GPO's Future Digital System:

System Releases and Capabilities

Version 4.0

FDsys Reference Document

December 18, 2006

Document Change Control Sheet

Date	Filename/version #	Author	Revision Description
08/24/2005	*Releases and Capabilities, v1.0*	FDsys team	First Draft for P&S review
08/30/2005	*Releases and Capabilities, v1.1*	Gil Baldwin	Version with matrix corrections
09/09/2005	*Releases and Capabilities, v2.0*	Mike Wash	Added Change / Configuration Chart
09/28/2005	*Releases and Capabilities, v2.0*	Lisa LaPlant	Changed Version Control per Comments from the Team Review of the Draft Version Control Specification.
03/31/2006	*Releases and Capabilities v3.0*	Gil Baldwin	Revision to Complement RD 2.0
04/03/2006	*Releases and Capabilities v3.0*	Kate Villano	Formatting
11/20/2006	*Releases and Capabilities v4.0*	*Isaac Jones*	*Revision to Complement RD 3.0*
11/29/2006	*Releases and Capabilities v4.0*	*Isaac Jones*	*Updates from first review of v4.0 draft*
12/18/2006	*Releases and Capabilities v4.0*	George Barnum	Final mechanical & copy edit

1.0 Introduction 1
1.1 System Purpose 1
1.2 System Scope 1
1.3 System Overview 1
1.4 System Releases 2
2.0 General System Description 3
3.0 FDsys REQUIREMENTS 7
3.1 SYSTEM, GENERAL 7
3.2 CONTENT METADATA 7
3.3 CONTENT PACKAGES 8
3.3.1 Submission Information Packages (SIP) 8
3.3.2 Archival Information Package (AIP) 9
3.3.3 Access Content Package (ACP) 9
3.3.4 Dissemination Information Package (DIP) 10
3.4 CONTENT PROCESSING 10
3.4.1 Pre-ingest Processing 11
3.4.2 Ingest Processing 12
3.4.3 Preservation Processing 12
3.4.4 Unique Identifier 13
3.4.5 Persistent Name 13
3.4.6 Authentication 14
3.4.7 Version Control 14
3.5 INFRASTRUCTURE 15
3.5.1 Workflow 15
3.5.2 Storage Management 15
3.5.3 Security 16
3.5.4 Enterprise Service Bus 16
3.5.5 Data Mining 16
3.6 CONTENT SUBMISSION 17
3.6.1 Deposited Content 17
3.6.2 Converted Content 17
3.6.3 Harvested Content 18
3.6.4 Style Tools 18
3.6.5 Content Originator Ordering 18
3.7 CONTENT ACCESS AND PROCESSING 19
3.7.1 Accessibility 19
3.7.2 Search 20
3.7.3 Request 20
3.7.4 Cataloging and Reference Tools 20
3.7.5 User Interface 21
3.7.6 User Support 21
3.8 CONTENT DELIVERY AND PROCESSING 22
3.8.1 Hard Copy Output 22
3.8.2 Electronic Presentation 22
3.8.3 Digital Media 22

1.0 Introduction

1.1 *System Purpose*

The U.S. Government Printing Office (GPO) Future Digital System (FDsys) will ingest, authenticate, provide version control, preserve and provide access to digital content from all three branches of the U.S. Government. FDsys is envisioned as a comprehensive, systematic, and dynamic means for preserving digital content free from dependence on specific hardware or software. The system should automate many lifecycle processes for digital content and make it easier to deliver content in formats suited to customers' needs.

1.2 *System Scope*

FDsys is unparalleled in scope. Included in the FDsys will include all known Federal Government documents within the scope of GPO's Federal Depository Library Program (FDLP). This content will be authenticated and catalogued according to GPO metadata and document creation standards. Content may include text and associated graphics, video, audio, and other forms of content that emerge. Content will be available for Web searching and Internet viewing, downloading and printing, and as document masters for conventional printing, on-demand printing, or other dissemination methods.

