44 U.S.C. 3301) includes official publications in scope for GPO’s dissemination activities. Therefore, content originating agencies have affirmative responsibilities to provide copies of

official publications to both GPO and to NARA, although in the current state the required actions occur at widely separated points in the information life cycle.

In 2003 GPO was recognized by NARA as an affiliated archive for the electronic records constituting the core content on *GPO Access*. Building on the affiliate status, GPO may offer agencies a service that allows Content Originators to take a single set of actions that will satisfy the legal requirements to provide official publications to both GPO and NARA.

Major System Components and High-Level Interconnection.

Archive Storage. Future Digital System Archival Storage preserves the content of official U.S. Government publications in digital format. The specific preservation processes required by GPO are a policy determination. Future Digital System must be capable of supporting activities necessary to keep content accessible and usable, including migration, refreshment, and emulation.

Interfaces to Systems and Procedures. Archival Storage receives AIPs from the ingested function, in the Content Processing layer. Preservation processes take place within the Content Processing layer, and the resulting AIPs are maintained in Archival Storage.

Archival Storage will interface with Content Processing to receive AIPs for preservation and to provide preserved content for the creation of derivative products, including Access Content Packages. The preservation sub-component of Content Processing executes preservation processes Future Digital System content.

The system will allow collaboration with Content Originators to ensure that their preservation needs are met.

Capabilities, Functions and Features of the Future Digital System.

- Content is passed between Content Processing and Archival Storage.
- Content is stored in Archival Storage.
- Preservation processes are executed in Content Processing.

5.3.5.8 Storage Management. This section covers the Access Storage Management for the system.

Operational Environment and Characteristics. Storage Management is a key component of Content Management that must provide and coordinate access, backup, and archiving of authentic and official Government information as well as ensure data reliability.

Storage Management will consist of facilities that are scalable and support increasing and changing storage requirements.

Major System Components and High Level Interconnection.

- Storage System Management: This component handles the actual management of the hardware storage devices.
- Storage Media Management: This component handles various media of the system for performance optimization.
- Back-up Management: This component is responsible for maintaining the system back-up and recovery.

Interfaces to Systems and Procedures

Figure 5-12, Storage Management in the Future Digital System, places storage management in the system context.

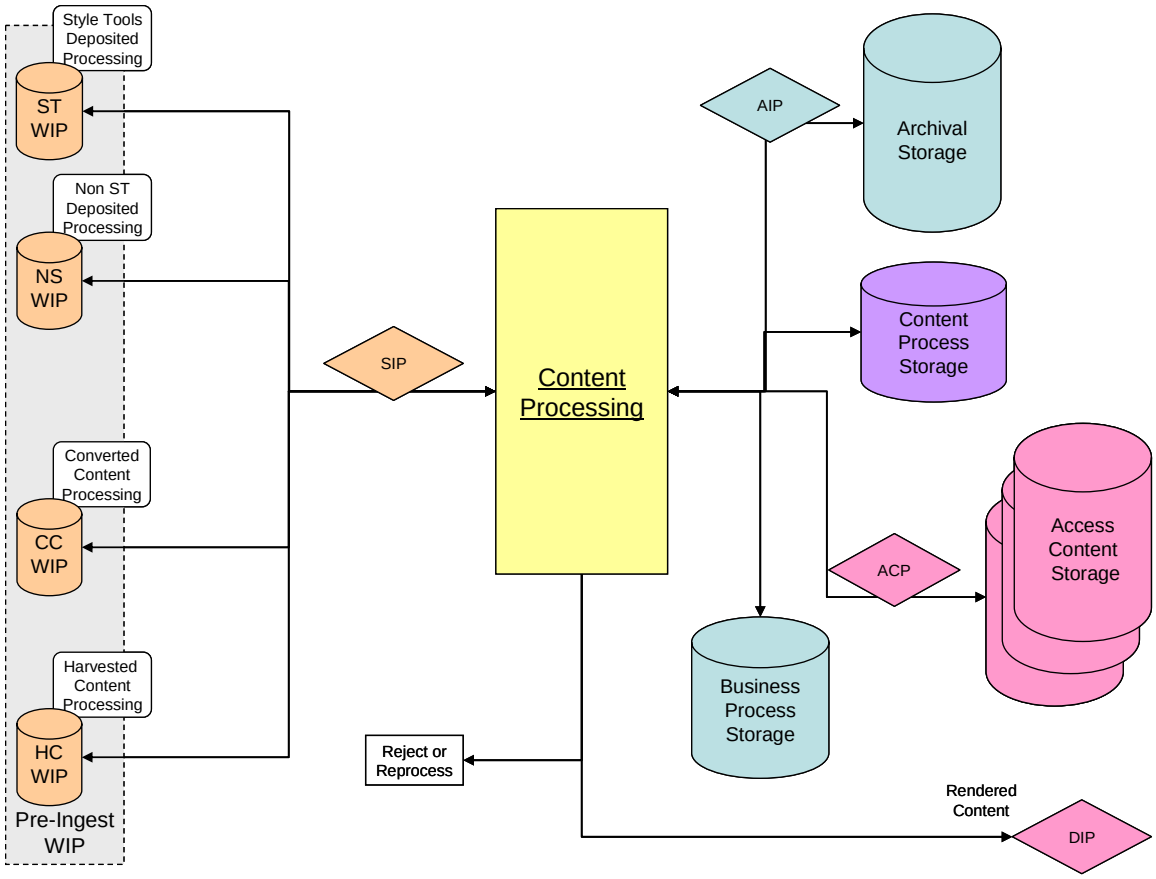


Figure 5-12. Storage Management in the Future Digital System

Capabilities, Functions and Features of the Future Digital System.

- Storage must be accessible from multiple sites.
- The Storage Management functionality must support rules/policies for retrieval.
- Storage accepts and delivers AIP, ACP and Business Process Storage information (BPI) from the CP
- Storage facilities must ensure uninterrupted availability to content packages
- Executes Hierarchical Storage Management on content packages
    - Near Line Storage for Objects that need fast access
- Manages WIP stores for Pre-Ingest functions.

5.3.5.9 Security

Operational Environment and Characteristics The Future Digital System must provide appropriate confidentiality, integrity and availability controls for Future Digital System information and processes. The Future Digital System security capabilities are also derived in the context of requirements for authentication, authorization, administration and auditing. Security requirements include enforcing restriction on access to content (both authentication and authorization), assigning user rights (authorization), and maintaining system security (auditing). In addition, Security requirements include those capabilities that also support the necessary technical controls, operational controls and management controls for Future Digital System, such as monitoring capabilities for security. Security concerns both internal and external system interfaces, as well as operational processes associated with the Future Digital System.

Major System Components and High Level Interconnection.

- User access controls
- User privacy controls
- Confidentiality
- System administration and logging

Interfaces to Systems and Procedures. The Security function works across Future Digital System. The Security function principally interfaces with the System Administration and Workflow to support user access in accordance with policy.

Capabilities, Functions and Features of the Future Digital System.

- The system shall have the capability to support identification of users
- The system shall have the capability to support user access in accordance with the user's authorization rights and with system security policy.
- The system shall have the capability to assure integrity of content within the system
- The system shall have the capability for authorized security administrators to set and maintain system security policy
- The system shall provide the capability conform to GPO's privacy policy and Federal privacy laws and regulations
- The system shall have the capability to provide confidentiality of user data
- The system shall the capability to reconstruct complete transactions

5.3.6 Content Delivery. Content Delivery refers to mechanisms for delivering content in a method or manner that fits the requirements of the Content Originator or End User. The delivery methods include hard copy, electronic presentation, and digital media.

The reference to content delivery is a reflection of the expansion of electronic presentation as the primary dissemination method for authentic and/or official content of the Government.

In the content delivery environment, there are two user class distinctions: 1) Content Originators who are End Users and 2) End Users.

- End Users are the end consumers who are not involved in the Ingest process. They do not submit anything into the system; however, they take content out in various presentations.
- Content Originators who are End Users are the agency and congressional customers who are the creators of content. In a future state, they will likely be subcategorized as follows:
    1. Customers who will submit content in and not take delivery of content.
    2. Customers who will submit content in and take delivery of content on a piece-by-piece or on-demand basis.

3. Customers who will submit content in and expect large quantities of deliverables (e.g., press runs of printed material.)

Interface With Content Management To Obtain Content For Desired Output. An electronic interface will be developed between the system and Service Providers that will allow electronic content and all information necessary for producing the required end product to be accessed electronically by Service Providers. Users requiring delivery of content will be able to request it via Access tools.

Capabilities, Functions and Features.

- Comprehensive content delivery toolsets (e.g. End User interface, presentation aid, output aid, etc.)
- Conform to GPO's best practices guide for content delivery.
- Support and deliver customized Content Packages.

5.3.6.1 Retrieval. Retrieval is defined as the tools and processes that retrieve content from storage and provide transformation information to Content Processing.

Operational Environment and Characteristics. Retrieval tools will interpret encoded information that contains the location of content and transformation information.

Major System Components and High-Level Interconnections. Retrieval provides a bridge between Request, Storage, and Content Processing. Retrieval retrieves requested Content Packages from storage and passes the Content Packages and Request instructions to Content Processing that will be used to transformation an ACP into a DIP.

Interfaces to Systems and Procedures. Retrieval interfaces with Request, Storage, and Content Processing. Content Processing interfaces with Content Delivery for delivery of DIPs to End Users.

Capabilities, Functions, and Features of the Future Digital System.

- Provides Content Package location information to storage.
- Retrieves ACP from storage based on a user request.
- Provides ACP to Content Processing for transformation to a DIP.
- Passes Request and transformation information to Content Processing.

5.3.6.2 Hard Copy Output. Hard copy is tangible printed content. Print on Demand is hard copy produced in a short production cycle time and typically in small quantities.

Operational Environment and Characteristics. Hard copy output may be requested either at the time of content submission or as an access order.

Content Originators submitting Content Packages will include information on the desired output such as color attributes, trim sizes, binding preferences, etc.

Service Providers will supply general business and workflow information, update their equipment and capabilities, provide unit pricing, and request modifications through the system.

Capabilities, Functions, and Features.

- Comply with quality assurance standards supported by the System Administration function (e.g. GPO Publication 310.1).
- Provide methods of authenticating hard copy output (e.g., watermarks, etc.).
- Meet delivery requirements of users.
- Determine most cost-effective method for hard copy output.

5.3.6.3 Electronic Presentation. Electronic presentation output is the dynamic and temporary representation of content in digital format.

Operational Environment and Characteristics. Electronic presentation will be a method of disseminating information. The primary goal will shift from making the digital format match the printed product to presenting the electronic content in the most useable format.

Capabilities, Functions, and Features.

- Compliance with standards.
- Assuring integrity of delivered content (e.g. virus-free, not corrupt, etc.).
- Provide capabilities based on user requirements (e.g. watermarking, bookmarking, links, indexing, security features, etc.).
- Display of integrity marks to verify content authenticity to users.

5.3.6.4 Digital Media. Digital media is a content delivery mechanism consisting of data storage devices. Digital media includes:

- Data storage devices (e.g., CD, DVD, etc.).
- Wireless handheld devices (e.g., PDA, MP3 players, e-books, etc.).
- Future media (e.g., flexible electronic displays, etc.).
- Storage at user sites (e.g. subscription services, etc.).

Operational Environment and Characteristics. Creation and replication of digital media must be available through Service Providers. The system will determine how to deliver content on the selected media.

I**nterfaces to Systems and Procedures**. Digital media may be generated on-demand by a Service Provider. Data may be pushed to the user's device, or requested and pulled from the proposed system.

5.4 MODES OF OPERATION

The modes of operation for the proposed system as currently known are:

- Nominal
- Degraded
- Maintenance
  - Remedial Maintenance
  - Preventive Maintenance
  - Code Upgrades
- Alternate Site.

Nominal mode of operation describes the system when working at the optimum, i.e., the system is operational and working as intended.

Degraded mode and maintenance mode of operation describe operations in time when the system is working using a reduced string of operations. For example, the system is placed in a maintenance mode in order to perform a software upgrade. Once the software has been loaded, tested, and verified to work, the system is placed back in the nominal mode.

Alternate site mode of operation can be described as occurring when one site has a failure that requires a user to access records from an alternate site.

5.5 USER CLASSES AND OTHER INVOLVED PERSONNEL

The following subsections describe the organizational structure and the class of users, including user capabilities, which are associated with the system.

5.5.1 Capabilities. The capabilities for the primary function within the user classes of the system are described above in the subsections of Section 5.3, Description of Proposed System.

5.5.2 Profiles of User Classes. A user can be defined as anyone who will interact with the system. A user class is determined by the ways in which the user interacts with the system. The major user classes identified for the system include:

- Content Originator – Develops information and content and generates requests for GPO services. The Content Originator works with the Content Evaluator to define the parameters of the Preservation and Dissemination Plan. Content Originator provides the content that will be transferred to the system for subsequent certification and preservation.

- Content Evaluator – Collaborates with the Content Originator to determine the content and if the content is in scope or not. The Content Evaluator establishes/defines the Preservation and Dissemination Plan and determines/makes decisions on what processing will occur, whether to use internal production or external contracting, and whether to include information in the Sales Program and/or FDLP.

- Service Specialist – Supports the customer and is expected to deliver the products and services as determined. The Service Specialist performs contracting, administrative, and preservation functions (e.g., creative services, contract writing and awarding, billing, quality control, cataloguing and indexing, preservation management, and dispute resolution.) The Service Specialist helps to describe the content and is involved with the creation of metadata and uses the system to preserve the content as required.

- Service Provider – The Service Provider delivers the expected services and products after receiving notifications. The Service Provider accepts print orders as an example and also certifies vendors as GPO vendors.

- Business Manager – Develop business expectations to meet Content Originator and End User expectations. Also works with GPO Sales Group to repurpose data in order to provide value added services.

- Systems Administration/Operations Manager – Systems Administration directly supports the overall operations and integrity of the system and its use and conducts such system activities as managing user access rights, monitoring system performance, and scheduling reports. The

Operations Manager interfaces with GPO personnel and makes decisions, including approval of workflow processes. The Operations Manager reviews system recommendations and makes decisions on when and how lifecycle activities related to specific records occur and who will perform the work. The Operations Manager has ultimate responsibility for the completion of tasks and the quality of the products.

- End User – Uses the system to search for and access records, to submit data requests, request assistance via mediated searches, communicate with GPO, and invoke system services.

5.5.2.1 User Class Capabilities. High-level Future Digital System capabilities correspond to specific GPO tasks and the users' needs and desires for the proposed system. These capabilities are organized according to user class, but some capabilities cross user class boundaries and might be employed by users in more than one user class. User classes do not correspond to the system position titles, nor does a user class correspond to a single individual user. Rather, each user class describes a role that a user assumes in interacting with the system. An individual user may assume different roles to accomplish different purposes.

Content Originator

These are the capabilities of the Content Originator (in some cases an agent for the actual Content Originator may act in the role of the Content Originator):

- Creates content
- Generates and submits order requesting GPO services
- Requests GPO assistance with content development or other services (such as training or hosting), as needed
- Makes content available to GPO
- Collaborates with GPO on the requirements for services on the preprocessing of content (creating the Preservation and Dissemination plan)

Content Evaluator

These are the capabilities of the Content Evaluator:

- Determines if document is in scope (GPO scope and GPO Dissemination Program scope)
- Works with Content Originator to determine what processing is needed, and to provide ingest aids
- Determines the content version
- Pushes certified content to Service Specialist
- Develops Preservation and Dissemination Plan
- Coordinates harvesting and format conversion
- Creates service orders for content
- Reports back to Business Manager on content flow for possible sales opportunities

Service Specialist

These are the capabilities of the Service Specialist:

- Provides special services
- Metadata development
- Cataloging and indexing
- Creates copy for Archival Storage
- Preserves content as required
- Certifies and/or authenticates
- Quality services on content delivered including Service Provider Quality monitoring

Service Provider

These are the capabilities of the Service Provider:

- Notifies the system with capability, status and completion of service
- Receives notification of service orders
- Accepts service orders
- Provides services

Business Manager

These are the capabilities of the Business Manager:

- Works with Content Originator to establish requirement and guidelines including pricing, dispute resolution process, etc.
- Works with Service Specialists/sales program to repurpose content for value added services
- Sets pricing levels for GPO services

Administration/Operations Manager

These are the capabilities of the Administration/Operations Manager:

- Establish and maintain user accounts
- Schedules reports as needed
- Uses tools to monitor the system
- Uses the system tools to develop workflows
- Modifies workflow when bottlenecks in the system occur
- Quality function administration

End User

These are the capabilities of the End User:

- Uses the system Access tools to locate content
- Uses the system tools to retrieve content
- Uses the system tools to customize content
- Uses the system tools to order content
- Uses the system to access selective BPI (e.g., profile information, order status)
- Works with GPO when needed to locate and order content

5.5.3 Interaction Among User Classes. The proposed system described herein is an overall conceptual workflow model that depicts where user classes should interact within the system and with each other. **Figure 5-13, Future Digital System User Classes**, illustrates this conceptual model. The Administration/Operations Manager User and relevant capabilities are embedded in all components of the proposed system.

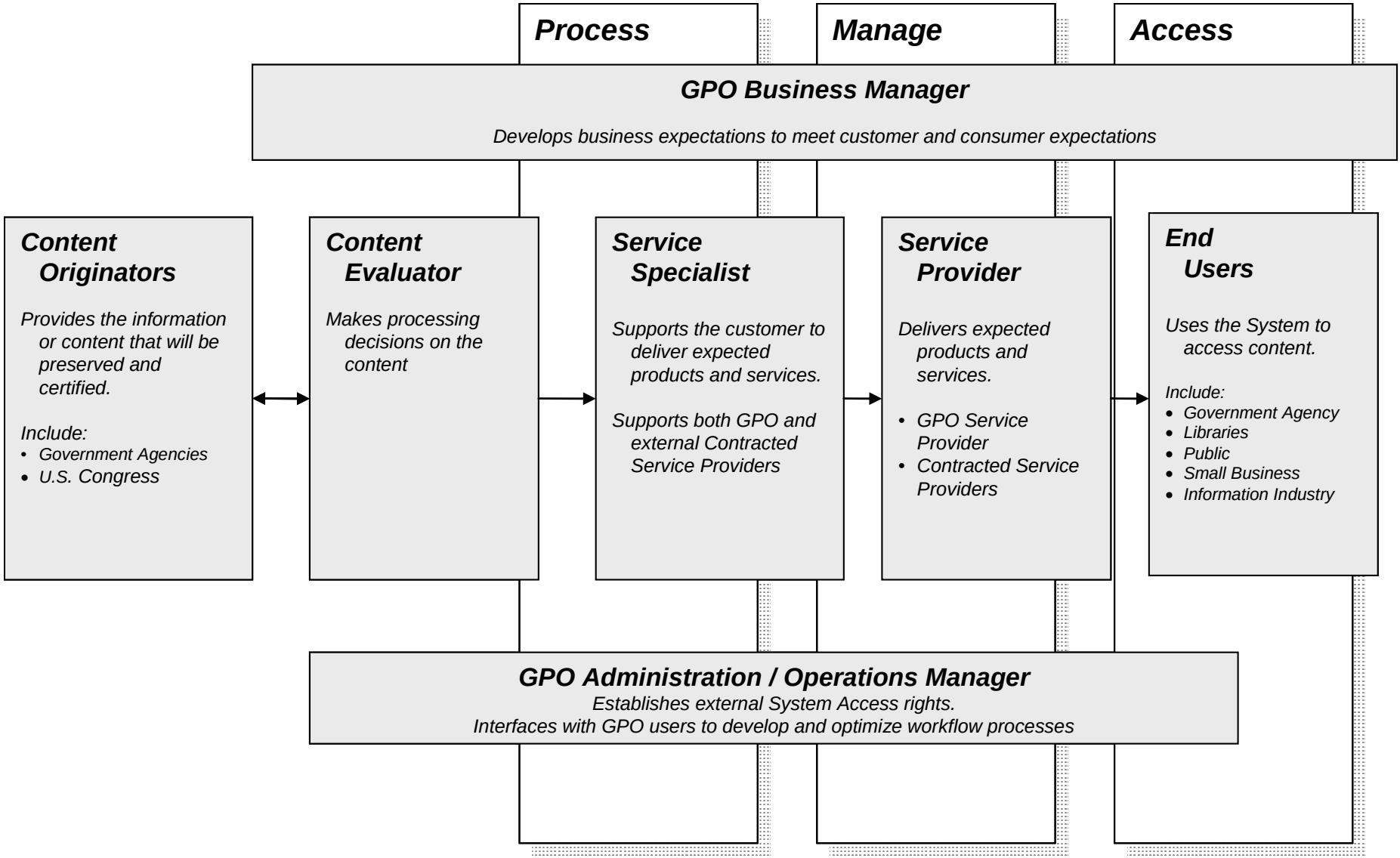
5.5.4 Other Involved Personnel. GPO will be developing an organizational structure for overseeing the Future Digital System program. When this information becomes available it will be included in a future update to the *Future Digital System ConOps.*

5.6 SUPPORT ENVIRONMENT

The support environment will not be determined until the conclusion of the systems analysis and design phase of the Future Digital System program. However, if another update to the *Future Digital System ConOps* is determined to be of value at that time, the required information will be provided.