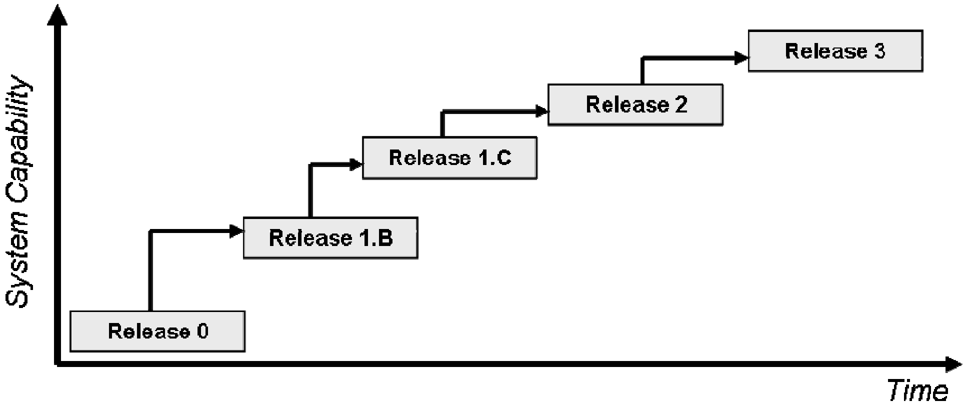
1.3 *System Overview*

FDsys will allow federal Content Originators to create and submit content that can then be preserved, authenticated, managed and delivered upon request. FDsys will automate many content lifecycle processes and make it easier to deliver digital content in formats suited to customers' needs. The system design will be based on the Reference Model for an Open Archival Information System (OAIS) (ISO 14721:2003) **.**

1.4 *System Releases*

FDsys will be implemented in a series of sequential releases. Each release includes improvements to system capability and underlying infrastructure, and is built incrementally on those preceding it .

FDsys Schedule of Releases



Release 0 – Completed March 2006
Supporting Digital Conversion Services for GPO Access

Release 1.A
Initial functionality *

Release 1.B – Target Date May 2007
Internal Prototype for Beta Testing

Release 1.C - Target Date Late 2007
First Public Release

Release 2 - Target Date Mid 2008
Enhanced Access and Capabilities

Release 3 – Target Date Late 2008
Enhanced Collaborative Tools

(** As a result of the procurement process leading to an award of a contract for a Master Integrator, the capabilities for Release 1.A were combined with Release 1.B.*)

2.0 General System Description

In order to meet GPO’s strategic goals, FDsys should be able to accomplish the following:

- Support GPO’s content submission, content processing, 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 managed by GPO;
- Accept/ingest content in a variety of complex formats;
- Accommodate future digital formats;
- Preserve digital content for future use;
- 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.

To meet the challenges of today and the future, the system should be able to

- Accept content in a wide variety of formats with the flexibility to easily adapt to future file formats;
- 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 a manner that is consistent with current technology and the changing expectations of GPO’s diverse user communities;
- Identify the essential characteristics of the content that is being preserved for the purposes of authentication and certification.

Table 1. Capabilities, Functions, & Features by Release (through Release 3)

System Function		Release 1B	Release 1C	Release 2	Release 3
System, General		Meets requirements for design & performance			
Content Metadata		Collects, edits, and shares			
Content Packages	*SIP*	Creates and accepts for ingest			
	AIP	Creates			
	ACP		Creates		
	DIP	Creates			
Content Processing	*Ingest Processing*	SIP validation and ingest			
	Preservation	Accepts AIPs into repository		Preservation processing	
	Unique Identifier	Assigned to all objects			
	Persistent Name	Assigned at ingest and entered in Metadata			
	Authentication		Integrity marks added to delivered content as required		
	Version Control		Accepts version information	Version triggers enabled	

Capabilities, Functions, & Features by Release (through Release 3)

System Function		Release 1B	Release 1C	Release 2	Release 3
Infrastructure	*Workflow*	Initiate	Enhanced as required	Enhanced as required	Enhanced as required
	Storage	Establish storage architecture	ACP storage added, Expanded as required	Expanded as required	Expanded as required
	Security	Enabled			
	ESB	ILS integration	Expanded integration		
	Data Mining			Complete functionality	
Content Submission	*Deposited Content*	Accepts Content			
	Converted Content	Accepts Content			
	Harvested Content	Accepts content		Harvester implemented	
	Style Tools				Implemented
	Content Originator Ordering		Supports job ordering, Service provider database	Full functionality	
Content Access and Processing					
	Accessibility (Sec. 508)	System Components Compliance	Assess content for Accessibility	Creates Accessible content as required	
	Search	Basic functionality	Enhanced functionality		
	Request	No-fee content delivery enabled			Fee-based content delivery enabled

Capabilities, Functions, & Features by Release (through Release 3)

System Function		Release 1B	Release 1C	Release 2	Release 3
Content Access and Processing (cont.)	*Cataloging & Reference Tools*	Integration with ILS		Basic reference tools	
	User Interface	GUIs as required and initial workbench developed	GUIs as required	Customization and Personalization	
	User Support	Context-Specific Help and Alerts		Help Desk, Training and Contact Management	
Content Delivery & Processing		Basic functionality	Enhanced functionality		
	Hard Copy	Basic functionality	Enhanced functionality	Enhanced functionality	Enhanced functionality
	Electronic Presentation	Basic functionality	Enhanced functionality	Enhanced functionality	Enhanced functionality
	Digital Media	Basic functionality	Enhanced functionality	Enhanced functionality	Enhanced functionality

3.0 FDsys REQUIREMENTS

This section provides a general description of the FDsys requirements; for a complete list of the requirements refer to *Requirements Document, v. 3.0.* The descriptions are organized into the six clusters: Content Access, Content Delivery, Content Preservation, Content Processing, Content Submission, and Infrastructure, plus overall system requirements, content package descriptions, and content metadata.

In the *RD* each requirement is identified by the release in which we anticipate its implementation (Release 1B, 1C, 2, and 3)its criticality:

- Must: The system cannot function without meeting this requirement. This requirement must be implemented in the Release listed.
- Should: Added functionality system users will expect. These are desirable features that will be implemented in the Release listed, whenever possible.
- Could: Additional functionality that is not critical to the system function or user experience.

3.1 *SYSTEM, GENERAL*

Core capabilities inherent to all areas of the system in order to ensure interoperability. The system will use open standards to ensure interoperability into the future. The system will be infrastructure independent, modular, policy neutral, scalable, extensible, comprehensive, and flexible.

3.2 *CONTENT METADATA*

Actions or processes in the Future Digital System use and create information about target content. This information is recorded, stored, and subsequently used as content metadata. Content metadata is a structured representation of information that facilitates interpretation, management, and location by describing essential attributes and significant properties of content. Generally, content metadata describes how, when, and by whom a particular content package was collected, what the content is, where it resides, and how it is formatted.

Content 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 (e.g., 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 processing, access, and use.

Content metadata is generally classified in the following broad categories, according to its function:

- Descriptive - such as bibliographic information describing, classifying, and characterizing the identity and context of the content.
- Administrative - describing rights, source, ownership, provenance, conditions of use and business rules.

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

GPO will adopt the most current version of the Metadata Encoding and Transmission Standard (METS) as the encoding standard for content packages in the system.

It is important to make the distinction that the FDsys requirements describe content metadata and how it will behave within the system, but do not address the use of Business Process Information and system metadata. These metadata types are described in the glossary and in other appropriate parts of the Requirements Document.

3.3 *CONTENT PACKAGES*

3.3.1 Submission Information Packages (SIP)

The Submission Information Package (SIP) contains target digital objects and associated descriptive and administrative metadata. It will be the vehicle whereby content packages are submitted to FDsys by Content Originators. It is necessary that a SIP follow established rules so that FDsys can validate and accept the content for ingest.

Associated with the SIP are three types of information:

- Content information (digital object(s) and Representation Information),
- Packaging information, and
- Descriptive information.

Packaging Information is the information that binds or encapsulates the Content Information. To accomplish this, a SIP will include a binding metadata file (sip.xml) that relates the digital objects and metadata together to form a system-compliant SIP. The Metadata Encoding and Transmission Standard (METS) schema shall be adopted as the encoding standard for the sip.xml file, and GPO will specify profiles for METS to drive its implementation for FDsys.

Descriptive information is the metadata that allows users to discover the Content Information in the system.

All file components of the SIP will be populated within a structured file hierarchy and aggregated into a single file or entity for transmission and ingest into the system.

3.3.2 Archival Information Package (AIP)

Archival Information Packages (AIPs) are preservation copies of digital objects with associated technical, descriptive, and preservation metadata. AIPs will be stored in a secure environment and acted upon by FDsys preservation processes to enable permanent public access to the official version(s) of U.S. Government publications in digital formats.