FINAL

Figure 5-13. The Future Digital System User Classes



6 OPERATIONAL SCENARIOS

The *Future Digital System ConOps* document expresses what users want and envision in the proposed system. Scenarios convey these needs in simple non-technical language. Overlap occurs between different scenarios as a result of interaction between different users or due to similarity between different activities. All of the scenarios represented in the following sections describe one example of how users may interact with the system. Scenarios have purposely been made to be far reaching in an attempt to include all possible actors within a designated class (of users), but the scenarios are not intended to identify all possible situations for any given user class. Additionally, the steps in the scenarios should not be interpreted as a fixed sequence of events; instead they should be interpreted as an illustration of capabilities the system will offer (any user class).

A scenario is a step-by-step description of how the system should operate and interact with both its users and external interfaces under a given set of circumstances. Scenarios are described in a manner that enables readers to walk through them and gain an understanding of how all the principal parts of the system function and interact. The scenarios tie together all parts of the system, the users, and other entities by describing how they interact. Scenarios cover the user's concept of all the operational modes and all classes of users identified for the proposed system and illustrate all the business processes that the system will support.

6.1 CONTENT ORIGINATOR SCENARIO

Content Originators are comprised of executive or judicial employees primarily consisting of authors, editors, and/or publication creators (Agency Customer), and legislative or congressional employees who manage the information assets of the Congress (Congressional Customer). Content Originators are responsible for making content available to GPO for certification, preservation and dissemination. In some cases, content is created without using the Future Digital System, and/or without the knowledge of GPO, in those cases, the Content Evaluator may make the decision to use harvesting tools to begin the ingest of that content.

Content Creation

- Content Originator develops/creates content.
- Content Originator generates and submits order requesting GPO services.
- The system provides ingest aides for content creation, creative services, content management, and content validation, as applicable.
- Content Originator develops content and initiates workflow through the system ingest aid that includes best practices guidelines for the creation and submission of the Content Package.
- Content Originator collaborates with GPO through the system to provide preprocessing information and specifications.
- Content Originator coordinates with Content Evaluator on the content service orders and the Preservation and Dissemination Plan and makes them available to Service Specialists and Service Providers.
- When GPO content creation services are fee-bearing, the system interfaces with the fee management system to determine fees and charge the Content Originator.

Content Validation and Ingest

- For deposited content ingest, text and/or other information is captured, managed, provided creative services as appropriate, and validated before the ingest toolset creates the SIP according to GPO-established best practices.

- For converted content ingest, converted content is made available, then is captured, managed, and validated before the ingest toolset creates the SIP according to GPO-established best practices.
- For harvested content ingest, content is harvested, then is then is captured, managed, and validated before the ingest toolset creates the SIP according to GPO-established best practices.

6.2 CONTENT EVALUATOR SCENARIO

The Content Evaluator may work with the Content Originator to assist in content development and determines what processing occurs. Once the content is made available to GPO, the Content Evaluator makes decisions regarding scope and preservation of content. The Content Evaluator may also use tools to coordinate harvesting and format conversion of content. Service orders are pushed to Service Providers by the Content Evaluator.

Preprocessing of Content

- When requested, uses the system and the system tools to work with Content Originators to develop content that conforms to GPO standards.

Version, Scope, and Certification

- Determines if content is within the scope of GPO.
- The Content Evaluator then determines if the document is in scope for GPO Dissemination Program. Working with a Business Manager determines if the document is in scope for the sales program. An example of GPO dissemination program is the FDLP.
- The Content Evaluator uses the system to request certified content from the Service Specialist.
- The Content Evaluator determines if the content is new or a version of previously delivered content and confirms this with the Content Originator as appropriate.
- The system applies storage management rules as appropriate for the content.
- The system maintains audit trails that document the location of the material.
- Based on scope and input from Content Originators, the Content Evaluator uses the system to develop the Preservation and Dissemination Plan.

Harvesting and Format conversion

- The Content Evaluator uses the system locating tools to locate content that is within scope.
- When content is located, coordinates the harvesting by either using harvesting tools or creating a service order for Service Providers to use the harvesting tools to gather the content and its related metadata.
- Content is packaged for ingest into the system.
- As required by an Operations Manager or by business rules, the Content Evaluator sends service orders to Service Providers to use the system tools to digitize existing tangible content. When digitization is complete the Content Evaluator packages that content for ingest.

Service Ordering

- The Content Evaluator determines if the service should be done by GPO or by an external Service Provider.
- Service orders are sent to the Service Provider.

6.3 SERVICE SPECIALIST SCENARIO

After content arrives in the system and the Content Evaluator has determined its preservation status, the Service Specialist performs the following actions as needed to perform preservation tasks on content for as long as necessary.

Provides Special Services

- Uses the system to work with Content Originators to track and resolve disputes with Service Providers.
- Works with Content Originators to assist in content creation by providing expert advice.
- Works with Business Managers to repurpose content. This includes repackaging of content for new business needs, such as determining what digital content should be made available in hardcopy. For those documents that should be made available in hardcopy, a service order is sent to the Service Provider.

Metadata development

- The system populates initial metadata information package (e.g., bibliographic record) for the content. The Service Specialist confirms or updates that catalogue and index entry.
- If this is a version of existing content, the system documents the necessary relationships between this version and previous ones.
- The Service Specialist reviews that metadata and updates it appropriately.
- The system performs conformity checking to ensure that metadata has been recorded, and meets appropriate standards.

Preservation Processing

- When content arrives in the system, and after a preservation and dissemination plan has been established, the system executes any preservation processing required by that plan.

- The Service Specialist uses the system to review the original content and the processing applied to them to determine if preservation objectives are being achieved effectively and consistently.

- When problems occur in executing a preservation and service plan, the Service Specialist determines whether the exceptions should be accepted and documented “as is.” Alternatively, the Service Specialist works with the Content Evaluator or Content Originator to determine appropriate corrective action or to modify the Preservation and Dissemination Plan. The system maintains an audit trail of activity.

- The Service Specialist ensures that the system captures and retains information about the digital content necessary to ensure its preservation, accessibility, and to certify whether it’s authentic and/or official. The system will provide appropriate tools, techniques, and methods to enable faithful reproduction of all digital content in the system.

- When new content has been preserved, electronic versions will be available to the depository libraries through the system.

Content Maintenance

- The Service Specialist uses the system to examine samples of digital content being preserved to ensure that nothing is lost or corrupted in storage. In the event of corrupted content (e.g. either due to

media degradation or media migration problems), the Service Specialist uses the system tools to assist in the recovery of the content.

- The Service Specialist works with administrative users to ensure that necessary changes, such as media migration, are implemented in the storage system. The Service Specialist reviews plans for, monitors, and evaluates updates or modifications of the storage system, including migration of digital content to new digital media.

Creation of the Collection of Last Resort

The system directs content to the appropriate storage location.

6.4 SERVICE PROVIDER SCENARIO

Service Providers provide the system with their capabilities and pricing information, receive notification of service orders, accept service orders, and provide services.

Service Capability Update

- The Service Provider updates the system with their capabilities.
- The Service Provider updates the system with pricing information.

Providing Service

- For each service order, the Service Provider receives notification that an order is coming.
- External providers use the system to accept or deny orders and to provide quotations. When possible, the system provides quotations automatically to the service requestor based on capability and pricing information already available within the system.
- The system notifies Service Providers when jobs are ready for initiation.
- The system tracks the progress of orders and their completion.
- The Service Provider performs the service and notifies the system of status and completion.

6.5 BUSINESS MANAGER SCENARIO

The Business Manager is responsible for determining what content might be made available for sale and setting pricing levels for GPO services.

Repurposing Content

- Based on knowledge of content, the system provided data on content usage, and other business information, the Business Manager decides what content should be made available for sale. This may involve repurposing various content into documents or publications that will be made available for sale. It may also involve making non-tangible content into tangible publications.
- When a decision is made to make new content available, the Business Manager collaborates with the Service Specialist (via the system) to determine how to repurpose the content.

Set Pricing Levels

- Based on business knowledge and the system data, the Business Manager uses the system tools to set pricing levels for GPO services and products.

6.6 ADMINISTRATION/OPERATIONS MANAGER SCENARIO

Administration/Operations Manager users handle such activities as assigning user rights and privileges, scheduling reports, monitoring the system, modifying workflow, and ensuring system availability. This scenario is included to demonstrate some of the capabilities that would be included in the system for the administrative user of the system as well as the Operations Manager of the system. Not all capabilities are described in the scenario and many of the system functions will be done without user involvement.

Assign user rights and privileges

- Using GPO predefined roles (which includes information regarding permissions granted and job roles), the administrative user creates the user account establishing requested access rights and privileges in the system (i.e., user profile is created). The user is granted appropriate access rights (e.g., access to data that may be restricted by certain access privileges or administrative access) and systems capabilities (e.g., ability to edit, input data, check security, produce user reports). Note that not all users will require accounts.

Schedule Reports

- This user logs on to the system and uses any data available in the system to create new reports or modify existing reports. The request for reports could be based on a specific requirement from GPO or from a system monitoring need. The reports could provide metric data for such activities as system usage, system capacity, performance, and workflow statistics.
- The reports are scheduled for regular distribution to the appropriate people or are created on as needed basis.

Monitor System

- The system provides this user with the ability to monitor system performance and security using system tools. These tools provide for monitoring storage, performance, space, load, security-related indicators, etc.
- This user, with help from support staff, diagnoses and troubleshoots problems implementing intrusion detection systems and virus control procedures. In parallel, the system is recording these events in system logs and establishing an audit trail.
- Once the problem has been corrected the user ensures that the system's operations are secure from intrusion, viruses, unauthorized access, etc.

Modify Workflow

- In some instances this user will be able to modify workflow. The user will be able to modify work flowing through the system at a point in time when problems within the system arise.
- When this user is alerted to a potential problem within the system (e.g., a problem with a server has occurred), the user notifies the appropriate support staff who diagnoses and troubleshoots the problem, and temporarily modifies system workflow(s) to ensure continued service.
- This user notifies the appropriate operation of the temporary modification to workflow. The user tracks the resolution of the problem for audit trail purposes and the modified system workflow(s) will exist in the system until the problem can be corrected.
- When notified by the system that identified steps are not occurring as scheduled, this user has the capability to examine the system in an attempt to understand and/or determine the nature of the problem. Possible problems could be related to bottlenecks in the system or due to inability of the Content Evaluator, Service Specialist or Service Providers to complete tasks.

- This user may recommend possible solutions (if due to a bottleneck in the system) or interface with GPO staff to determine the nature of the problem and recommend solutions.
- This user has the authority to implement an agreed upon solution in order for the tasks to continue.

Job Pending

- This user logs onto the system and receives a notification from the system that a specific task is ready to be performed. The system, using predefined GPO business process rules is able to determine what activities need to occur. Based on these rules, the system can decide to create a task, assign tasks to staff, assign due dates, and provide relevant information about the task.

Review System Assignments

- This user reviews the assignments identified by the system and selects from the options that are presented:
    - Confirm Assignments
        - Upon confirmation the system notifies staff of their assignments including milestones and begins to track the task, which includes capturing performance statistics.
        - As the task proceeds, the system is able to send notifications, collect approvals, detect when processing has been suspended, make additional assignments, or notify this user that the job is complete.
    - Modify Assignments
        - Upon inspection of the task, the manager has the capability to modify the steps, adding or removing steps, or changing the order of the steps to be performed to process the job as a candidate workflow.
        - The system will either confirm the modification or may determine that additional steps are necessary requiring this user to make additional modifications.
        - Upon approval, notifications are sent to staff alerting them of their assignments.

Approval and Closure of Tasks

- As the task progresses through the system, there are various junctures when approval may be required by this user. The manager will inspect tasks on a periodic basis and provide approval as appropriate, including final approval that the job has successfully been completed.
- Upon final approval, the system captures this information and stops tracking the job.

6.7 END USER SCENARIO

The End User will undertake the following steps in using the system to obtain content. (The steps listed below should not necessarily be interpreted as a sequence of events.)

Search

- All End Users will have the ability to search for and access content within the system. Access to content may be dependent on End User rights and privileges.
- The End User searches descriptive metadata and/or content within documents. Within the End User's given access rights and privileges, the End User may use available functions and features. The system provides the capability for the End User to view and/or sort the results of the search, modify the search, and refine or save search results. If needed the user may interact with GPO to receive help.

Retrieve/Receive Content

- From search results that identify relevant content, the system allows the End User to view and access available content. The End User may request the system to deliver content to an available medium.
- If content is located that is not available electronically, the system provides the End User with bibliographic and location information, and options for accessing the tangible content.

Assisted Access

- The End User may request help from GPO while using the system. Assisted Access may include such activities as answering questions, conducting and handling searches, processing special requests, expediting requests, and similar issues. This may also include online help tools or referrals to libraries, etc. the system tracks the communication and information about the Assisted Access. Some of these services may involve fees.

Fee for Service

- End Users may request products or services that require them to pay a fee. If a fee must be collected, the system tracks, reports upon, and routes any required financial transaction information to all appropriate billing/accounting systems, and provides the requested product or service on authorization by the billing/accounting system.

6.8 OTHER

There are no additional user scenarios at present. If additional users are identified, their interaction with the system will be provided in a future update of the *Future Digital System ConOps*.

7 SUMMARY OF IMPACTS

Implementation of the Future Digital System may have wide ranging impacts on both GPO and its customers. The subsections below identify potential operational impacts, organizational impacts, and impacts during development that should be considered as GPO develops its plans.

7.1 OPERATIONAL IMPACTS

Until the Future Digital System undergoes systems analysis and design, operational impacts of the proposed system are not known; therefore, impacts to the following have been omitted.

- Interfaces with primary or alternate computer operating centers
- Changes in procedures
- Use of new data sources
- Changes in quantity, type, and timing of data to be input into the system
- Changes in data retention requirements
- New modes of operation based on emergency, disaster, or accident conditions
- New methods for providing input data if the required data are not readily available
- Changes in operational budget
- Changes in operational risks

However, it is anticipated that GPO as an organization will have to implement changes to the way it conducts business in order to achieve the agency's mission, goals, and objectives. The system will facilitate this endeavor. When implemented, the system will be able to handle vastly different content in a variety of formats that GPO has been incapable of addressing in the past.

The system should provide decision support for GPO management processes for the content management lifecycle processing of all content. This additional capability includes supporting processes for such activities as evaluating, creating descriptions of and metadata for, and preserving content of all types. Additional operational impacts may include the following items:

- Data architecture modeling;
- Disaster or catastrophic recovery;
- Advances in technology; and
- Changes to operational procedures.

7.2 ORGANIZATIONAL IMPACTS

GPO is examining current policies and business practices and may have to develop and/or modify policies and business practices as necessary. The depth and breadth of the organizational impact is unknown at this time. Information with respect to the following has not been provided for this reason and includes such items as the number of personnel, skill levels, position identifiers, and locations of personnel. Additionally, the interaction of personnel with the system may necessitate revising position descriptions to reflect the anticipated changes in GPO's business practices.

With this in mind, GPO has identified a number of possible organizational impacts, as described below:

- An assessment of how the system will fit organizationally within GPO and how the system will relate to other GPO components and/or the agencies they interact with;
- The commitment of resources (e.g., funding, time, staff) by GPO to establish working relationships with other Government agencies as the system is rolled out;

- The need for cross-functional, inter-disciplinary staff teams;
- The development of education and increased training for both GPO staff and End Users;
- The need for additional personnel for the creation and maintenance of a help desk facility for GPO staff and End Users; and
- Improved opportunities for career development for GPO staff.

7.3 IMPACTS DURING DEVELOPMENT

The full extent of impacts during development will not be known until completion of the systems analysis and design phase, and, as such, this information has not been provided; however, impacts considered thus far include:

- Articulation of business rules and other controls needed for operational implementation;
- Development of training for requirements to be implemented in an increment; and
- Training necessary for rollout of the increment.

When known, information on impacts such as the following will be provided as required:

- Involvement in studies, meetings, and discussions prior to award of the contract;
- User Support and involvement in reviews and demonstrations, evaluation of initial operating capabilities and evolving versions of the system, development or modification of databases, and required training;
- Parallel operation of the new and existing systems; and
- Operational impacts during system testing of the proposed system.

8 ANALYSIS OF THE PROPOSED SYSTEM

Various improvements, disadvantages and limitations, and alternatives and trade-offs considered are covered in this section.

8.1 SUMMARY OF IMPROVEMENTS

The system, when implemented, will subsume and enhance the existing functionality provided by the legacy systems identified in Section 3.3.1; however, the proposed system will provide a completely new set of capabilities as described in Section 4.2.3. The full extent of the capabilities will not be known until the completion of the systems analysis and design phase of the program and will be addressed at that time as required. It is anticipated, however, that the proposed system will offer numerous benefits to GPO and End Users and may include the following items:

- Consolidated content management lifecycle administration and streamlined internal workflow;
- More involvement with agency customers in the early phases of the content creation and management lifecycle;
- New tools to support processing and evaluation of content:
    - Tools to aid in the ingest process,
    - Tools for preservation and access, and
    - Tools for the creation of descriptions and metadata;
- A wider variety of available content in GPO assets;
- Faster access to content;
- Enhanced capabilities for searching content;
- The ability to service additional End Users;
- Increased responsiveness to End Users;
- Remote access to content; and
- The harvesting, locating and subsequent preservation of fugitive documents that would otherwise be lost.

8.2 DISADVANTAGES AND LIMITATIONS

Disadvantages and limitations related to developing the Future Digital System include:

- High development costs;
- High costs associated with security;
- Staff anxiety brought about by new responsibilities;
- Poor GPO staff morale without proactive change management; and
- Impact on transferring customer agencies (resources required to prepare for transfer of materials to GPO, greater content management responsibilities).

8.3 ALTERNATIVES AND TRADEOFFS

Alternatives to the Future Digital System are TBD.

9 NOTES

The following sections provide acronyms and definitions of terms used herein.

9.1 ACRONYMS

The technical terms used in this document are defined in IEEE Std 610.12-1990*, IEEE Standard Glossary of Software Engineering Terminology.* **Table 9-1, Acronyms**, provides a list of acronyms used herein.