Associated with the AIP are four types of information:

- Content information (digital object(s) and Representation Information)
- Preservation Description Information (PDI)
- Packaging information
- Descriptive information

Preservation Description Information (PDI) is the information needed to accurately describe the Content Information and provide an understanding of the environment in which the Content Information was created. The PDI includes several types of additional information that are needed to help preserve the Content Information. These are:

- Reference: How users can uniquely identify the Content.
- Provenance: Who has had custody of the Content Information and what was its source.
- Technical environment
- Context: How the Content Information relates to other information objects.
- Fixity: Information and mechanisms used to protect the Content Information from accidental change.

Packaging information is the information that binds or encapsulates the Content Information and Preservation Description Information for transmission.

Descriptive information is the metadata that allows users to discover the Content Information in the system.

An AIP is composed of target digital object(s) and metadata about the digital object(s), and a binding metadata file (aip.xml) that relates the digital objects and metadata together to form a system-compliant AIP. The Metadata Encoding and Transmission Standard (METS) will be the encoding standard for the aip.xml file, and GPO will specify profiles and extension schema for METS as required.

3.3.3 Access Content Package (ACP)

Access Content Packages (ACPs) are internal system copies of digital objects with associated content metadata to support access and delivery. The ACP may include access copies, as-ingested version(s) of files, and optimized copies of content to facilitate and optimize access and delivery to End Users. ACPs will follow the concept of content packages outlined in the OAIS reference model. In addition, ACPs will address GPO's business needs including:

- Timely and efficient access to official Federal Government information through search, cataloging, and reference tools;
- Delivery of content and metadata in a way that meets Content Originator and End User expectations for structure, format, and presentation as specified through Content Originator ordering and End User request.

The ACP is created as part of ingest processing and may be modified a result of preservation processing and access processing. ACPs will be stored in high availability / high access storage, , to enable timely search and retrieval. The system must have the capability to send ACPs to delivery processing for creation of DIPs which are then delivered to users.

The ACP consists of digital objects and content metadata about the digital objects, including rich descriptive information to facilitate access. The ACP may also include a binding metadata file that relates the digital objects and content metadata together to form a package. The Metadata Encoding and Transmission Standard (METS) schema has been adopted for the SIP and AIP and may be used as the encoding standard for the binding metadata file, if a binding metadata file is required by the system.

3.3.4 Dissemination Information Package (DIP)

Dissemination Information Packages (DIPs) are transient copies of digital objects, associated content metadata, and business process information that are delivered from the system to fulfill End User and Service Provider requests and Content Originator orders. DIPs will follow the concept of a content package outlined in the OAIS reference model.

The DIP should include all digital objects and/or metadata necessary to fulfill requests and orders. The DIP may also include a binding metadata file that relates the digital objects and metadata together to form a package. The Metadata Encoding and Transmission Standard (METS) schema has been adopted for the SIP and AIP and may be used as the encoding standard for the binding metadata file, if a binding metadata file is required.

3.4 *CONTENT PROCESSING*

FDsys content processing identifies, manages, and verifies digital content as it moves through the system, from creation to dissemination and archiving. Content processing consists of pre-ingest processing, ingest processing, access processing, preservation processing and delivery processing.

Pre-ingest processing prepares content for ingest into the system.

Ingest processing compares submitted content to established criteria, and either accepts the content and creates initial Access Content Packages and Archival Information Packages or rejects it.

Access processing facilitates the finding, analyzing, ordering, and retrieving content and content metadata. During access processing, the system shall execute and manage the following functions:

- Manage ACPs
- Cataloging and reference tools
- Assign persistent names to content packages

Delivery processing facilitates the transfer from the stored form of a digital object in a repository to a user. During delivery processing, the system shall execute and manage the following functions:

- Create DIPs for Service Providers and other users.
- Create pre-ingest bundles (PIBs) from content in pre-ingest WIP to support the publisher approval process (e.g., proofing).
- Apply accessibility processes to create DIPs compliant with GPO accessibility policies.
- Apply integrity marks to DIPs to create packages compliant with GPO authentication policies.

Preservation processing facilitates the maintenance of publications for use, either in their original form or in some verifiable, usable form.

3.4.1 Pre-ingest Processing

FDsys pre-ingest processing includes the processes necessary to identify, manage, and verify digital content prior to entry into the system.