ACRONYM	DEFINITION
ABLS	Automated Bid List System
ACES	Access Certificates for Electronic Services
ACSIS	Acquisition, Classification, and Shipment Information System
AIP	Archival Information Package
AP	Access Processor
ARK	Archival Resource Key
ASCII	American Standard Code for Information Interchange
BAC	Billing Address Code
BPI	Business Process Information
CA	Certification Authority
CCSDS	Consultative Committee for Space Data Systems
CD	Compact Disk
CD-ROM	Compact Disk Read Only Memory
CE	Content Evaluator
CFR	Code of Federal Regulations
CGP	Catalog of U.S. Government Publications
CMS	Content Management System
CP	Content Processor
CPI	Content Packet Information
CSV	Comma Separated Variable
DARD	Departmental Account Representative
DIP	Disposition Information Package
DO	Digital Objects
DOI	Digital Object Identifier
DoS	Denial of Service
DPI	Dots Per Inch
DVD	Digital Versatile Disc
ePub	Electronic Publishing Section
FAQ	Frequently Asked Question
FBCA	Federal Bridge Certificate Authority
FDLP	Federal Depository Library Program
FIFO	First In First Out
FOIA	Freedom of Information Act
FTP	File Transfer Protocol
GAO	General Accounting Office
GILS	Government Information Locator System
GPEA	Government Paperwork Elimination Act
GPO	Government Printing Office
HTML	Hypertext Markup Language

ACRONYM	DEFINITION
Hz	Hertz
ID	Information Dissemination
IEEE	Institute of Electronics and Electrical Engineers
ILS	Integrated Library System
ISO	International Organization for Standardization
IT	Information Technology
JDF	Job Definition format
LOC	List of Classes
LPI	Lines Per Inch
MARC	Machine Readable Cataloging
MOCAT	Monthly Catalog of Government Publications
MPCF	Marginally Punched Continuous Forms
NARA	National Archives and Records Administration
NB	National Bibliography
NET	New Electronic Titles
NFC	National Finance Center
NIST	National Institutes of Standards and Technology
NLM	National Library of Medicine
OAIS	Open Archival Information Systems
OCLC	Online Computer Library Center
OCR	Optical Character Recognition
PCCS	Printing Cost Calculating System
PDA	Personal Data Assistant
PDF	Portable Data Format
PDI	Preservation Description Information
PICS	Procurement Information and Control System
PKI	Public Key Infrastructure
POD	Print On Demand
PPR	Printing Procurement Regulation
PURL	Persistent URL
RI	Representation Information
ROI	Return on Investment
RPPO	Regional Printing Procurement Office
SF	Standard Form
SIP	Submission Information Package
SGML	Markup Language
SMP	Storage Management Processor
SMS	Storage Management System
SPA	Simplified Purchase Agreement
SuDocs	Superintendent of Documents
URL	Uniform Resource Locator
U.S.C.	United States Code
USGPO	United States Government Printing Office
WAIS	Wide Area Information Servers
WAP	Wireless Application Protocol
WML	Wireless Markup Language
XML	eXtensible Markup Language

9.2 GLOSSARY

Terms used herein are defined to promote readability of the document.

Access: Tools and processes associated with finding, analyzing, ordering, and retrieving CPI or BPI.

Access aids: Tools and processes associated with finding, analyzing, retrieving, and ordering CPI or BPI.

Access Content Package (ACP): The result of ingest processing; i.e., validation, authentication, version control, transformation, verification of scope, validation or assignment persistent name, and metadata generation/capture.

Access (or service) copy: A digital publication whose characteristics (for example a screen-optimized PDF file) are designed for ease or speed of access rather than preservation.

Accessibility: Making tools and content available and usable for all users including those with disabilities; the degree to which the public is able to retrieve or obtain Government publications, either through the FDLP or directly through an digital information service established and maintained by a Government agency or its authorized agent or other delivery channels, in a useful format or medium and in a time frame whereby the information has utility.

Activity: A description of a piece of work that forms one logical step within a process. An activity may be a manual activity, which does not support computer automation, or a workflow (automated) activity. A workflow activity requires human and/or machine resources(s) to support process execution.

Application Security: The protection of application data and systems against unauthorized access to or modification of information, whether in storage, processing or transit, and against the denial of service to authorized users or the provision of service to unauthorized users, including those measures necessary to detect, document, and counter such threats at the application level. See also Security.

Archival information package (OAIS): Content information and its associated PDI needed to preserve the content over the long term, bound together by packaging information.

Archive: A collection with related systems and services, organized to emphasize the long-term preservation of information.

Archive management - See Preservation.

Authentic: Describes content that is verified by GPO to be complete and unaltered when compared to the version approved or published by the Content Originator.

Authentication: Validation of a user, a computer, or some digital object to ensure that it is what it claims to be. In the specific context of the Future Digital System, the assurance that an object is as the author or issuer intended it. See also Certification.

Authenticity: A digital publication's identity, source, ownership and/or other attributes are verified. Authentication also connotes that any change to the publication may be identified and tracked.

Availability: The degree to which information is obtainable through an intentional or unintentional provision of information and services.

Born digital: In the Future Digital System context, digital objects, created in a digital environment, with the intention of multiple eventual output products, potentially including hard copy, electronic presentation, and digital media. Born digital object will exist in an entirely digital lifecycle; relating to a document that was created and exists only in a digital format.

Browse: To explore a body of information on the basis of the organization of the collections or by scanning lists, rather than by direct searching.

Business process: A set of one or more linked activities which collectively realize a business objective or policy goal, normally within the context of an organizational structure defining functional roles and relationships.

Business process information: Administrative information, non-content specific information that is used within the business process and package description (PD) to support access aids and data mining.

Cataloging and indexing: Cataloging is comprised of the processes involved in constructing a catalog: describing information or documents to identify or characterize them, providing "entry points" (terms) peculiar to the information or document, e.g., author, title, subject, and format information, by which the information can be located and retrieved. The immediate product of cataloging is bibliographic records, which are then compiled into catalogs. Indexing is the process of compiling a set of identifiers that characterize a document or other piece of information by analyzing the content of the item and expressing it in the terms of a particular system of indexing. In GPO context, cataloging and indexing is the statutory term for the processes that produce the Catalog of U.S. Government Publications and its indexes. In the FDSys context, the process or results of applying bibliographic control to final published versions. Certification: Proof of verification or authority. Process associated with ensuring that a digital object is authentically the content issued by the author or issuer. The "certificate" is a mark of veracity which is in some way joined to the object itself.

Certified: Providing proof of verification of authenticity or official status.

Collaboration: Allowing for multiple authors or content sources while maintaining digital asset and document control and provenance.

Collection of Last Resort – See National Collection of U.S. Government Publications

Collection plan, or Collection management plan: The policies, procedures, and systems developed to manage and ensure current and permanent public access to remotely accessible digital Government publications maintained in the National Collection.

Compose: The ability to style/format content

Composition: Creating content using FDsys applications.

Content: Information presented for human understanding.

Content Analysis: Interpretation of intended context.

Content Information (OAIS): The set of information that is the primary target for preservation, composed of the data object and it's RI.

Content Package Information (CPI): Information that directly relates to the content and is ultimately used in the dissemination and preservation of the content to the end users.

Converted content: Digital content created from a tangible publication.

Cooperative Publication: Publications excluded from GPO's dissemination programs because they are produced with non-appropriated funds or must be sold in order to be self-sustaining. See 44 USC 1903.

Dark archive (digital): The site or electronic environment wherein a second "copy" or instance of all master and derivative digital files, data, and underlying enabling code resides and is maintained, under the control of the managing organization or its proxy. The dark archive must be inaccessible to the general public. Access to the dark repository contents and resources ("lighting" the archive) is triggered only by a specified event or condition.

Dark archive (tangible): A collection of tangible materials preserved under optimal conditions, designed to safeguard the integrity and important artifactual characteristics of the archived materials for specific potential future use or uses. Eventual use of the archived materials ("lighting" the archives) is to be triggered by a specified event or condition. Such events might include failure or inadequacy of the "service" copy of the materials; lapse or expiration of restrictions imposed on use of the archives content; effect of the requirements of a contractual obligation regarding maintenance or use; or other events as determined under the charter of the dark archives.

Data mining: Discovery method applied to large collections of data, which proceeds by classifying and clustering data (by automated means) often from a variety of different databases, then looking for associations. Specifically applied to the analysis of use and user data for GPO systems, data mining includes the tools and processes for finding, aggregating, and associating BPI to enhance internal and external business efficiencies.

Deposited content: Content received from content originators in digital form.

Derivative: A new presentation of existing content optimized for access. This does not include language translation.

Device: Content delivery mechanisms for digital media, such as data storage devices (e.g., CD, DVD, etc.), wireless handheld devices, future media, and storage at user sites.

Digital media: An intermediary mechanism consisting of data storage devices to deliver content to users' storage or display devices.

Digital object: An item stored in a digital library or other digital collection of information, consisting of data, metadata, and an identifier.

Digital signature: A cryptographic code consisting of a hash, to indicate that data has not changed, encrypted with the public key of the creator or the signature.

Dissemination: The transfer from the stored form of a digital object in a repository to the client or user.

Dissemination information package (DIP): An information package that contains parts of all or one or more archival information packages, to be distributed to the user or consumer as requested, or to service providers to produce various outputs.

Distribution: Applying GPO processes and services to a tangible publication and sending a tangible copy to depository libraries.

Document: A digital object that is the analog of a physical document, especially in terms of logical arrangement and use.

Draft: A preliminary version of content, not yet in its finalized form.

Dynamically Changed Workflow: Workflow process that is changed during executing.

Electronic presentation: The dynamic and temporary representation of content in digital format; strongly dependent upon file format and user's presentation device

Emulation: Replication of a computing system to process programs and data from an earlier system that is no longer is available.

Existing digital: In GPO's current situation, publications or digital objects which are produced solely for digital dissemination, such as documents on agency web sites for which there is no printed equivalent. Faithful digital reproduction: Digital objects that are optimally formatted and described with a view to their quality (functionality and use value), persistence (long-term access), and interoperability (e.g. across platforms and software environments). Faithful reproductions meet these criteria, and are intended to accurately render the underlying source document, with respect to its completeness, appearance of original pages (including tonality and color), and correct (that is, original) sequence of pages. Faithful digital reproductions will support production of legible printed facsimiles when produced in the same size as the originals (that is, 1:1).

FDLP Electronic Collection, or EC: The digital Government publications that GPO holds in storage for permanent public access through the FDLP, or are held by libraries and/or other institutions operating in partnership with the FDLP. These digital publications may be remotely accessible online publications, or tangible publications such as CD-ROMs maintained in depository library collections.

FDLP partner: A depository library or other institution that stores and maintains for permanent access segments of the Collection.

Final Published Version: Content in a specific presentation and format approved by its Content Originator for release to an audience. (See also Government Publication; Publication).
Fixity: the quality of being unaltered (e.g. "fixity of the text" refers to the durability of the printed word).
Format: In a general sense, the manner in which data, documents, or literature are organized, structured, named, classified, and arranged. Specifically, the organization of information for storage, printing, or display. The format of floppy disks and hard disks is the magnetic pattern laid down by the formatting utility. In a document, the format includes margins, font, and alignment used for text, headers, etc. In a database, the format comprises the arrangement of data fields and field names.

Format management - See Preservation.

Fugitive document: A U.S. Government publication that falls within the scope of the Federal Depository Library Program (FDLP), but has not been included in the FDLP. These publications include tangible products such as ink-on-paper, microforms, CD-ROM, or DVDs. Fugitive documents most commonly occur when Federal agencies print or procure the printing of their publications on their own, without going through GPO.

Government publication: A work of the United States Government, regardless of form or format, which is created or compiled in whole or in part at Government expense, or as required by law, except that which is required for official use only, is for strictly operational or administrative purposes having no public interest or educational value, or is classified for reasons of national security.

Granularity: The degree or level of detail available within content in the system

Granularity policy: The system shall have the ability to certify related or continuous piece of content in context

Handle System: A comprehensive system for assigning, managing, and resolving persistent identifiers, known as "handles," for digital objects and other resources on the Internet. Handles can be used as Uniform Resource Names (URNs).

Hard copy: Tangible printed content.

Harvest: The gathering and capture of content resident on official Federal Government Web sites that falls within the scope of GPO dissemination programs.

Harvested content: Digital content within the scope of dissemination programs that is gathered from Federal agency Web sites.

History: A record of all system activities.

Information granularity: The degree or level of detail available in an information system. With reference to authentication, the level of detail or specificity (e.g., page, chapter, paragraph, line) to which veracity can be certified.

Ingest (OAIS): The OAIS entity that contains the services and functions that accept SIPs from Producers, prepare Archival Information packages for storage, and ensure that information packages and their supporting descriptive information packages are established within OAIS. In the FDSys, ingest processing includes validation, authentication, version control, transformation, verification of scope, validation or assignment persistent name, and metadata generation/capture.

Integrity Mark: Emblem that is used to convey authentication information to users. The mark may be visible or invisible, and all content delivery methods should have associated marks.

Interoperability: Compatibility of workflow across standards (e.g., WFMC to BPEL) and, compatibility of workflow within a standard and across programming languages (e.g., Java and C++ working in WFMC).

Item: A specific piece of material in a digital library or collection; a single instance, copy, or manifestation.

Job: An instance that will result in a product or service supplied by the system.

Light archive: A collection of tangible materials preserved under optimal conditions, designed to safeguard the integrity and important artifactual characteristics of the archived materials while supporting ongoing permitted use of those materials by the designated constituents of the archives. A light archive normally presupposes the existence of a dark archive, as a hedge against the risk of loss or damage to the light archives content through permitted uses. A light archive is also distinct from regular collections of like materials in that it systematically undertakes the active preservation of the materials as part of a cooperative or coordinated effort that may include other redundant or complementary light archives.

Localized Presentation: Temporary representation of layout or structure on a user's local presentation device.

Locate (discover): The organized process of finding Web-based documents or publications that are within scope for a particular collection.

Manage: In Information Technology contexts, to add, modify, or delete content.

Manifestation: Form given to an expression of a work, e.g., by representing it in digital form.

Message: Communication between a process and the Workflow Management System.

Metadata: Metadata is a structured representation of information that facilitates interpretation, management, and location by describing essential attributes and significant properties. Metadata describes the content, quality, condition, or other characteristics of other data. Metadata describes how, when, and by whom information was collected, where it resides, and how it is formatted. Metadata helps locate, interpret, or manage. In current usage several types of metadata are defined: descriptive, which aids in locating information; structural/technical, which records structures, formats, and relationships; administrative, which records responsibility, rights, and other information for managing the information; and preservation, which incorporates elements of the other types specific to preserving the information for the long term.

METS (Metadata Encoding and Transmission Standard): Essentially a standard DTD (document type definition) for interpreting XML as metadata.

Migration: Preservation of digital content where the underlying information is retained but older formats and internal structures are replaced by newer.

Modified Workflow: Workflow process that is changed during process development, or not at runtime.

National Collection of U.S. Government Publications (NC): A comprehensive collection of all in-scope publications, content that should be (or should have been) in the FDLP, regardless of form or format. The NC will consist of multiple collections of tangible and digital publications, located at multiple sites, and operated by various partners within and beyond the U.S. Government.

No-fee access: There are no charges to individual or institutional users for searching, retrieving, viewing, downloading, printing, copying, or otherwise using digital publications in scope for the FDLP.

Notification: A message in Workflow between a process and the WMS that indicates when an identified event or condition, such as an exception, has been met.

OAIS (Open Archival Information System Reference Model) (ISO 14721:2003): A reference model for an archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a designate community. The model defines functions, activities, responsibilities, and relationships within this archive, sets forth common terms and concepts, and defined component functions which serve as the basis for planning implementation.

Official: A version that has been approved by someone with authority.

Official content: Content that falls within the scope of the FDLP EC and is approved by, contributed by, or harvested from an official source in accordance with accepted program specifications

Official source: The Federal publishing agency, its business partner, or other trusted source.

ONIX (Online Information eXchange): A standard format that publishers can use to distribute electronic information about their books to wholesale, e-tail and retail booksellers, other publishers, and anyone else involved in the sale of books.

Online: A digital publication that is published at a publicly accessible Internet site.

Online dissemination: Applying GPO processes and services to an online publication and making it available to depository libraries and the public.

Permanent Public Access, or PPA: Government publications within the scope of the FDLP remain available for continuous, no-fee public access through the program.

Persistent Name: Provides permanence of identification, resolution of location, and is expected to be globally (e.g., internationally) registered, validated, and unique

Preliminary Composition: Preparatory representation of content format or structure

Presentation Device: A device that can present content for comprehension

Preservation: The activities associated with maintaining publications for use, either in their original form or in some verifiable, usable form. Preservation may also include creation of a surrogate for the original by a conversion process, wherein the intellectual content and other essential attributes of the original are retained. For digital materials, preservation includes the management of formats of information (including possible migration to newer versions), the storage environment, and the archival arrangement of information to facilitate preservation.

Preservation description information (OAIS): Information necessary for adequate preservation of content information, including information on provenance, reference, fixity, and context.

Preservation master: A copy which maintains all of the characteristics of the original publication, from which true copies can be made.

Preservation master requirement: A set of attributes for a digital object of sufficient quality to be preserved and used as the basis for derivative products and subsequent editions, copies, or manifestations. Requirements for use, users, and state/condition/format of the source of the original object need to be noted.

Preservation processes: Activities necessary to keep content accessible and usable, including Migration, Refreshment, and Emulation.

Print on demand (POD): Hard copy produced in a short production cycle time and typically in small quantities.

Process: A formalized view of a "business process", represented as a coordinated (parallel and/or serial) set of process activities that are connected in order to achieve a common goal.

Provenance: The chain of ownership and custody which reflects the entities that accumulated, created, used, or published information. In a traditional archival sense, provenance is an essential factor in establishing authenticity and integrity.

Publication: (N) Content approved by its Content Originator for release to an audience.
See also Government publication.

Reference tools: Finding aids, bibliographies, and other services to assist in the locating and use of information, often less formally organized than catalogs and indexes.

Refreshment: A preservation process for data extraction, cleaning and integration, and the triggering events of these activities.

Render: To transform digital information in the form received from a repository into a display on a computer screen or other presentation to a user.

Replication: Make copies of digital material for backup, performance, reliability, or preservation.

Repository: A computer system used to store digital collections and disseminate them to users.

Requirements: In system planning, a requirement describes what users want and expect according to their various needs. Requirements draw a comprehensible picture to facilitate communications between all stakeholders in the development of a system, and outline the opportunities for development of successful products to satisfy user needs.

Rich media: An electronic presentation incorporating audio, video, text, etc.

Rider: Request by GPO, agency, or Congress that adds copies to a Request or C.O. Order placed by a publishing agency or Congress.

Search: Process or activity of locating specific information in a database or on the World Wide Web. A search involves making a statement of search terms and refining the terms until satisfactory result is returned. Searching is distinct from browsing, which facilitates locating information by presenting references to information in topical collections or other logical groupings or lists.

Secondary dark archive (digital): Multiple "copies" or instances of the dark repository, maintained as assurance against the failure or loss of the original dark repository. The secondary dark repository must provide redundancy of content to the original dark repository, and the systems and resources necessary to support access to and management of that content must be fully independent of those supporting the original dark repository content.

Secondary service repository (digital): The secondary service archive is a "mirror" of the service archive, created to provide instantaneous and continuous access to all designated constituents when the access copy or service archive is temporarily disabled.

Security: The protection of systems against unauthorized access to or modification of information, whether in storage, processing or transit, and against the denial of service to authorized users or the provision of service to unauthorized users, including those measures necessary to detect, document, and counter such threats. The measures and controls, including physical controls in conjunction with management, technical and procedural controls, that ensure the confidentiality, integrity and availability of information processed and stored by a system. See also Application Security.

Service archive (digital): The site or electronic environment wherein the derivative, or "use," files and metadata created from source objects (here, tangible government documents), as well as the software, systems, and hardware necessary to transmit and make those files and metadata accessible, are maintained for public display and use. The service repository contains the current and most comprehensive electronic versions of those source materials.

Shared repository: A facility established, governed, and used by multiple institutions to provide storage space and, in some instances limited service for low-use library materials, primarily paper-based materials that do not have to be readily available for consultation in campus libraries.

Status: A representation of the internal conditions defining the state of a process or activity at a particular point in time.

Storage: The functions associated with saving digital publications on physical media, including magnetic, optical, or other alternative technologies.

Storage management - See Preservation.

Sub-versions of content: The state of content within the style tools and prior to ingest.

Submission information package (OAIS): The information package identified by the producer for ingest into an OAIS system.

Subscription: An agreement by which a user obtains access to requested content by payment of a periodic fee or other agreed upon terms.

System: An organized collection of components that have been optimized to work together in a functional whole.