Pre-Ingest processing manages the functions that prepare content for ingest into the system. Content Originators and Service Specialists have the capability to submit content to WIP storage. Content can be submitted from deposited processing, Content Originator ordering, style tools, conversion processes and harvesting. The system will assign unique identifiers, identify versions, detect duplicate content, and allow for publisher approval processes. Pre-ingest processing performs the following functions:

- Version control
- Content Originator ordering
- Assign unique IDs to content
- Assign unique IDs to system jobs
- Scope assessment, per the Information Dissemination Scope Determination policy.
- Integrity checking on content
- Accessibility assessment

- Style tool, non-style tool, converted content, harvested content processing to create a SIP
- Publisher approval processes (i.e., proofing) to move content to ingest processing.

3.4.2 Ingest Processing

FDsys ingest processing includes functions to identify, manage, and verify digital content as it moves into the system.

Ingest processing manages content and content metadata as it is received into the system as a Submission Information Package (SIP). Content Originators and Service Specialists will have the capability to submit SIPs created from deposited, harvested, and converted content and content created using GPO style tools. Ingest processing creates AIPs and ACPs from SIPs and transfers the resulting content packages to storage. During ingest processing, the following functions are performed.

- Accept and validate SIPs
- Create AIPs from SIPs
- Create initial ACPs from SIPs
- Apply digital time stamping to content

3.4.3 Preservation Processing

FDsys preservation processes will enable comprehensive, timely, permanent public access to the official version(s) of U.S. Government publications in digital formats. Only content in scope for GPO's dissemination programs will be accepted into FDsys archival storage and managed by preservation processes.

Preservation copies of digital publications, Archival Information Packages (AIPs), with associated technical metadata, will be maintained in FDsys Archival Storage. During preservation processing, the following functions are performed.

- Manage AIPs through refreshment, migration, and emulation.
- Manage ACPs to ensure ongoing consistency with AIPs.
- Create DIPs from AIPs.

In order of preference, the outcomes desired are:

- Faithfully duplicated files rendered using the original application.
- Files which faithfully reproduce content, behavior and presentation of the original, rendered using other software than the original application.
- Files which exactly convey the content but may alter behavior and/or presentation; rendered using other software than the original application.

Although digital preservation is an emerging discipline, GPO expects to employ such strategies as:

Refreshment (copying) of content to new media. Refreshment is the systematic transfer of stored digital information to newer, fresher media.

- Migration of data in formats or versions that are in danger of becoming or have become obsolete to newer versions of that application or format. Migration is a process in which the underlying information is retained but older file formats and internal structures are replaced by newer;
- Emulation preserves the essential behaviors and attributes of digital objects by using current software to mimic the original environment;
- Hybrids of these approaches or new approaches.

The preservation process employed in any given situation should be the least intrusive; i.e. that which alters the original AIP the least.

3.4.4 Unique Identifier

The system will create and assign unique IDs to content as defined by GPO business rules. All assigned unique identifiers will be recorded and used in metadata. Once assigned, a unique ID cannot be reused within the system.

- Digital Objects: A unique ID will be assigned to all digital objects upon ingest into the system.
- Content Packages: A unique ID will be assigned to Content Packages (SIP, ACP, AIP)
- Jobs: A unique ID will be assigned to Jobs.

3.4.5 Persistent Name

In order for the digital content managed by FDsys to be easily found and shared by a wide range of users, there must be a system for reliably and unambiguously identifying each resource independent of its location.

Persistent naming allows for an interoperable system of identifiers that uniquely identify content, support permanent access to that content, and support access to information about the content. A resolution system will locate and provide access to content and metadata associated with assigned persistent names.

The system will assign persistent names to content packages at ingest. Once assigned, a persistent name cannot be reused within the system.

3.4.6 Authentication

The content authentication functional element will assure users that content made available by GPO through FDsys is authentic and/or official. This includes identifying content that has been approved by, contributed by, or harvested from an official source such as a Federal publishing agency, its business partner, or other trusted source. GPO generally defines its products as official if the content was issued by the United States Government at Government expense or as required by law.

Content authentication will help GPO establish a clear chain of custody for deposited, harvested, and converted content that is ingested into the system Content authentication will assure users that content is authentic meaning that it has been verified by GPO to be complete and unaltered when compared to the version approved or published by the Content Originator.