Tangible publication: Products such as ink-on-paper, microforms, CD-ROM, or DVDs, characterized by content recorded or encoded on a physical substrate.

Transformation: The process, or the results of a process, of reformatting or otherwise changing the way content is digitally encoded.

Trusted content: Official content that is provided by or certified by a trusted source.

Trusted source: The publishing agency or a GPO partner that provides or certifies official FDLP content.

Unique Identifier: A character string that uniquely identifies digital objects, content packages and jobs within the system.

User: The person who uses a program, system, or collection of information to perform tasks and produce results.

Validation: A process that ensures data entered into the system conforms to standards for format, content and metadata.

Verification: The process of determining and assuring accuracy and completeness.

Version: Unique manifestation of a publication.

Version control: Relating to a specific manifestation, revision, issuance, or edition of a previously published or issued document or publication. Changes beyond an agreed upon threshold or tolerance constitute a new version. That threshold is a version trigger, and the activity of scanning for changes and activating the trigger is "version control."

Version detection: Activity of inspecting a content package for changes and responding to version triggers. Also, activity of polling the system to identify if an identical version already exists in the system.

Version identifier: Information stored in metadata that identifies version.

Version trigger: Changes beyond an agreed upon threshold or tolerance which constitute a new version.

Work Item: The representation of the work to be processed (by a workflow participant) in the context of an activity within a process.

Workflow: The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules.

Workflow Management System (WMS): A system that defines, creates and manages the execution of workflows through the use of software, running on one or more workflow engines, which is able to interpret the process definition, interact with workflow participants and, where required, invoke the use of IT tools and applications.

Workflow Participant: A resource, human or computer tool/application, which performs the work represented in an activity.

Worklist: A list of "work items" associated with a given workflow participant (or in some cases with a group of workflow participants who may share a common worklist). The worklist forms part of the interface between a workflow engine and the worklist handler.

FINAL

APPENDIX A

DESCRIPTIONS AND CAPABILITIES OF EXISTING SYSTEMS

A.1.1 Descriptions and Capabilities of Existing Systems. The following subsections provide a functional description, including any known limitations, of each legacy system. The high level functionality of these systems is also summarized.

A.1.1.1 Content Ingest. In the context of GPO's function and authority, content is the informational matter received through a chain of responsibility and authority from Federal agencies that GPO processes to the agency's specifications into publications in various formats, with the intention of disseminating it to the public; or informational matter that is discovered and harvested from the Web for preservation or access. A content object is formed by combining content with RI that maps or organizes the data into a useful, meaningful presentation. Thus a page of a Congressional bill is composed of data (in this case text) and the RI that arranges that text into columns with line numbers on the printed page or screen. In the case of a printed bill, the RI arranges the content meaningfully on the page and is fixed in time by the printing process, after which the RI is discarded. The bill presented online depends on its RI being preserved alongside the content to generate a meaningful rendering each time it is invoked.

GPO receives content from agencies in a variety of formats, intended for a variety of output products. This diversity influences the treatment content receives in the course of processing by GPO.

Content may be received by GPO in digital form, structured for hard- or soft-copy output, or in analog form from which digital files for printing are created by GPO staff. Digital inputs may range from structured files intended for producing hardcopy output or web presentations, to minimally structured ASCII text to be loaded into a searchable database. GPO's practice has been to accommodate agency requirements and processes by not limiting input forms.

Content in the current environment is managed on two largely unconnected levels. In the short term, content management consists of the management of jobs in the production process, the organizing principle of which is GPO jacket, literally a jacket or folder onto or into which specifications and other information are recorded. Most of this metadata is acquired or created by GPO and is of limited value, with a short lifespan, ending with the delivery of the completed work or service.

In the longer term, management of content is focused on distribution of publications through several channels: sale, deposit of publications in libraries, and delivery of online publications through *GPO Access* alongside searching and locator services. Metadata supporting these longer term activities is created by GPO and maintained and augmented over time.

A.1.1.1.1 Hard Copy Content. The process of hard copy content ingest is described through the exploration of functions, capabilities, and limitations.

A.1.1.1.1.1 Capabilities, Functions, and Features of the Current System. The current Production and Printing Procurement systems are very flexible regarding the receipt of hard copy content (i.e., GPO accepts everything). However, GPO’s in-plant Production spends an unnecessary amount of extra effort in order to generate output due to a lack of uniformity in what is provided as hard copy. Contracts established with contractors through Printing Procurement frequently incur additional lost time and costs because of the lack of guidance for submitting hard copy.

A.1.1.1.1.2 Operational Environment & Characteristics. Contracting vehicles available to agency customers include:

- The SPA and AgencyDirect: both allow agencies to procure their own printing.
- GPO’s procurement process (Small Purchase and One-Time Bids): a specification is written and bids are solicited.
- Term Contracts: for requirements that will be repeated throughout a specific time period.

- Direct Deal Term Contracts: are term contracts established by GPO but managed by the ordering agency. GPO handles the back-end administrative duties (e.g., billing, quality control, legal).

In-house Production capabilities:

- Whenever Congress is in session, GPO produces the Congressional Record and Congress's daily business (bills, etc.).
- In-house production capabilities are also available to agencies.
- Services are available at GPO for customers who submit hard copy and would like digital files returned to them.
- A POD initiative has been created and implemented with the SuDocs in coordination with Production.

Hard copy input is allowed for all of these production and procurement methods. When it comes to Production, any and all types of hard copy are accepted. GPO does not place any restrictions or requirements on what can be submitted for output. GPO also does not provide any guidelines for submitting hard copy.

In regard to Printing Procurement, the only guidelines provided that indicate how hard copy should be submitted are found in GPO Publication 300.6, "Best Practices for Preparing & Submitting Electronic Design & Prepress Files":

> *Camera Copy: Customers who submit camera copy in lieu of electronic files should follow the following guidelines:*
> - *Output at a minimum of 600 dots per inch (DPI). 1000 DPI or higher is preferred.*
> - *If the document contains photographs, tint screen areas or copy with fine detail, output with a minimum line per inch (LPI) screen of 75 and a maximum line per inch screen of 100 LPI.*
> - *For better results with tints and/or photographs, simply fill the area with 100% black (called a window) and indicate the percentage desired, or the appropriate photograph. A vendor can underlay the appropriate screen tint, or use the window to "drop" in a photograph.*

Use of publication 300.6 (rev. August 2001) is not mandatory, only recommended. The publication is available on line but is not widely utilized. In addition, the most recent revision does not include information about hard copy submission. (The document was created for customers submitting only electronic formats; therefore the reference to hard copy was eliminated.) Therefore, GPO will accept any type of hard copy and does not provide any guidelines to, or place any requirements upon, the customer agency.

A.1.1.1.1.3 Major System Components and High Level Interconnection. The major components involved with hard copy input are:

- Agency customers, who supply hard copy;
- Production, which accepts hard copy;
- Printing Procurement, which writes contracts specifying hard copy input; and
- The contractors affiliated with GPO who accept hard copy.

Within each of these areas are groupings of personnel with different job functions and levels of expertise that have access to different types of hardware and software. In addition, the Production and contractor components possess equipment with varying capabilities.

A.1.1.1.1.4 Interfaces to Systems and Procedures. The current information systems do not make appreciable distinctions between the types of copy submitted (i.e., Camera Copy or Electronic), either for

in-plant production or contracted work. Consequently, there is no way to break down work by types of copy submission. (The limited exception is for work contracted on Small Purchase and One-Time procurements in the central office. For each of these procurement methods, a limited amount of data is captured specific to furnished electronic files.)

A.1.1.1.2 Legacy Scanning. The tangible resources, held in depository libraries since the early days of the FDLP, are the legacy materials that are currently being considered for scanning. This material is in a variety of manifestations, including books, loose-leaf pages, posters, maps, etc., and microfiche, primarily 2nd generation diazo duplicates. Since many libraries have incorporated into their Government documents collections material that was acquired through channels other than the FDLP, there are "fugitive" products being scanned that have no corresponding metadata in GPO's Cataloging and Indexing Program database.

Much of the older material in the tangible collections is in need of preservation and curatorial care. Scanning of the material in the collections will allow wider access to resources and will at the same time provide libraries with the opportunity to reduce the physical volume of their collection.

GPO is working with the library community on a national digitization plan, with the goal of digitizing a complete collection of legacy materials. The initial size of the paper product collection is approximately 2 million documents, estimated at over 60 million pages. The objective is to ensure that the digital collection is available, in the public domain, for no-fee permanent public access through the FDLP. The project will ensure that the collection is digitally reformatted for preservation purposes. The digital preservation masters and the associated metadata will be preserved in GPO electronic archive (in addition to any other places that the materials might be held), and there will be no-fee public access to the content through derivative files on *GPO Access*.

Because the files created by scanning existing printed documents will likely be created in a collaborative environment, and because their preservation and the generation of derivative products for access will also be a shared effort, there is a need for documented specifications for the digitization process and for metadata to be used to record digitization processes and events. A published specification will allow not only for internal quality control but for wide interoperability and the development of best practices for preservation of digital versions.

A meeting was held in March 2004 at GPO that brought together practicing experts in the field of digital format conversion and digital project development to discuss the current standards and specifications for the creation of digital objects for preservation and to propose a set of minimum requirements for digitizing documents for this project. A report from the meeting was produced, which included a set of requirements for the scanning of Federal Government documents to produce digital preservation masters. This report is located at *www.gpoaccess.gov/about/reports/preservation.html*.

An additional meeting was held in June 2004 at GPO that brought together practicing experts on metadata for digital library collections to discuss the current standards and specifications for the creation of metadata for digital objects. While the summary report of the meeting has not yet been completed, there was a perceived need for metadata outside the existing MARC records, including descriptive, administrative, technical, and preservation metadata. Further interchange with the library community is needed before a listing of recommended metadata requirements can be established.

A.1.1.1.2.1 Digitization and Metadata. The following subsections describe the digitization and metadata processes in the context of capabilities, functions, and features that currently exist, major system components, and interfaces.

A.1.1.1.2.1.1 Capabilities, Functions and Features of the Current System. Although planning has begun for the project of scanning legacy materials in the depository libraries, no system is currently in

place at this time. There are no capabilities, functions, or features of a current system to describe. However, when the project begins, most digital scanning projects will produce files of the documents and additional information to help establish bibliographic control over the scanned material, including MARC records with URLs and other links.

A.1.1.1.2.1.2 Major System Components and High Level Interconnection. Although planning has begun for a project of scanning legacy materials in depository libraries, no system is currently in place. When the project begins, cooperating institutions that undertake digital scanning projects will produce images, derived textual files, technical and structural metadata, and descriptive metadata, including MARC records, to establish bibliographic control over the scanned material.

In addition to the CMS, the following will be the major components for a system for legacy scanning:

- Scanning in GPO Plant;
- Scanning by outside contractors;
- Workflow management;
- Metadata creation; and
- Storage of the scanned images.

A.1.1.1.2.1.3 Interfaces to Systems and Procedures. Currently there are only limited systems in place to do the scanning, and therefore no interfaces to systems and procedures currently exist.

A.1.1.1.3 Existing Digital. In the current environment, approximately 75% of the content that GPO makes available to depository libraries is in digital format on line. Most of these products are pushed to GPO and its dissemination programs and do not have to be located.

The largest subset of GPO online universe consists of files from core legislative and regulatory products on *GPO Access*; most of these are available to GPO as ancillary outputs of the traditional printing process. In these cases there is no metadata created at the individual product level, although in some cases library-style descriptive metadata, as described in section 3.3.1.2.2, is created by ID staff and included in the CGP and the cataloging record distribution scheme.

Another significant portion of the titles available from GPO online are executive branch information that GPO either hosts or links to, such as the reports on the Department of Energy's *Information Bridge* site. Again, these require no locating or harvesting activity on GPO's part. Also, the supplying agency generally applies a descriptive metadata scheme and a search engine for product retrieval. These are consolidated with, or replicated in, GPO's cataloging operation.

There is a significant online resource location activity that takes place in GPO's ID operation. Agency online resources that are located and harvested by GPO's ID staff have descriptive metadata created for them.

A.1.1.1.3.1 Locating. The following sections describe the process of locating data in the context of capabilities, functions, and features that currently exist, major system components, and interfaces

A.1.1.1.3.1.1 Locating Capabilities, Functions and Features of the Current System. Over the past few years, GPO has become increasingly aware that many publications being published by Federal agencies are not being included in the FDLP; these documents have come to be known as "fugitive publications." With increasing frequency, agencies are publishing information only in electronic formats and, when this occurs, they frequently fail to inform GPO of these new publications for inclusion in the FDLP. In addition, agencies sometimes procure their printing directly from private sector companies or use in-house facilities rather than coming to GPO, and they then fail to inform GPO of these publications,

although there may be electronic counterparts on the publishing agency Web sites that could and should be included in the FDLP.

In attempt to partly alleviate this problem, a significant online resource location activity takes place in GPO's ID operation. The current system relies almost entirely on manual intervention and other parties pushing content or notifications about content to GPO. Individual online products are located by ID staff using:

- Cues from print products, which may refer to an online version;
- Cues from news media;
- Routine mining or visits to agency Web sites, resulting in the location and identification of new online products;
- Cues provided by the library community, typically through the "LostDocs" reporting channel; and
- Teleport Pro commercial, off-the-shelf software.

A.1.1.1.3.1.2 Major System Components and High Level Interconnection. Currently, the locating function carried out by ID staff is predominantly a manual process that merely requires a standard Web browser and an e-mail connection.

In light of the large number of publications that have become fugitive, GPO is in the process of identifying and eventually obtaining Web crawler and data mining tools that can provide an automated solution for the identification and harvesting of fugitive documents and publications from agency Web sites. GPO is actively seeking these technologies as a future means of locating and harvesting publications on agency Web sites that fall within the scope of the FDLP and other dissemination programs but that have not been cataloged by GPO.

A.1.1.1.3.1.3 Interfaces to Systems and Procedures. Once the online resources are located, they are harvested and assigned access metadata as outlined in section 3.3.1.1.4. Resources on *GPO Access*, or from agency Web sites, are typically accessed by the interface specific to that Web site but may not receive additional processing from ID staff.

A.1.1.1.3.2 Harvesting. The following subsections describe the process of harvesting data in the context of capabilities, functions, and features that currently exist, major system components, and interfaces.

A.1.1.1.3.2.1 Capabilities, Functions and Features of the Current System. Two systems are currently in place to harvest documents from Federal Government Web sites, the Teleport Pro System and the OCLC Digital Archive. The current system enables users to search the CGP, online library catalogs that use GPO cataloging records with active links, or the OCLC Digital Archive to connect to the harvested resources.

A.1.1.1.3.2.2 Major System Components and High Level Interconnection. GPO uses two methods to harvest data, the Teleport Pro System and the OCLC Digital Archive. The Teleport Pro is an all-purpose high-speed tool for harvesting data from the Internet to capture a copy of the target digital resource, which resides dormant on an "archive server" until such time as the original version on the publishing agency Web site is no longer available. Then the PURL resolver table is adjusted to resolve the original version to GPO "archive server" copy.

GPO has only recently begun to use the OCLC Digital Archive to harvest files. The OCLC Digital Archive is a suite of tools to prepare and add objects to a secure offsite storage facility. The major system components include tools for record creation, bibliographic records, and preservation metadata records that are linked to the bibliographic records.

A.1.1.1.3.2.3 Interfaces to Systems and Procedures. Using the Teleport Pro System, the harvested copy is downloaded onto a PC and then sent by FTP to an "archive server' in GPO designated *permanent.access.gpo.gov*. A PURL is assigned by ID staff to the harvested resource. A PURL is an HTTP link employing a resolver application which routes the user to the correct location of the resource. Once the online resources have been uploaded to the "archive server," which is not OAIS-conforming, access is maintained through a link from the bibliographic record.

The OCLC Digital Archive works with known HTML and PDF objects and creates XML-encoded preservation metadata that is packaged with the object for ingest into the offsite archive. The tool also facilitates archive management. GPO maintains full control of the content and metadata in the archive and controls access privileges and mechanisms.

A.1.1.1.4 Born Digital. GPO accepts and processes electronic files in a full spectrum of formats. The files are normally furnished manually, with much of the internal processing being done by transferring hard copy and storage media by hand. Agencies currently process work through GPO based on *Title 44 U.S. Code*.

Although suggestions for desktop publishing software applications are provided to customers via *GPO Publication 300.6*, there are no mandatory requirements regarding file submissions. Adobe Acrobat is currently the suggested file format for customers creating documents in Office Graphics that require hard copy production, digital media production, or Web-based access. The Electronic Publishing Section (ePUB) provides suggested practices, file format settings, and processes to create acceptable documents.

GPO Agency Direct (formerly the "OMB Compact") procurements specify that vendors create a Print Optimized PDF file from the furnished production files. PDF files are delivered manually via hard storage media; no automated method for submitting PDF files is in effect. ID then uses these PDF files or Print-On-Demand production and the files are archived. SPAs and one-time purchases are occasionally requiring contractors to return digital deliverable final production files, but only at the agency's request. Many term contracts with vendors now require PDF files to be created from the native production files and returned on all orders.

Customers often demand that GPO find a way to make the files work based on the end product they require, or they will procure the products on their own.

However, many commercial print vendors avoid bidding on procurements that contain office graphics, due to re-flow, color shift, and general file integrity issues. The industry is beginning to adapt to processing non-standard desktop publishing files, but production delays and additional costs are common due to quality and consistency issues.

Various documents provided to GPO have statutory requirements for producing hard copy, digital media, and Web access/soft display capabilities.

GPO requests that customers furnish GPO Desktop Publishing Form 952 to provide basic information regarding layout software, version, points of contact, and any system work that is required to the furnished layout files. GPO Standard Form 1 (SF1) is normally provided, with specific requirements regarding production specifications. GPO Form 3868, "Notification of Intent to Publish," is also utilized to gather information about the document. These three forms provide information such as operating environment; printing specifications; and creator. This information is not stored electronically and is generally kept in hard copy format with no relative database references. GPO creates the procurement specifications based on this gathered information. Contractors utilize the bid specifications as the standard for what is required and what is being supplied.

Although customers are encouraged to provide a completed GPO Form 952, it is not mandatory. Agencies have sometimes resisted the excess work required to provide GPO with metadata. Many orders are currently processed with no supporting documentation regarding the originator, production specifications, or operating environment. Short production schedule requirements have created situations where files are not pre-flighted at GPO, and contractors are forced to delay production due to file format incompatibilities.

Metadata is collected from the SF1 and Form 3868 relative to one-time usage to create the printing specifications. The main information used relates to the data transferred to PICS (i.e., agency Billing Address Code (BAC), Requisition Number, Product Code, Quantity, Date Received, etc.). The majority of information is discarded beyond a single use function. Portions of the data used to create the specifications are searchable in the databases on local levels (i.e., MS Access - ability to search by title, etc.) within GPO. Metadata from PICS is also used to create initial records in ID's ACSIS (Acquisition, Classification, and Shipment Information System), which maintains metadata on FDLP publications at the piece level. Metadata from the Form 952 is used to describe the characteristics of the layout files in procurement specifications, and it is then discarded. ePUB and various offices track layout software and computer platform information from this data.

Often the originator of the electronic files has no specific instructions on what information GPO or contractors require. Formal documentation is lacking regarding cataloging, indexing, and minimum requirements for file submission.

A.1.1.1.4.1 Content. The following sections describe born digital data in the context of capabilities, functions, and features that currently exist, major system components, and interfaces.

A.1.1.1.4.1.1 Capabilities, Functions, and Features of the Current System. Born digital files currently allow processing and production of virtually any method. When requirements dictate hard copy or digital media production, layout and formatting issues are addressed.

The files are normally transferred throughout GPO manually via hard storage media. Once the order is received, it is manually carried to in-house production, or it is procured commercially. Contracts are awarded to commercial vendors via competitive purchases, existing term contracts, or direct buys via the agency utilizing the SPA or GPO Direct. The electronic media is manually picked up by the vendor or sent via overnight courier service. Once received at the vendor's plant, the files are copied to the vendor's system and pre-flighted. Pre-flight refers to the operation of evaluating an electronic format to determine if all of the elements necessary for producing the desired output are included and correctly created. Pre-flighting can be done manually by a computer operator who evaluates the native application format and each element within it, or automatically by utilizing a computer program that evaluates the native application format and advises of possible problems. Production generally takes place after proofs are approved.

Once the vendor completes production, the electronic files are returned to the originating agency. Occasionally, digital deliverables or PDF files are returned by the vendor. The agency will post the PDF on the Web and use the final production files on future re-procurements. SuDocs collects the file for POD usage and provides online access in a variety of derivative formats.

GPO in Washington, DC, has limited storage capabilities for a portion of the data submitted through Printing Procurement (only small purchase and one time bid jobs are reviewed) on their server, and ePub and various sections access the data. It is active for a minimum of three months and than backed up to tape. Some RPPOs store copies of the data that is forwarded to contractors for short-term access.

A.1.1.1.4.1.2 Major System Components and High Level Interconnection. The agency creates the digital file and provides the file to GPO. Files are occasionally transferred via FTP, email, etc., but more

commonly files are received on on CD-ROM, zip disks, or floppy disks. Upon receipt, GPO reviews the customer requirements and either sends the job to in-house production or procurement. Production normally consists of hard copy, digital media, or Web access soft display. **Figure A-1, Born Digital Process Flow**, depicts the born digital process from file creation to output.

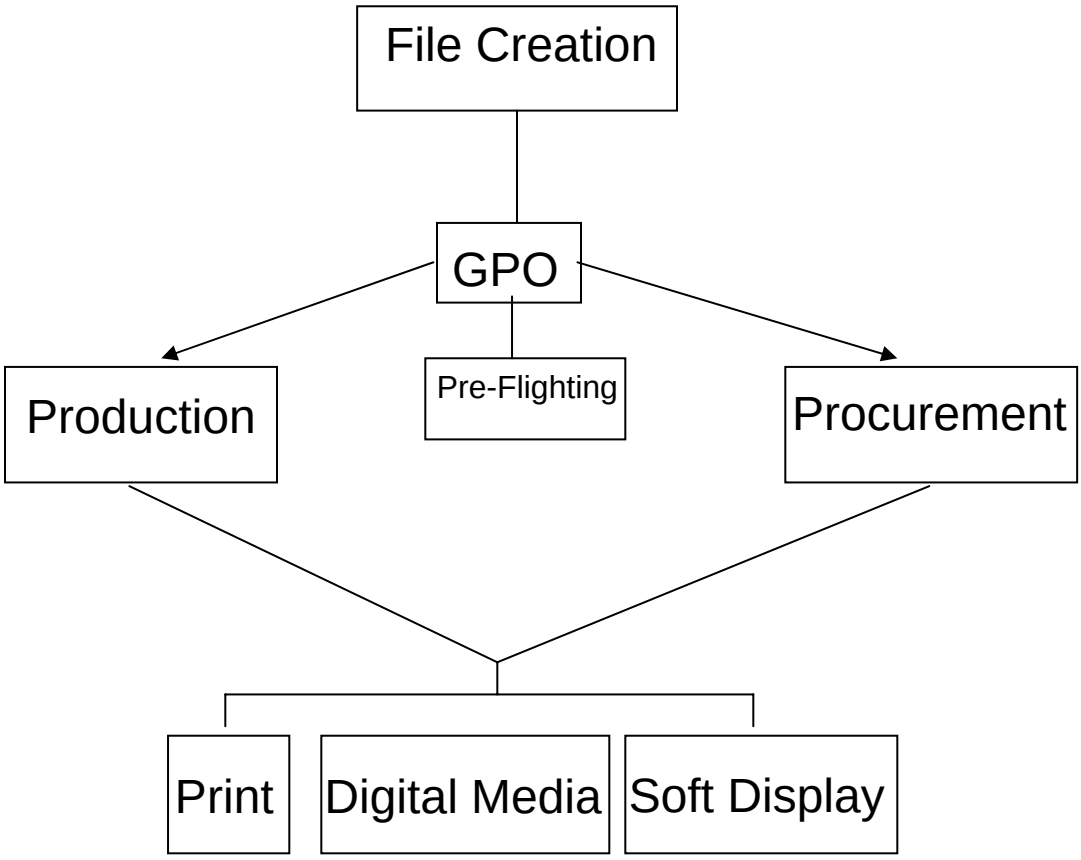


Figure A-1. Born Digital Process Flow

A.1.1.1.4.1.3 Interfaces to Systems and Procedures. Currently, GPO has in-house capabilities to convert files into multi-purpose documents (print-on-demand, CD-ROM, PDF, XML, etc.). Print-on-demand is a major initiative being developed as a flexible and cost-effective way for customers to produce only the minimum required quantity of hard copies. The majority of electronic files are currently processed from conventional storage media (CD-ROM, zip disk, or floppy). GPO reviews the electronic files and makes suggestions regarding changes to file formats or alternate production methods.

A.1.1.1.4.2 Metadata. The following subsections describe the process of gathering metadata in terms of current system capabilities, functions, and features, major system components, and interfaces.

A.1.1.1.4.2.1 Capabilities, Functions, and Features of the Current System. Metadata, primarily administrative in nature, covering output specifications, responsibility, and billing, is gathered or created to support the management of the job on several paper forms, including SF1, GPO Form 868, and GPO Form 952. The information is summarized for the particular job on GPO Jacket, which carries a control number. This metadata survives only for the life of the job -- until a product is delivered or a service is completed.

Automated methods of gathering metadata are lacking in the current system. With limited required documentation, the metadata received on current orders is gathered manually, with data management focused mainly on processing the orders based on the end product requested.

A.1.1.1.4.2.2 Major System Components and High Level Interconnection. Information about the electronic files is gathered from GPO Form 952, SF1, GPO Form 3868, and basic file information direct from the files. The printing specifications are then drafted for procurement. Born digital metadata requirements are depicted in **Figure A-2, Born Digital Data Requirements**.

The current environment for capturing metadata is a reactive workflow, gathering most information after the data has been created and received by GPO. With the exception of the data transferred to PICS (i.e., BAC, Requisition Number, Product Code, etc.), all information gathered from the hard copy is utilized to perform a single use function. Standards and required formatting information are not documented for agency reference.

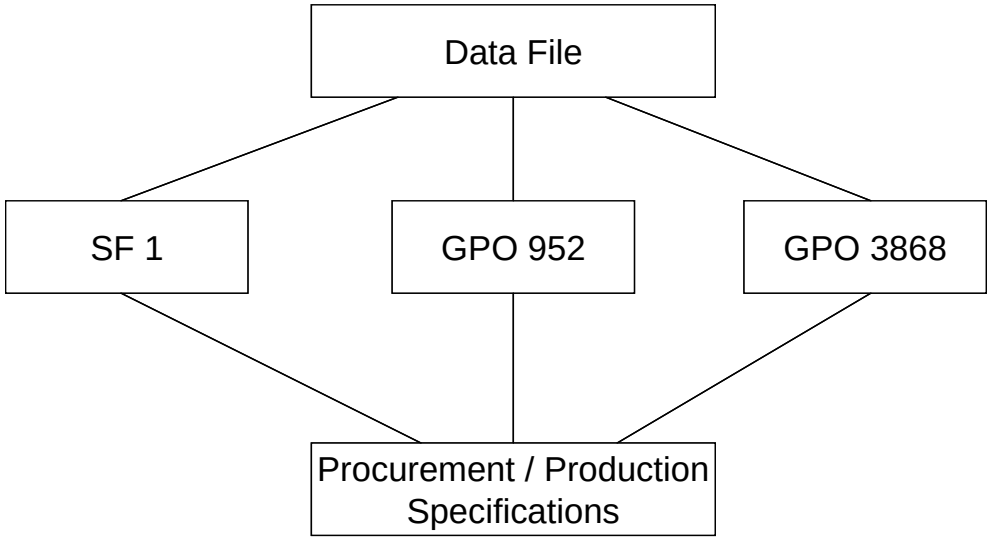


Figure A-2. Born Digital Metadata Requirements

A.1.1.1.4.2.3 Interfaces to Systems and Procedures. Although some data elements (e.g., title, dates, author) are common with other metadata created and used in other processes (e.g., cataloging and indexing), information from the forms or the jacket may be adapted manually. No automated exchange capability exists.

A.1.1.2 Content Management. GPO's current CMS consists of four independent areas (Customer Services, In Plant Production Services, ID, and IT). Once information is processed from the customer, it is subsequently stored and accessible from multiple areas within GPO.

Customer Services and In Plant Production Services work in tandem via a mainframe system to process customer orders and deliver products, as specified, in tangible or intangible formats. A limited portion of these orders are transferred to IT or ID for additional processing and archiving.

IT maintains GPO's servers and Web services (e.g., *GPO Access* and agency Web hosting) in a non-automated fashion. ID shares tangible and intangible documents to the FDLP and the Sales Program.

Recognizing the need for an enterprise-wide solution, GPO, which currently has multiple individual systems that do not integrate, has investigated possible enterprise-wide CMSs.

Information needs to be located and accessed by the public. Historically, GPO has created descriptive metadata files in the form of cataloging and bibliographic information. However, due to fugitive documents as well as limited financial and human resources, the comprehensive nature of this program has eroded into cataloging the documents in the FDLP. This descriptive information is integrated into reference and search tools such as the MOCAT, a suite of tools called Locator Tools and Services, OpenText, and the CGP application on *GPO Access*.

GPO gets information for reports from a database called the PICS. Information is manually keyed into PICS from forms or transferred from other software.

Once the public locates a publication they assume that it is official and authentic. GPO has begun implementing a PKI in order to make sure documents are not changed without GPO knowing about it. Unfortunately, the public also expects to have the most recent version of a document, which might not be possible since GPO requires no formal verification for documents that are processed. The approval process is one way to establish the version of a document, but a better way might be to use a CMS. However, GPO does not currently have the financial resources to implement a CMS.

A.1.1.2.1 Authentication. Customers expect that information made available from the Federal Government over the Internet is official and authentic. Additionally, legislative mandates such as the Government Paperwork Elimination Act (GPEA) and others have requirements that include the need to maintain confidentiality, authentication, data integrity, and non-repudiation of information made available online. In response, GPO has recently taken steps to implement a PKI; one of the business processes associated with the completion of this project is to ensure customers that the information made available through *GPO Access* is official and authentic through the implementation of digital signatures. Implementing digital signatures on *GPO Access* documents will enable customers to determine that the files are unchanged since they were authenticated by GPO. GPO has already completed work on the infrastructure. The CIO staff is currently testing digital signatures in order to apply them internally to GPO files and externally to *GPO Access* files. This effort is being performed in parallel with work to cross-certify GPO's PKI with the Federal Bridge Certificate Authority (FBCA). Once certification is completed, live signatures will be made available on *GPO Access* files. It is expected that this granularity of authentication will be applied to PDF documents in the beginning, to commence with Congressional Bills at the full document/file level. Granularity refers to the level of specificity to which the document is authenticated. Granularity ranges from authentication of an entire file to authentication of each page in the document.

A.1.1.2.1.1 Chain of Command. The following subsections describe the chain of command in the context of capabilities, functions, and features that currently exist, major system components, and interfaces

A.1.1.2.1.1.1 Capabilities, Functions and Features of the Current System. The planned PKI will provide the means for the following authentication/certification practices:

- Proving the authenticity and integrity of electronic Government information;
- Providing the means to recognize and trust the participants in an electronic transaction;
- Providing the means to track and record that the intended person received accurate information;
- Providing the means to keep business transactions confidential when using a public communications channel (such as the Internet); and
- Providing the means to reduce paperwork processing and archival storage of physical media by replacing physically signed documents with electronically signed documents.

A.1.1.2.1.1.2 Major System Components and High Level Interconnection. GPO's PKI will operate in an environment that supports the transmission, delivery, and receipt of digital communications over the Internet, using digital key pairs that encrypt and decrypt information. GPO will act as its own Certification Authority (CA) for implementing digital signatures on documents available via *GPO Access* and will act as a trusted third party that creates, signs, and issues public key certificates to subscribers (potentially other Federal agencies). The signature of the CA ensures that the contents of the certificate cannot be altered without detection. Additionally, GPO will perform the role of a Registration Authority (RA), collecting and verifying an identity and information entered into a public key certificate. The digital identity for End Users is known as a digital certificate and is digitally signed by the CA. The certificate itself is a public structure containing public information such as the identity of the subscriber and an assertion that the CA has properly identified the subscriber's identity according to a published set of management controls; and it binds the identity in an inalterable manner to the subscriber's public key. This public key is published in the certificate and can be used by anyone in the process of validating an identity. The presence of the signature on the certificate from GPO as an authorizing source will enable the certificate user (relying party) to establish trust in the certificate owner's (subscriber's) identity. In addition, the presence of the public key in the certificate signed by the CA enables verification that the certificate owner possesses the corresponding private key.

A.1.1.2.1.1.3 Interfaces to Systems and Procedures. Confidential Diagrams at Present Not Available for this Report –Cannot Release this Information due to Cross-Bridge Certification Process GPO is going through. Access to Information is only available to Trusted Role PKI Staff in accordance with management control documentation and audit compliance. The infrastructure will include the CAs, Directories, and RA Workstations in a PKI facility that physically houses CAs, Directories, Registration, Roaming, Verification and GetAccess Servers.

A.1.1.2.2 Versioning. Customers expect to have the most recent version of a document or at least to know what version they are viewing. Processes associated with versioning/version triggers for electronically available files are currently manually based and require subjective human intervention because GPO does not have a CMS. Version control is limited to the extent that files and Web pages can be overwritten on servers, and it is impossible to tell who made the changes or when they were made. In the print environment, manual-based paper processes and associated electronic tracking processes for dissemination of different versions of documents are available to a very limited extent through mainframe-based applications and extensive knowledge of cataloging records.

Currently, for GPO procurements, the customer is not asked about the document creator or authentication of files. No formal verification is provided, or required, by GPO to customers regarding documents that are processed. The workflow expects that the customer will be responsible for verification of the document being produced (except for ensuring documents are not copyrighted) and abide by Joint Committee on Printing regulations.

By receipt of the customer's signed SF1 or direct deal print order Form 2511, the form is then used as a means of certifying or authenticating the document. Commercial procurements also normally require that proofs be submitted for approval. Once the proof has been signed off, this approval process is one of the ways to establish the chain of responsibility and the version of the product. After a contractor has delivered the final product, and it has been accepted by the agency, they have established the version of the product in printed form.

Many Federal Government documents are produced in hard copy and are available electronically on multiple Web sites. The public must have confidence that the electronic file that it accesses from a Government Web site is the most current version available. The ease of making changes to posted files by unidentified GPO personnel has created a need for certification of the integrity of the files. For example, changes to electronic files for uploads to GPO's web services lack an integrated method for

developers to coordinate changes. Consequently, content developers can overwrite the changes of other developers without notification.

Agency procedures range from no formalized provenance or lifecycle management to structured archiving and authenticating of data. Many procurement divisions within agencies expect the local level (originator) of a document to ensure integrity and to interface with forms managers. Directives exist within most groups, but procedures are lacking to validate these requirements.

The motivation for establishing certified versioning is to ensure the public can have a method to confirm format integrity of the files.

A.1.1.2.2.1 Chain of Responsibility. The following sections describe the chain of responsibility in the context of capabilities, functions, and features that currently exist, major system components, and interfaces.

A.1.1.2.2.1.1 Capabilities, Functions, and Features of the Current System. Automation for authentication and versioning does not exist in the current procurement system. Customers are responsible for data management and for working with the majority of documents they produce.

The Government has a requirement that GPO needs to fill regarding authentication and storing of documents. The public would then have a consistent location or method to search for Government information and research materials. Data management would take place on a single level, as opposed to the current system, which creates multiple versions of documents across the entire spectrum.

Functions are required to establish thresholds for current documents that will create version triggers in the data management cycle. **Figure A-3, Potential Trigger Points for Versioning**, depicts the version triggers in the data management cycle.

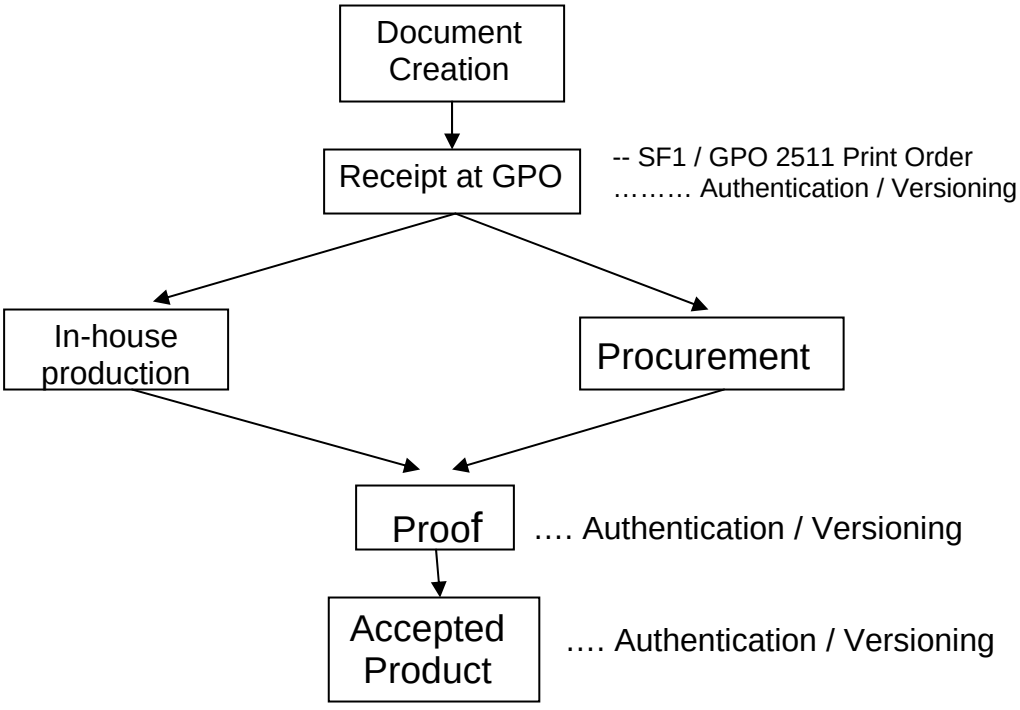


Figure A-3. Potential Trigger Points for Versioning

A.1.1.2.2.1.2 Major System Components and High Level Interconnection. Currently, no documentation or standards exist for requiring certification of customer documents that are procured by GPO. The SuDocs division creates extensive metadata records such as creator, dates, classification information, etc. Fugitive documents have created a void in public access because they are not available to GPO FDLP.

A.1.1.2.2.1.3 Interfaces to Systems and Procedures. Currently, no verification of versioning for agency-submitted documents occurs for procured work. The chain of responsibility is resident at the agency level only. One of the current methods used to establish the version of a document occurs when an agency places a signed SF1 or direct deal Print Order. Approved proofs, or the final delivered approved product, constitute another method of ensuring that the version/responsibility for authentication have been established.

A.1.1.2.3 Access. GPO has a legal mandate under 44 U.S.C. 1710-11 to prepare and publish a "comprehensive index of public documents," including "every document issued or published…not confidential in character." Historically this mandate has been fulfilled by creating descriptive metadata, specifically cataloging, or bibliographic records that conform to accepted national library standards and practices.

Over time GPO's Cataloging and Indexing Program has become a catalog of the publications distributed in the FDLP. This erosion of comprehensiveness has resulted from several factors, including fugitive documents not available to GPO, the advent of and increase in digital publishing, and ongoing constraints on program fiscal, IT, and human resources.

GPO cataloging records include:

- A description of the publication, such as the author, publisher, dates, etc., following the Anglo-American Cataloging Rules, 2nd edition.
- Subject terms derived from the Library of Congress Subject Headings and other thesauri, which characterize what the publication is about.
- A unique identifier. For tangible items this is the SuDocs classification number that provides a library shelf location. For online resources, GPO includes a PURL, which is an HTTP link employing a resolver application that routes the user to the correct location of the resource.

All of this data is encoded in MARC21 format, a standard, structured metadata communications format used worldwide.

In addition, there is a very small and relatively inactive group of Government Information Locator System (GILS) records available on *GPO Access*. Also, there are several search and reference tools available to help locate Federal documents.

Reference tools help library staff and the general user to access the Federal information products provided to libraries under the FDLP. The major reference resource, the MOCAT, and the online equivalent, CGP, are addressed under Cataloging & Indexing. For many years MOCAT was a tangible product, which was initially published in book format and then on CD-ROM before it was converted to a Web format. With the move to online electronic resources, additional reference tools were developed to meet the specific needs of patrons wishing to access the electronic products. These resources started out as a collection of services entitled Pathways Services, but they are now called Locator Tools and Services. The resources include the CGP, New Electronic Titles (NET), Browse Topics, U.S. Government Online Bookstore, Federal agency Internet sites, GILS, Subject Bibliographies, and reference tools on *GPO Access* Web site. These services are a mixture of GPO resources and partnerships with agencies and libraries.

A second set of resources includes the List of Classes (LOC) and related files located on the Federal Bulletin Board. The files are the LOC, Inactive or Discontinued Items, List of Active Items and Class Items, Classes No Longer Active (LOC, Appendix II), and Alphabetical Listing of Government Authors.

The final resources are the OpenText resources. WebTech Notes includes information about corrections and changes to classification numbers and updates on information dissemination products. These resources are shared with the depository community through WebTech Notes. The FDLP Directory, which is an online application that also utilizes the OpenText product, provides current information about the depository libraries participating in the FDLP.

Search tools available include the databases available on *GPO Access*, which provides free electronic access to a wealth of important information products produced by the Federal Government. *GPO Access* is comprised of about 3,200 databases over 100 applications. These databases include those associated with Congressional publications such as the Federal Register, Congressional Record, and the House and Senate Reports and Documents, Hearings, Code of Federal Regulations (CFR), Unified Agenda, House and Senate Calendars, Public and Private Laws, and the U.S. Code. Hosted databases include the Supreme Court Reports, GAO Reports, and Equal Employment Opportunity Commission documents.

In addition to the resources available via the FDLP and *GPO Access*, ID also administers GPO Sales Program. Through this program, users can locate and order publications (books, maps, serials, videos, CD-ROMs, subscriptions) available for sale through GPO. Orders may be placed online in a secure environment via the U.S. Government Online Bookstore, and also via mail, fax, phone, teletype, or by visiting the main bookstore at the main GPO facility in Washington, DC.

GPO generates reports via data mining from PICS, which is the primary mainframe database for information used by the Printing Procurement Department to provide tracking and control of orders. PICS processes two basic types of orders. One type is recurring, i.e., print or work orders placed against term contracts or SPAs, respectively, that are in effect for a specific time period. The other type is one-time orders that are procured via small purchase or sealed bid/negotiated methods. As each of these types of orders has different keys (identifying fields) by which they are accessed, PICS was designed to accommodate them.

Associated mainframe system files are:

- The Automated Bid List System (ABLS), which is the initial source for GPO's information on contractors' names, addresses, and printing capabilities.
- The Printing Cost Calculating System (PCCS), which provides an automated estimating system for work to be placed against applicable term contracts.
- The Marginally Punched Continuous Forms (MPCF), which provides an automated estimating system for work placed against the main MPCF term contract, 1026M. This contract is administered by the central office, with regional offices utilizing the same contractors.

Due to the continued reliance on hard copy/manual order origination and processing, PICS is the focal point for extracting ordering histories and projections for future business. Information is either manually keyed in or uploaded through software interfaces that help eliminate some duplicate key stroking.

Reports are generated using specific reference data, much of which was manually extracted from the SF1 or job jacket and keyed into the system. Search capabilities include the ability to use information such as billing address codes, product codes, and GPO office codes. A total of 126 different computer generated reports have been created to data mine the PICS system, with approximately 22 of these reports no longer being used. The results of these hard copy reports are used to document trends in purchasing,

volume and cycle of orders processed, dollar values, etc. This information can then be used to develop new business opportunities and help project future staffing requirements.

Many agencies come to GPO managers with reports based on their workflow history that are more detailed and specific in relation to order history than GPO is currently able to provide due to hard copy/manual workflow limitations.

From the ID perspective, various data mining capabilities exist that help to track the success of various dissemination components, including *GPO Access*, the U.S. Government Online Bookstore, and the FDLP. ID uses a web monitoring service to track average transaction times on *GPO Access*, error rates, and availability rates. ID is also able to track sales through GPO Sales Program in various ways. Furthermore, ID mines information relating to depository distribution of titles in electronic only and tangible formats.

A.1.1.2.3.1 Cataloging and Indexing. The following subsections describe the cataloging and indexing processes in the context of capabilities, functions, and features that currently exist, major system components, and interfaces.

A.1.1.2.3.1.1 Capabilities, Functions and Features of the Current System. The current Cataloging & Indexing System provides the means to:

- Create library-standard bibliographic records;
- Distribute those records to library customers at minimal cost;
- Provide a searchable public access tool on *GPO Access*;
- Compile and publish hard copy output; and
- Enable users to connect directly to online content or locate tangible products in depository libraries.

A.1.1.2.3.1.2 Major System Components and High Level Interconnection. Since 1976, all GPO cataloging data has been entered into the OCLC (Online Computer Library Center, Dublin, OH) WorldCat database, a comprehensive MARC database of some 50 million library cataloging records, which was created by and has received contributions from over 20,000 member libraries worldwide. GPO is by far the single largest contributor of cataloging records for U.S. Government documents, having created some 510,000 records in WorldCat.

OCLC provides the shared database and the Web-based tools for creating structured bibliographic records in MARC. Records created on OCLC via the Web are batched and returned to GPO by daily FTP transmission. These records are batched and processed by the Office of Chief Information Officer to produce printing files in GPO markup language.

These print files are used to produce two principal outputs of the C&I program: the printed MOCAT and its online counterpart, the CGP, a Wide Area Information Servers (WAIS) application on *GPO Access*.

Monthly batches of records in MARC format are made available for download from the Federal Bulletin Board application on *GPO Access*. These batches, each equivalent to a Monthly Catalog issue, are acquired by individual libraries, value-added resellers, and the Library of Congress. This redistribution enables users to search GPO cataloging records through online catalogs at an untold number of local libraries. In the current environment the use of GPO cataloging records in local library catalogs is far more important to users than the CGP.

GPO has already begun the implementation of ILS software, the Aleph 500 and Metalib/SFX products from Ex Libris, USA. Once implemented, these applications will replace most of the Web-based input

tools and will provide all new public catalog capabilities, linked through *GPO Access* and all current system outputs.

A.1.1.2.3.1.3 Interfaces to Systems and Procedures. **Figure A-4, Access System Interfaces**, depicts the interfaces to the core GPO system for access.

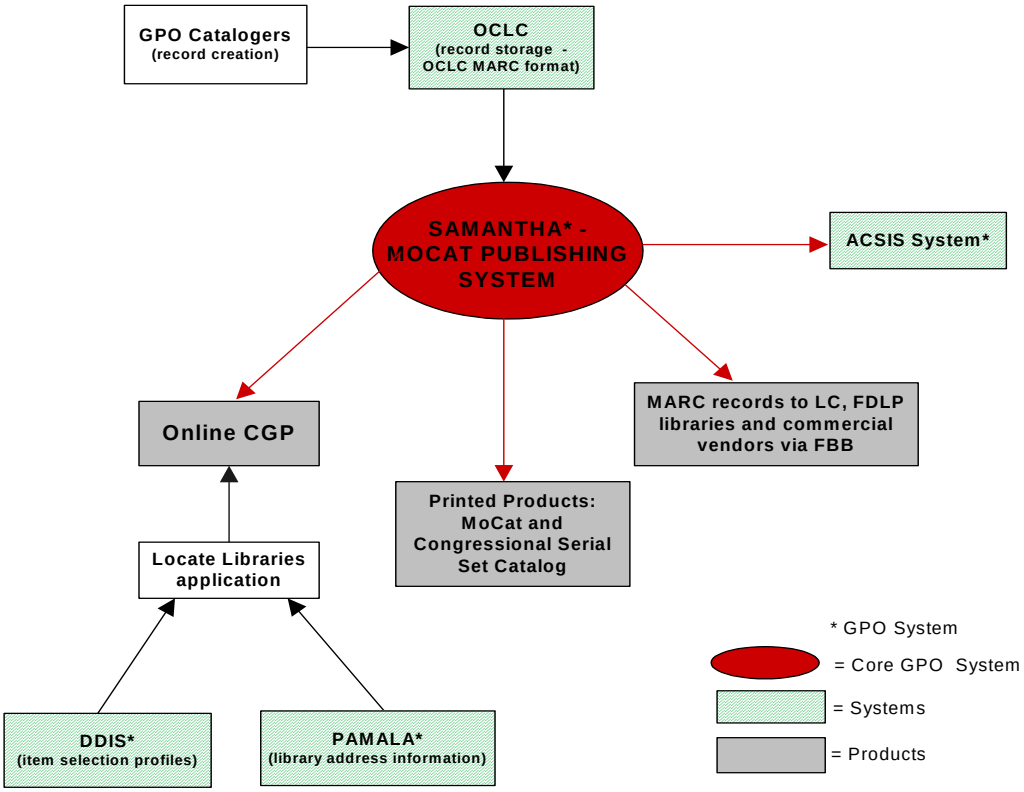


Figure A-4. Access System Interfaces

A.1.1.2.3.2 Reference Tools. The following subsections describe reference tools in the context of capabilities, functions, and features that currently exist, major system components, and interfaces.

A.1.1.2.3.2.1 Capabilities, Functions and Features of the Current System. Reference tools are comprised of bibliographies and other Web tools that assist users in locating Government information that is within the scope of the FDLP. Reference tools include metadata, references to metadata, and/or references to content. Some of the reference tools are HTML based, while other tools employ search technologies such as WAIS or OpenText. The tools in WAIS format have limited future growth capabilities, as that format is no longer being supported outside of GPO. This is also the case for the OpenText software. Plans are proceeding to obtain replacement service for all the WAIS applications and the OpenText resources. Some of these will fit under the architecture-wide Oracle applications. The ILS, currently being procured by GPO, will replace resources currently created from legacy databases: the CGP, NET, WebTech Notes, and the LOC.

A.1.1.2.3.2.2 Major System Components and High Level Interconnection. The major system components for the reference tools include GPO WAIS servers for the CGP, NET, U.S. Government Online Bookstore, Subject Bibliographies, GILS, and reference tools for *GPO Access* Web site.

Partnership sites are responsible for many reference tools, including Federal agency Internet sites, Browse Topics, and the Cybercemetery.

The major system component for the LOC files is the Federal Bulletin Board, which is an NT server with FTP capability.

The major system components for the WebTech Notes and FDLP Directory are the OpenText applications and the databases located on GPO servers. Note that these are not WAIS applications.

A.1.1.2.3.2.3 Interfaces to Systems and Procedures. The CGP and NET utilize flat text files indexed by WAIS and output from the mainframe legacy system, SAMANTHA.

The U.S. Government Online Bookstore interfaces with STAIRS, PRF, and ROPS. The Subject Bibliographies are produced with information from the STAIRS mainframe system. The information is loaded into a work file from the mainframe, and updates/changes are performed and loaded onto a floppy for loading onto the U.S. Government Online Bookstore.

The databases for the WebTech Notes and FDLP Directory are developed on workstations in the ID organization and then updated on a regular basis from information loaded onto floppy diskettes.

A.1.1.2.3.3 Search. In 1993, Congress passed *Public Law 103-40*, which amended GPO's duties to include provisions for the dissemination of information in electronic format. According to *Title 44*, GPO is required to:

- Maintain an electronic directory of Federal Government electronic information;
- Provide a system of online access to the Congressional Record, the Federal Register, and, as determined by the SuDocs, other appropriate publications distributed by the SuDocs; and
- Operate an electronic storage facility for Federal Government electronic information.

As its programs evolved into Web technology, GPO needed an online system that would allow Federal agencies, library partners, and the general public to access its information resources. Utilizing search and browse technologies, GPO's online resources have evolved and expanded to the point where they offer traditional print publications, such as the CFR, as well as new innovations in Government publications, such as the dynamic daily update of the CFR, known as the eCFR. The popularity of GPO's Web services has led to the growth of its online services, such as *GPO Access*, which today consists of over 100 applications, consisting of more than 2,500 databases, as well as the expansion of GPO's Federal services to include Web hosting and design.

A.1.1.2.3.3.1 Capabilities, Functions and Features of the Current System. GPO's online system, *GPO Access*, was launched on June 8, 1994. At the time of launch, the system provided fee-based access to three databases. The system became available to the public free of charge on December 1, 1995. Traditionally, the vast majority of information made available via *GPO Access* has been derived from databases used in the printing of Government publications. The databases are delivered by a distributed text searching system called Wide Area Information Servers (WAIS).

While this text searching system was cutting-edge at the time, the creator of WAIS, Thinking Machines, went bankrupt in 1995. Since then, WAIS has been used supported entirely by GPO's IT staff. The system has since been customized to meet the needs of GPO. Ten years later, WAIS is still the primary search engine deployed by GPO. Files are posted online directly following the receipt of the information from the publishing agency or Congress and they are generally by-products of print products. While GPO has been experimenting with providing dynamic content as demanded by today's user, WAIS cannot support dynamic content delivery. In addition, GPO's Web design and hosting customers as well as internal customers have demanded search capabilities that go beyond the abilities of WAIS, such as PDF

indexing and search. This situation has led to the use of two other search platforms—OpenText and Microsoft. Like WAIS, OpenText is an obsolete search platform.

A.1.1.2.3.3.2 Major System Components and High Level Interconnection. While all of GPO's Web services utilize search functionality, five major entities have unique search demands:

- *GPO Access* disseminates information for all three branches of the Federal Government for Federal agency, library partner, and general public use. All of the resources on *GPO Access*, with the exception of one beta application (the eCFR), utilize WAIS technology.

- The FDLP Desktop provides cataloging and reference records or the Federal information resources distributed through the FDLP. One of the primary applications for the FDLP Desktop, the CGP, is provided through WAIS. Other applications provided by the FDLP Desktop, however, go beyond the capabilities provided by WAIS and have been developed using OpenText. Recently, GPO procured an Integrated Library System (ILS) to replace many of the FDLP Desktop applications. Work is progressing to implement an ILS at GPO.

- The U.S. Government Online Bookstore allows users to purchase tangible publications on line. The search functionality of the bookstore is solely implemented by WAIS; however, GPO's Oracle implementation is addressing the replacement of WAIS for the Bookstore.

- GPO's Web hosting and design services offer a mixture of WAIS and Microsoft search platforms. More recently, sites hosted by GPO (such as the U.S. Supreme Court Web site) have been utilizing the Microsoft platform. Previously, Web sites (such as the Export Administration Regulations Web site) were utilizing the WAIS platform.

- Customer Services is in constant communication with Federal agencies and vendors to disseminate information products. Currently, Customer Services provides an online system called PICSWEB that allows customers to track, estimate, and obtain information on any of their printing jobs. PICSWEB is provided through GPO's mainframe.

A.1.1.2.3.3.3 Interfaces to Systems and Procedures. Plans are proceeding to obtain replacement service for all the WAIS applications and the OpenText resources in order to provide GPO with an enterprise-wide search capability. GPO is currently implementing a data and search disaster recovery program for selected Web services. In the event that GPO's WAIS engine becomes unavailable, a backup search will kick in until service is restored. While this system has not yet been designed or developed, initial discussions indicate that the search will run on the FAST platform. In addition, the implementation of Oracle and an ILS are progressing.

GPO's WAIS system is currently provided on a mixture of platforms and operating systems. WAIS, in particular, runs on Compaq Tru64 servers that are no longer being supported by Hewlett Packard.
GPO's search system currently exists within the following server structure:

- *GPO Access*: 2 Windows boxes (one primary, one backup), 9 Webgates on UNIX platform (only 7 are operational)
- U.S. Government Online Bookstore: 2 UNIX boxes (one primary, one backup)
- GPO Hosted Sites: 2 Windows boxes (one primary, one backup)
- GPO.gov/Unmigrated *GPO Access* pages: 2 UNIX boxes (one primary, one backup).

Each of these system's interfaces, in one way or another, interacts with WAIS, but not in an integrated, streamlined fashion. Instead, GPO's WAIS system is currently provided on a mixture of platforms and operating systems as mostly independent servers. Serious hardware and security risks persist due to the

nature of the outdated WAIS system. For example, WAIS runs on Compaq Tru64 servers that are no longer being supported by Hewlett Packard.

A.1.1.2.4 Data Mining. The following subsections describe the Data Mining processes in the context of capabilities, functions, and features that currently exist, major system components, and interfaces.

A.1.1.2.4.1 Capabilities, Functions and Features of the Current System. With the current manual order processing and the various databases in use within Printing Procurement and In Plant Production, automated methods of gathering extensive information about customers are lacking.

From the ID perspective, various data mining capabilities exist that help to track the success of various dissemination components. ID uses analysis and aggregation of log files on *GPO Access* to track the number of document downloads from *GPO Access* Web pages and databases, the number of referrals to *GPO Access* Web pages from external Web sites, top referring sites overall, top referring sites categorized by domain type and/or name, and top referral counts from Federal Depository Library Web sites.

In addition to this activity, ID uses a Web monitoring service to track average transaction times on *GPO Access* (the amount of time it takes to search for and download a document within a *GPO Access* database), error rates, and availability rates. The capability also exists to identify exact times in which a large amount of errors or low availability rates were experienced. These same metrics can be applied at a more granular level, measuring out transaction times, error rates, etc., for specific geographic locations or Internet Service Providers.

ID is also able to track sales through GPO Sales Program in various ways. A mainframe system tied to order processing systems can track sales from orders received online and via telephone, mail, and fax. For online orders, a daily cumulative file is uploaded that contains all data from orders received from the U.S. Government Online Bookstore.

ID also mines information relating to depository distribution of titles in electronic only and tangible formats. A report is compiled monthly and disseminated to the appropriate GPO personnel. The report includes listings of classes broken up into the following reports: Lists of Classes in All Formats, Electronic Only Classes, Classes Available in Multiple Formats, and Added and Dropped Classes.

A.1.1.2.4.2 Major System Components and High Level Interconnection. Various databases are connected to PICS via interfaces on a local level, but no agency-wide database information is readily available, except mainframe data. A high level view of data mining opportunities for GPO is provided in **Figure A-5, Data Mining Opportunity Map**.

Various Regional Printing Procurement Offices are using MS Access for generation of specifications, print orders, and purchase orders. Search capabilities at the local level can provide more specific information (Product Titles, Ordering History, Specific Product Descriptions), but other offices cannot readily access this information on the network. GPO Central Office is currently working with NEPTUNE Web-based specification writing, which should provide easier information processing.

Currently Regional Printing Procurement Offices have an existing decentralized configuration consisting of 40 disparate servers (one primary and one backup in each office). These servers run the same databases but are not connected.

A.1.1.2.4.3 Interfaces to Systems and Procedures. Due to the current workflow, in which orders originate and are processed from hard copy into localized databases, large-scale interfacing is not possible.

GPO has an inventory of approximately 126 computer generated reports (22 of these are no longer used) that can be run in conjunction with the PICS system. Examples of these reports include the Billing Address Code (BAC), which lists the Requisition #, Jacket, program/print order, actual billing amount for each order, and total orders with billed amount. Cumulative Award Stats are another series of reports that provide total awards for offices, dollar amount, contractors, and term contract statistics.

Trends are for GPO to migrate to robust and scaleable software such as NEPTUNE for order processing, which will enable electronic order placement with Web-based workflows. This advance will allow for much easier access to the large collections of data GPO processes.

Agencies are requiring GPO to accept and process data directly from there systems, which will eliminate redundant key-stroking and improve the efficiency of the workflow.

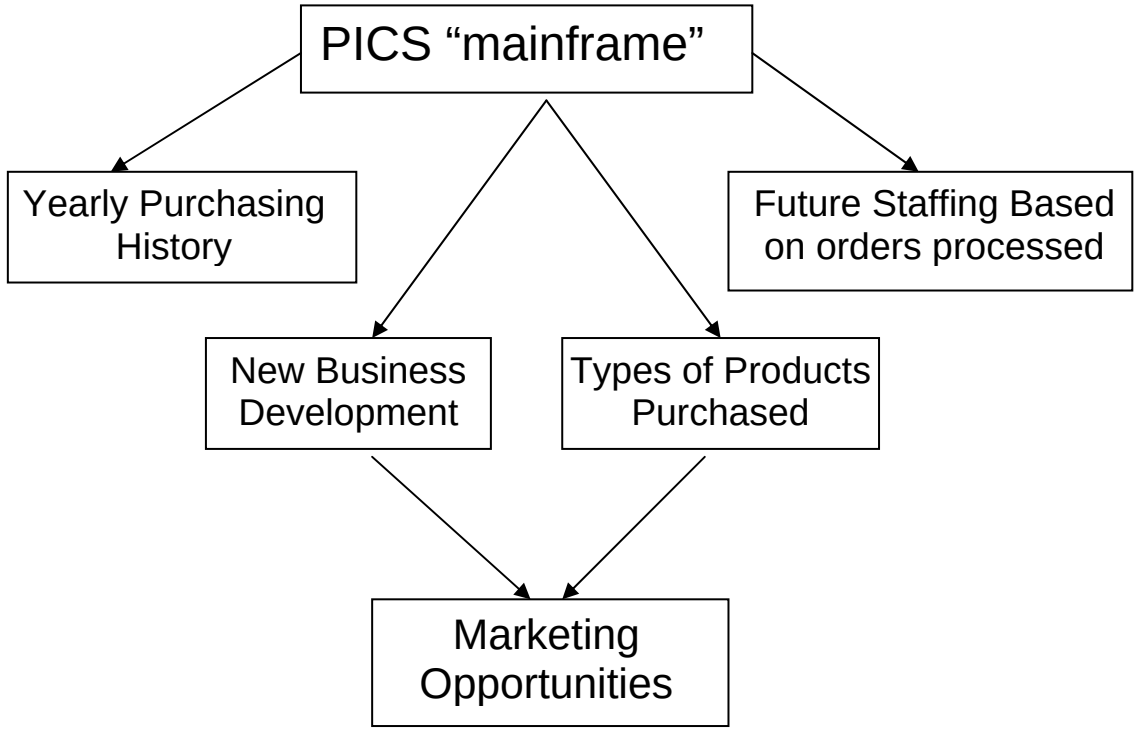


Figure A-5. Data Mining Opportunity Map

A.1.1.2.5 Preservation. The following sections discuss preservation techniques currently used or being pursued by GPO.

A.1.1.2.5.1 Storage Management. The following subsections discuss storage management in terms of capabilities, functions, and features, operational environment and characteristics, major system components, and interfaces.

A.1.1.2.5.1.1 Capabilities, Functions, and Features of the Current System. Besides the noted Depository Library, GPO currently has no real institutional systematic storage management policy that addresses comprehensive archival/preservation storage and retrieval. The capabilities and features are limited to content programmatic applications, or local functional needs. However, GPO is pursuing a mirror capability for all Internet available data, which will reduce the risk of both data loss and availability.

A.1.1.2.5.1.2 Operational Environment and Characteristics. Traditionally, GPO has created and disseminated information for the Federal Government in a tangible format. As technology has improved, GPO has expanded its role to include the production and dissemination of information in diverse, customer-driven digital formats. Today, GPO serves as a clearinghouse for the printing and online dissemination needs of the Federal Government and, as such, receives and produces content in a wide variety of formats from Federal agencies and private vendors.

Customer Services, In Plant Production, Office of Chief Information Officer, and the SuDocs each operate as separate entities using differing technologies and rules for their content storage. Some are product derived, others are simply based on information retention standards, some are archived off-site, and others are based on operational backup and recovery needs.

Tangible and electronic product development and manufacturing formats are retained and archived off site for some products, such as Congressional material; however, other product formats often are simply removed when the job has been completed and exist only in the context of server backups.

The FDLP collections include preservation and access copies of digital objects and tangible publications. These collection components are geographically dispersed, serve different functions, and are managed according to their specific roles in the overall program for public access to Government information.

Furthermore, the Sales Program executes the retirement of publications and publication metadata on the basis of demand; thus, retention of the publication is not based on standard preservation considerations.

Publications handled through the Reimbursable Distribution program are transient and subject only to the instructions provided by the customer agency.

Facilities at this time are diverse and are driven by organizational and/or operational needs. They include UNIX-driven WAIS databases using redundant hard disk systems, which are also replicated to backup systems, Microsoft XP supported Intel systems using standard hard disk systems, and mainframe file backups.

A.1.1.2.5.1.3 Major System Components and High Level Interconnection. Customer Services receives content from Federal agencies and private vendors for printing or electronic products that can be received from any publishing tool and version. Data is temporarily archived and then retired as storage becomes available on a First In First Out (FIFO) basis. Otherwise there is no policy for establishing a central system of storage, tracking, control, backup or retirement of these files either internally or by contractors.

Meanwhile, information that is disseminated on line (e.g., *GPO Access*, Regulations.gov, U.S. Supreme Court) is stored on two redundant servers in the same physical location. These servers are available on UNIX and Microsoft platforms. If the primary fails, the backup is designed to provide 24/7 availability until the primary is brought back online. At present, there is no offsite storage of this data for disaster recovery. However, redundant, network-attached array storage devices are being installed within GPO and at an alternate facility off-site, and Internet storage services are being rolled out for *GPO Access* data.

A.1.1.2.5.1.4 Interfaces to Systems and Procedures. Upon receipt of content for publication or by prior agreements with the requesting agency, Library Programs and the Sales Program will decide whether or not to incorporate the publication in distribution programs. Library Program specialists will review the requests and evaluate them by title or series, and the Sales Program will decide if the publication warrants inclusion on the basis of its potential sales.

In Plant Production develops electronic formats in accordance with the instructions provided, generally archiving material based on the classification of the content. In regard to electronic format, instructions for further distribution of the content are either based on agency instructions or are established through the arrangements of various programs within the SuDocs. Backups or archival copies are made in accordance with local administrative instructions.

Depending on publication format, finished products may become part of the tangible collections of the FDLP or *GPO Access* or included in the Sales Program. Metadata is entered into the appropriate administrative control systems or classification systems depending on the program. Preservation for Depository Collections is applied by the FDLP and is governed by Title 44 USC Chapter 19. Sales Program material is retired based on decisions made by inventory management specialists.

A.1.1.2.5.2 National Collection of U.S. Government Publications. GPO's National Collection is still in the planning stage. It is expected to provide comprehensive, timely, permanent public access to the official publications of the U.S. Government in all formats. The draft plan is available at *http://www.gpoaccess.gov/about/reports/clr0604draft.pdf*.

A.1.1.2.5.2.1 Capabilities, Functions and Features of the Current System. The FDLP collections include preservation and access copies of digital objects and tangible publications. These collection components are geographically dispersed, serve different functions, and are managed according to their specific roles in the overall program for public access to Government information. The National Collection serves three roles in the conceptual overview, serving as the dark archive for preservation of tangible publications and digital objects as well as providing online access.

A.1.1.2.5.2.2 Major System Components and High Level Interconnection. The functional component of the National Collection is the FDLP Electronic Collection, which is described in Section 3.3.1.1, Content Ingest. Tangible publications are currently stored, although with little active preservation, in the distributed network of depository libraries and in Record Group 256 at NARA. The relevant GPO records consist of copies of publications that have been processed for the Cataloging and Indexing Program described in Section **A.1.1.2.3.1 Cataloging and Indexing**.

A.1.1.2.5.2.3 Interfaces to Systems and Procedures. At this time publications are being acquired to form the core of the National Collection. Tangible product copies are being set aside as a by-product of the depository library acquisition and distribution scheme. In addition, GPO is acquiring copies or sets of selected titles that are being removed from depository library collections. Most of these copies are being held in secure storage in the main GPO building.

A.1.1.3 Content Delivery. Under legal authority of Title 44, Section 1702 of the United States Code, GPO's Office of Information Dissemination (SuDocs) has sold official Federal Government publications through its Sales Program. Due to decreased demand and high expenses, the Sales Program has not been able to recover costs during the last several fiscal years. Additional analysis of inventory data has shown the Sales Program is only selling 35% of the inventory ordered. As a result, GPO is developing a plan to reduce costs by making a substantial portion of its products available through POD. When a request is made for a printed copy of a document, an SF1 is manually generated and a jacket number is created for billing purposes. Production then locates the file on the repository and outputs to a digital press.

The ability to produce a publication on demand via digital printing hinges upon the availability of a press-optimized digital file for that publication. GPO's POD Committee (including representatives of ID, Customer Services, and Plant Operations), is actively obtaining press optimized PDF files through various methods. If a press optimized PDF file does not exist for a given publication, hard copies of that publication are obtained and scanned. Due to budget constraints of the project (the concept being to spend as little as possible digitizing titles that do not sell in high-volumes in the first place), approval has

not been granted to procure a formal CMS to manage these files. As an interim solution until a CMS can be implemented, the files are currently being stored in a content repository, which grants authorized users full access to the files, allowing them to create, change, move, and delete all files on the server.

Not every document is available as POD and agencies always have new jobs to be completed. Therefore, agency customers regularly request conventional printed hard copy jobs through GPO. These customers currently have two options: Printing Procurement and Production. If they do not specify which is preferred, GPO chooses the procurement method. In all cases the agency must complete and submit an order form. The majority of forms must be submitted manually, with the exceptions of the pilot program GPODirect (*http://gpodirect.gpo.gov*, formerly known as the OMB Compact) and the Web-enabled mainframe application PICSWEB (*http://govprint.access.gpo.gov*), which allows registered users to electronically submit Direct Deal term contract order forms (2511). No other online processes for submitting printing requirements are currently in operation.

Customers submitting order forms to Printing Procurement have multiple contracting vehicles available to them. The SPA and AgencyDirect both allow agencies to procure their own printing. If an agency chooses to go through GPO's procurement process, a specification is written and bids are solicited. The lowest responsible, responsive bidder as determined by the Contracting Officer within the guidelines of the PPR is awarded the job. Term contracts may be written if the requirement will be repeated throughout a specific time period (usually a year). Direct Deal contracts are managed by the agency, with GPO handling only administrative duties. GPO-placed term contracts are available to all agencies, with all contracting and administration handled through GPO. Any work submitted through Printing Procurement will be completed by commercial printing vendors.

Many agency customers are also requesting from GPO, and GPO's affiliated contractors, digital files that they can place on line for viewing and/or download. Agencies requiring digital files may go through Production or Printing Procurement. Both will accept hard copy and digital input for digital file creation. These files may replace the printed documents. However, the vast majority are created as supplements to the printed pieces. The current system is very printing-centric. Digital layout files are usually created in professional design programs following guidelines for a printed end result (CMYK color, high resolution images, etc.). Once the printed product is completed the files are repurposed for screen usage. Few files are currently created for strictly digital output.

Production has the capability of creating digital files from both hard copy and digital input. If hard copy is provided, Production can scan into a digital format. OCR scanning and subsequent search features are available upon request by the agency. While some hard copy is sent to production for scanning, the current system is inadequate to handle a large influx of agency documents. Personnel, training, and equipment issues must be addressed if this area of revenue is to grow.

Typesetting from hard copy into a digital format is also available. Once a digital file has been created from the scanned hard copy or the digital input, additional features such as bookmarks, links, and indexing may be added to certain digital file types (PDF, etc.). A choice of file formats including PDF, JPG, ASCII, etc., may be provided back to the customer, with PDF being the most commonly requested format. Fill-able PDF files for online use may be created from both hard copy and existing digital files. The agency may also choose from a variety of media, with CD-ROM being the most widely used.

When a digital file is requested in addition to the printed product, the file type and media to be returned to the customer are identified on the order form, along with any additional requirements. Once the printed publication is completed, the specified digital file should be created from the production files and should be an exact representation of the printed product (format, structure, etc.). The file will be supplied to the agency in the format and on the media requested on the order form. This process should be followed by GPO's in-house production facilities and outside contractors.

Many customers also request Web page capabilities in the form of digital files, typically HTML. Printing Procurement can request digital files such as HTML from outside vendors. Various areas of Production can also produce digital files and, in addition, can create and/or maintain Web sites for agency customers.

Agencies requiring digital file creation may submit an order form to Printing Procurement for bidding by contractors. The written specifications include information on the hard copy or file type and media submitted to the contractor, as well as the file type and media to be returned to the customer. Additional requirements for deliverables include searching capabilities, metadata creation, etc. Additional requirements for digital files with Web page capabilities (such as HTML) include coding, version compatibility, etc. Repurposed Deliverables Specification Language is available for inclusion into specifications. This language includes specific requirements for PDF and HTML deliverables. Once specifications are complete, bids are solicited and accepted in the same way they would be for a hard copy job described above.

In addition to customer agencies, GPO's ID section also utilizes digital files for soft copy display. Files submitted to ID (from both In Plant Production and outside contractors) are evaluated to determine if they fall within the scope of ID programs. When a file is determined to be within scope, it is further processed for display on *GPO Access*. Additional derivatives may be created for different purposes from the existing digital file. Examples of processing options include WAIS database indexing and optimizing for placement on a web server. Additional requirements such as OCR scanning, bookmarks, breaking large files into smaller more easily downloaded components, etc., are available to the ID department.

The primary media type currently used at GPO is the CD. CDs may be formatted for both Macintosh and Windows computers. Virtually all computer systems are able to read CDs, making them the most widely accepted form of hard media available today. GPO accepts, outputs, duplicates, and replicates CDs.

GPO is working on expanding its in-house DVD capabilities. Some areas of Production currently have the ability to read and/or write DVDs while others do not. The section that handles CD replicating has DVD equipment in place and is conducing testing on DVD replication. Until testing is completed, a contract with an outside vendor is in place to handle DVD replication work from customers.

Many outdated types of media, such as floppy disks, ZIP drives, etc., are no longer typical at GPO. However, customers with older media may be able to submit their digital files to GPO's Production section for hard copy output and in addition have the files returned to them on a CD.

Jobs that do not go through GPO's in-house facilities are submitted to Printing Procurement. Specifications can be written for any type of digital media provided by the agency customer and will be procured from an outside contractor. While the most common form of media accepted by contractors is the CD, some contractors have the capability of utilizing other media.

A.1.1.3.1 Hard Copy.

A.1.1.3.1.1 Production. The following sections describe hard copy production in terms of capabilities, functions, and features, major system components, and interfaces.

A.1.1.3.1.1.1 Capabilities, Functions, and Features of the Current System. The Production and Printing Procurement sections of GPO provide diversified ways for agency customers to receive printed publications, digital files, and materials from multiple sources. However, the current system requires the majority of interaction between customers and GPO to be performed by phone, e-mail, and physical paper. Interaction between the different departments in GPO, as well as with contractors, is also limited to these methods of interaction. Automated methods of dispersing information are lacking in the current system.

A.1.1.3.1.1.2 Major System Components and High Level Interconnection. On a basic level, the major components for hard copy production jobs are the agency customer, Production, Printing Procurement, and the contractors affiliated with GPO. Within each of those areas are groupings of personnel with different job functions and levels of expertise. These groups have access to different types of hardware and software. In addition, the Production and contractor components possess equipment with varying capabilities. These components are illustrated in **Figure A-6, Hard Copy Production Components**.

Requirements for printed products are currently processed manually with hard copy documentation in the form of various order forms, with the limited exceptions of the pilot program GPODirect and the Web-enabled mainframe application PICSWEB, which allows registered users to electronically submit Direct Deal term contract order forms (2511).

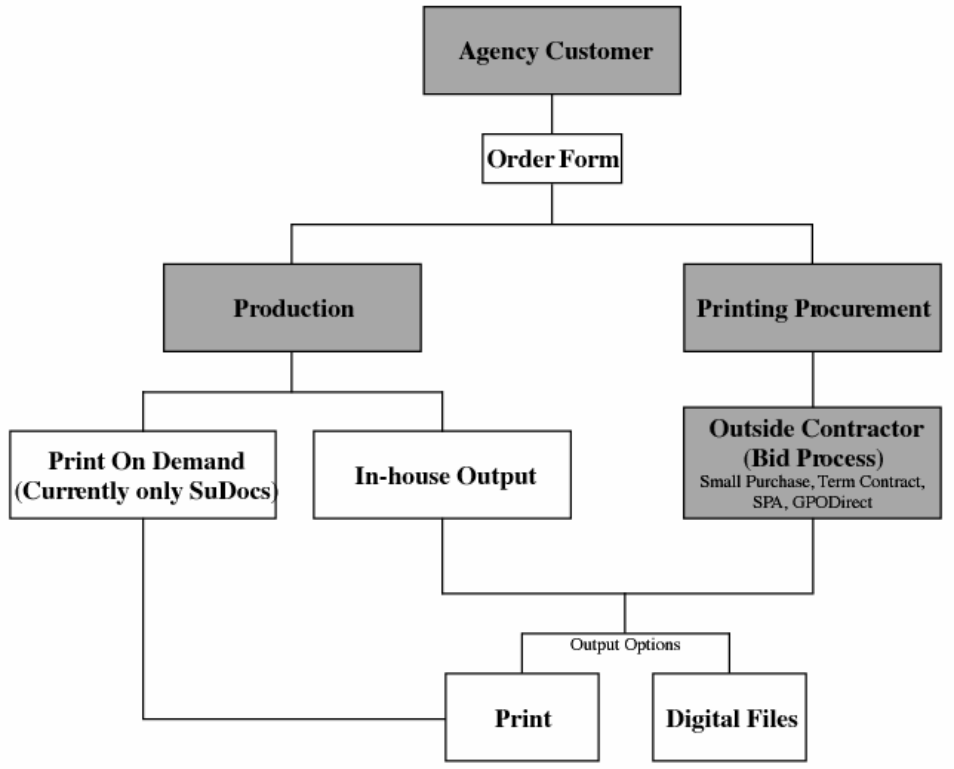


Figure A-6. Hard Copy Production Components

A.1.1.3.1.1.3 Interfaces to Systems and Procedures. Currently customers interface with GPO using various order forms that are filled out manually and submitted to GPO along with the digital files or hard copy that will be used to produce the printed product. The limited exceptions are the pilot program Agency Direct and the Web-enabled mainframe application PICSWEB, which allows registered users to electronically submit Direct Deal term contract order forms (2511). No other automated, online systems for submitting printing specifications are currently available.

The interface between Printing Procurement and outside contractors is the written specification that is created based on the order form and the Disk Information Form 952. These specifications should include all information necessary for producing a printed product. A hard copy of the specifications is provided to the contractor. No automated, online system for submitting printing specifications is currently available.

Figure A-7, Hard Copy Production – Interfaces and Additional Detail, illustrates the hard copy production interfaces.

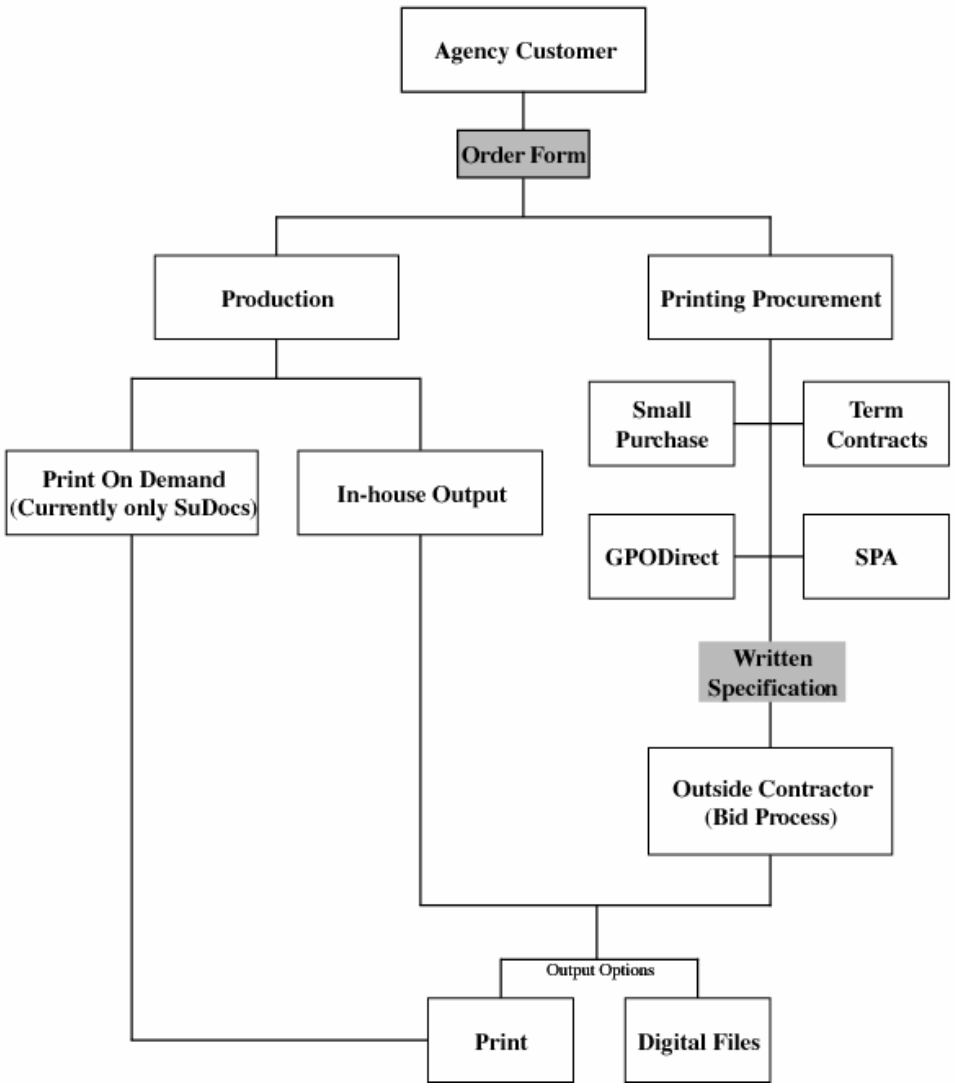


Figure A-7. Hard Copy Production – Interfaces and Additional Detail

A.1.1.3.1.2 Print on Demand. The following subsections describe POD in terms of capabilities, functions, and features, major system components, and interfaces.

A.1.1.3.1.2.1 Capabilities, Functions and Features of the Current System. The capabilities, functions, and features of the repository are currently meeting the basic requirements of storing and retrieving POD digital files. However, due to the lack of a formal CMS, the overall capabilities are quite limited in the areas of data management and user interface. The files reside on the server and are accessed on an as-needed basis.

A.1.1.3.1.2.2 Major System Components and High Level Interconnection. The repository is located in the Datacenter at \\MARS\repository. The files stored on the server are version 1.5 PDFs that contain

raster data, captured as TIFF and stored in the PDF wrapper. The files in the repository range in size from 10's of megabytes to 100's of megabytes. The current size of the entire directory (as of 7/28/2004) was approximately 18GB, and the overall capacity is 250GB. The file names are the sole identifying factor as to the nature of the documents. Through standard naming conventions have been put in place for files within the repository, the processes of ingest, storage, search, and retrieval of documents is simplified. The file name is structured so that it identifies the authoring agency, the ISBN/ISSN of the publication, and the date of file creation.

The repository exists according to procedures established by GPO's IT Organization that address the Datacenter. Authorized users access the repository via Windows XP and Mac OS X operating systems. These users have full access to the files, allowing them to create, change, move, and delete all files on the server.

A.1.1.3.1.2.3 Interfaces to Systems and Procedures. System interactions and workflows associated with creating, naming, storing, and retrieving files on the repository follow very simplistic models at this point. As the server resides on the network, authorized users push and pull content as needed. For example, when a press optimized PDF file is created and/or acquired as part of the POD program, authorized users in Plant Operations, ID, and Customer Services load the file onto a "new files" folder within the repository.

GPO's ePUB Customer Services then pulls the file and performs any preflight work that needs to be done. Once the file is pre-flighted, it is then moved into a "finished files" folder. As the current capabilities of the system only lend themselves to simplistic processes, any flowchart that could be developed for the current process would be incomplete and therefore has not been created.

A.1.1.3.2 Soft Display. The following sections describe soft display in terms of capabilities, functions, and features, major system components, and interfaces. The capabilities, functions, and features of the soft copy display jobs are very similar to those of hard copy jobs, as are the major system components and interconnections and the interface to systems and procedures.

A.1.1.3.2.1 Accessibility. In 1998, President Clinton signed the Workforce Investment Act into law. This Act amended Section 508 of the Rehabilitation Act of 1973 to include accessibility requirements for electronic and information technology. Section 508 now requires that electronic and information technology used by the executive branch of the Federal Government, including their Web sites, be made as accessible for people with disabilities as it is for people without disabilities. The Federal Access Board outlined 16 specific accessibility standards implementing Section 508 as it applies to Web pages. The deadline for meeting these standards was June 21, 2001. Although Section 508 currently only applies to executive branch agencies, GPO is committed to ensuring that new and existing *GPO Access* Web pages along with pages on hosted sites meet the 16 accessibility standards outlined in Section 508. Therefore, GPO is currently working to ensure that existing pages on *GPO Access* and hosted sites are Section 508-compliant, and that future pages will be created specifically with these accessibility standards in mind. However, GPO does not currently have a standard policy for approaching 508 compliance.

While a variety of documents may be created for online purposes, Adobe's PDF has become the standard for online documents and is requested by many customers. There is no standard requirement for making online PDF documents created by GPO's Production facilities compliant. Scanned documents may or may not be OCR scanned in order for searching and accessibility to be possible. Whether or not accessibility is built into a PDF depends on the customer agency, time restrictions, and the individual creating the document. When 508 is incorporated, it is usually limited to Acrobat's MakeAccessible Plug-In. This tool creates a tagged PDF file that may be read by a screen reader. However, a number of limitations exist to the current technology:

- Font Encodings: The MakeAccessible Plug-In program maps fonts used in the document to a Unicode value. In some cases, a font encoding used within a PDF document does not contain enough information for the Plug-In to map to a Unicode equivalent.
- Table Recognition: Simple table recognition is supported; however, complex tables and tables that do not include complete rules around each cell may not be recognized as table content.
- Complex Graphics: Documents containing complex graphics (especially vector graphics) may take a long time to process.
- Complex Regions: The Plug-In may not be able to ascertain the appropriate logical read order from documents containing complex layouts (such as magazines, newspapers, etc.).
- Layer Order: Layers of objects within a complex PDF may be rearranged in a different stacking order, resulting in a tagged PDF that does not look identical to the original.

When any of the above issues occur, additional work must be done to create a 508-compliant file. Alternately, a text file may be created and placed online in addition to the PDF.

Web sites created by GPO typically are 508-compliant. However, exceptions are made depending on the customer agency, time restrictions, and the individual creating the Web site. When specifically requested by the agency, Web sites are made 508-compliant. Also, when agency-created files are submitted to GPO for placement on a Web site, any current accessibility features are retained. All the Web pages on *GPO Access* are made 508-compliant by following the 16 Web standards outlined by the Access Board.

Printing Procurement only incorporates language into its contract specifications requiring 508 compliance when agencies specifically request it. This language typically requires only that the contractor guarantee that all files comply with Section 508 of the Rehabilitation Act. It does not usually specify how compliance is to be attained or validated.

The following subsections describe accessibility in terms of capabilities, functions, and features, major system components, and interfaces.

A.1.1.3.2.1.1 Capabilities, Functions, and Features of the Current System. Section 508 states that when an agency claims undue burden and exempts a project from Section 508, the basis for undue burden must be documented and an alternative means of access is to be provided to individuals with disabilities. GPO leaves the responsibility of creating alternative methods of access to the individual agencies. GPO also strongly recommends that agencies not claim undue burden.

While GPO has standard methods in place for implementing 508 for in-house personnel, no official policy is in place. Sometimes documents are not made 508-compliant due to time constraints or the individual creating the document. ID has the ability to go back to digital documents that do not incorporate 508 and make them accessible (or require that they be made accessible by the contractor or customer agency.)

A.1.1.3.2.1.2 Major System Components and High Level Interconnection. All sections of Production creating web documents and Web sites should incorporate handicap accessibility features. Printing Procurement contracts and specifications should have a standard way of indicating 508 requirements for outside vendors bidding on jobs. The various groups of people in these GPO departments, Federal agencies, and outside contractors have varying knowledge of what 508 compliance is and how it is implemented. System components are depicted in **Figure A-8, Accessibility System Components for Soft Copy**.

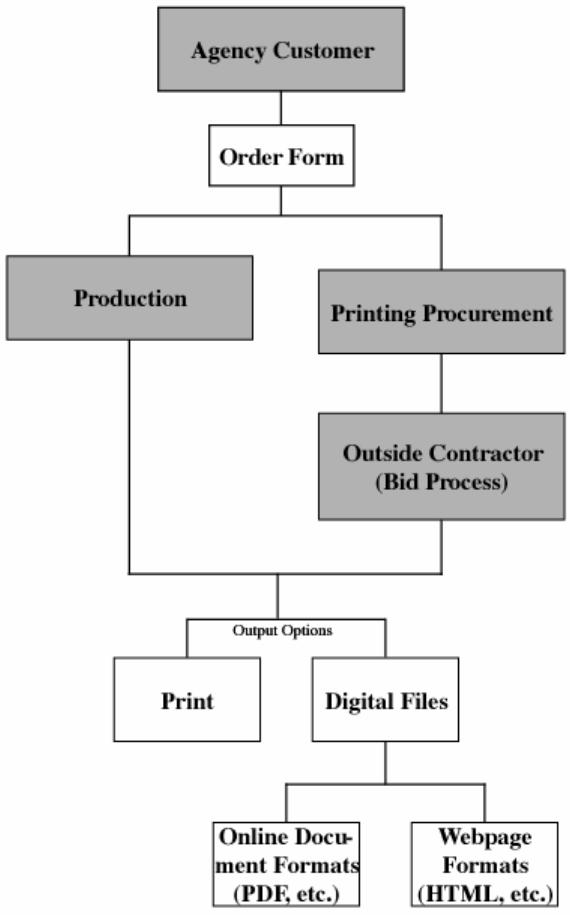


Figure A-8. Accessibility System Components for Soft Copy

A.1.1.3.2.1.3 Interfaces to Systems and Procedures. The interface to systems and procedures of accessible soft copy is similar to that of normal soft and hard copy jobs since 508 compliance information is seldom included.

The interface between Printing Procurement and outside contractors is the written specification that is created based on the order form and the Disk Information Form 952. These specifications usually do not indicate what accessibility should be incorporated into the returned digital files unless specifically required by the agency. System interfaces are depicted in **Figure A-9, Accessibility Interfaces to Systems and Procedures**.

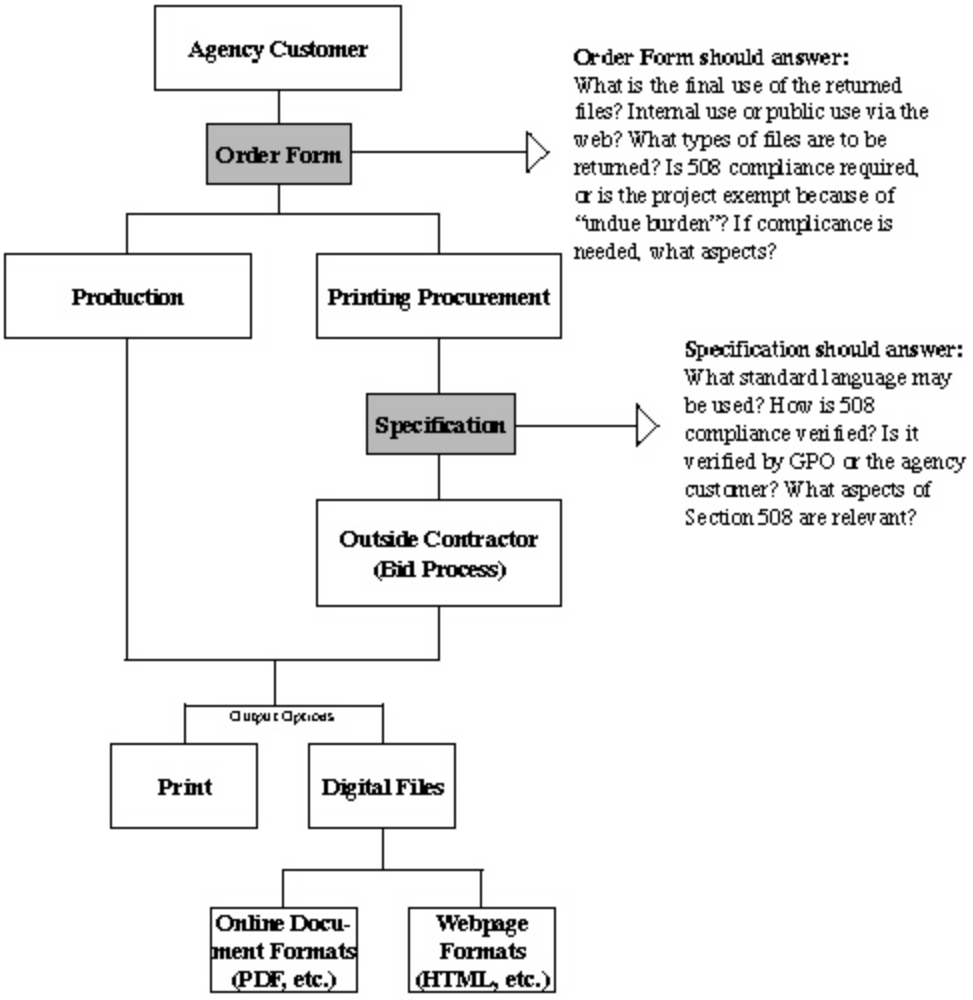


Figure A-9. Accessibility Interfaces to Systems and Procedures

A.1.1.3.3 Digital Media. The following sections describe digital media in terms of capabilities, functions, and features, major system components, and interfaces.

A.1.1.3.3.1 Capabilities, Functions, and Features of the Current System. GPO's use of CDs allows for an almost universal way of transmitting data using hard media. However, the CD also has two big weaknesses as a preservation format. The long term life of the CD is unknown, but potentially very limited. In addition, the data may be encoded using proprietary formats that cannot be migrated to future media. Exploration into future types of media is necessary but currently very limited. In addition, DVD authoring and dissemination capabilities are currently limited. Other then the occasional files sent via FTP or e-mail, an electronic method of sending/receiving files directly to/from agencies and contractors without hard media does not currently exist.

A.1.1.3.3.2 Major System Components and High Level Interconnection. Agency customers, GPO, and GPO's affiliated contractors are the major components currently using various types of hard media. Each component contains multiple groupings of people with various equipment and levels of expertise. These groups utilize media in various ways for viewing, storing, and sharing information. Different groups also have varying levels of access to data stored on hard media. Some are only able to

view data contained on hard media while others can copy, alter, and store that information. Major system components are presented in **Figure A-10, Digital Media Production Components**.

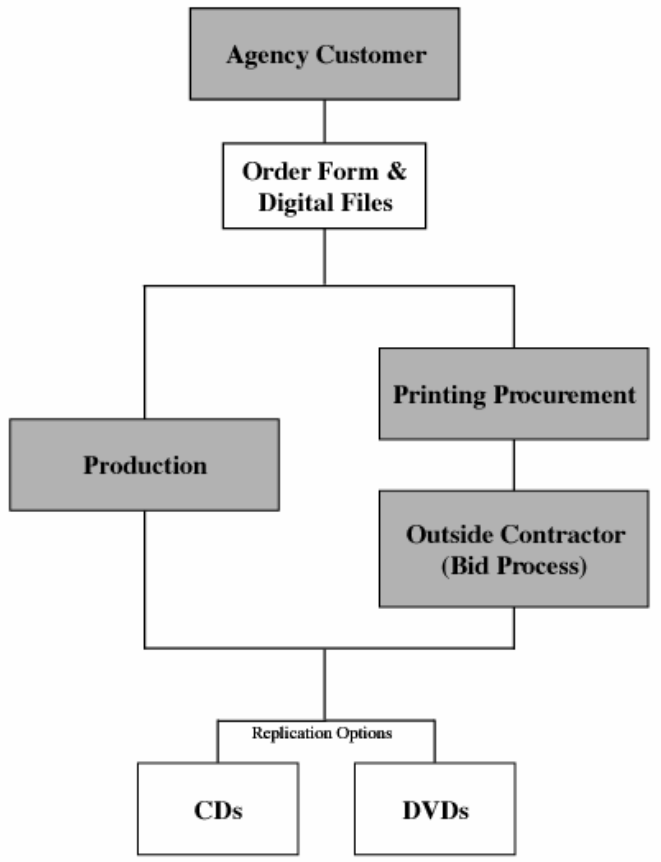


Figure A-10. Digital Media Production Components

A.1.1.3.3.3 Interfaces to Systems and Procedures. Digital media is typically submitted manually to GPO via an overnight delivery service. If the media is to be sent to an outside contractor, it is sent from GPO via an overnight delivery service or the contractor can pick up the materials from GPO's mailroom. Customers using Direct Deal term contracts and the pilot program GPODirect submit media directly to the contractor. Occasionally Printing Procurement specifications will state that the customer will submit digital media directly to the contractor. In addition, digital files are sometimes transferred over FTP or e-mail from the agency customer to either GPO or the contractor. Interfaces are depicted in **Figure A-11, Digital Media Interfaces to Systems and Procedures**.

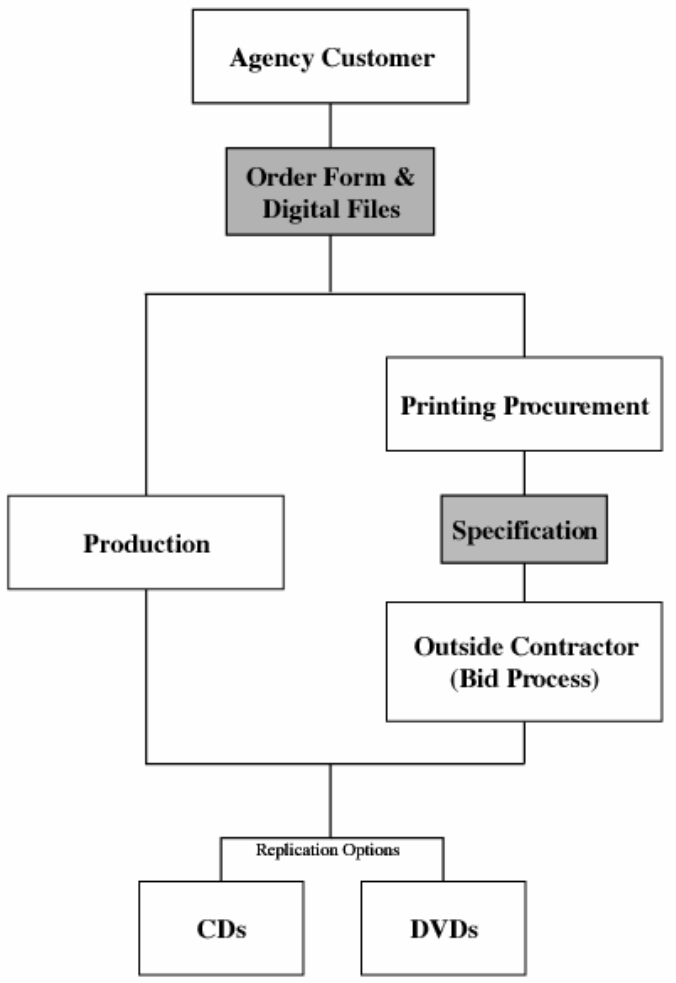


Figure A-11. Digital Media Interfaces to Systems and Procedures