The system will verify content integrity by assuring users that content has not been altered in an unauthorized manner. The system will verify content integrity at various points throughout the content lifecycle. If content is modified, the content authentication functional element will have the ability to notify designated users when, where, by whom, and what changes were made to content. The system will have the capability to certify content at both the document and granular levels, and certification will be conveyed to users through the use of integrity marks such as digital signatures and watermarks.

3.4.7 Version Control

Version control in FDsys will evaluate and establish the version of a piece of content and subsequently track it through its entire life cycle. Version control will be called upon to analyze Content Packages and assign the appropriate version identifier, consistent with requirements for version triggers and chain of custody.

Users, including all categories in the FDsys User Class model, want to be certain that they are using the version of information that meets their needs and to be able to track the history of changes that may have occurred. In the case of Federal information, multiple versions of Government publications may be available on public Web sites. This can be confusing and potentially damaging to users who are not aware of the version of the content. Version control is a necessary operation in the management and dissemination of digital content to ensure that users are accessing the appropriate or desired content.

Version control is a critical function of FDsys for which GPO will define what constitutes a unique manifestation of a publication across all publication formats (e.g., monograph, serial).

GPO envisions that the process of version control will include acquiring, cataloging, storing, preserving, indicating relationships among, and retrieving different versions of content. This process may be accomplished by assessing various document attributes (e.g., structure, content, and format), creating metadata about these attributes, monitoring changes to the attributes, updating metadata to indicate changes to the attributes, and creating links to related documents. The version control process within the FDsys will be automated whenever possible, but subjective evaluation and interpretation by service specialists may be a critical requirement at various points through the process.

3.5 *INFRASTRUCTURE*

3.5.1 Workflow

Workflows are utilized in the FDsys to automate business processes. The system will provide capabilities to define, execute and monitor the workflows at various granularity levels. The system will provide GUI tools for users to perform workflow management tasks.

3.5.2 Storage Management

Storage management will 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.

Storage Types

- Failover Storage - Separate storage location to allow access to all data in the event of an emergency with primary storage.
- Back-up Retrieval Media Storage - Off-site backup of critical data.
- Mid-term Archival Storage - Moderate capacity of offline storage with archival capabilities for at least 10 years.
- Long-term Permanent Archival Storage - Large capacity of offline storage with archival capabilities for at least 100 years.
- Content Delivery Networks – External networks used to deliver high-demand content without overloading GPO equipment.

Storage Categories

- Work In Progress Storage (WIP)
- Archival Information Package Storage (AIS)
- Access Content Storage (ACS)
- Business Process Information Storage (BPS)

The storage architecture is designed to be as flexible as possible within cost constraints. The storage architecture for FDsys must also be highly scalable in order to continue ingesting content. Requirements for storage architecture are based on Storage Categories rather than Storage Types in order to maintain this flexibility.

Storage Type	**WIP**	**AIS**	**ACS**	**BPS**
Access Time	< 2sec.	< 2sec.	<2 sec.	< 2sec.
Failover Storage	No	Yes	Yes	Yes
Back-Up Retrieval Media Storage	No	Yes	Yes	Yes
Long-Term Permanent Archival Storage	No	Yes	No	No

3.5.3 Security

The security functional element provides the appropriate confidentiality, integrity, and availability functions for FDsys information and processes. It also governs access to content (both authentication and authorization), assigning user rights (authorization), and maintaining system security (administration and auditing). Finally, the security element provides mechanisms for the necessary technical, operational, and management controls for FDsys, including interfaces that it will have with other systems.

3.5.4 Enterprise Service Bus

The system shall consist of many internal individual functional elements (i.e. services), each specializing in a business functional area. The system shall also provide the capability to interact with external applications. The concept of the Enterprise Service Bus (ESB) is the preferred approach and shall be employed to facilitate flexible and scalable integrations between the services and applications.

The system shall provide the capability to plug-in services or applications deployed in different hardware and software platforms. The interoperability is facilitated by the underlying integration infrastructure – the ESB. The system shall provide the capability to add, replace or remove service components declaratively via configurations in XML. The system shall provide the administrative GUI tool to manage the integrated internal and external service components.

The ESB is a relatively new technology in the enterprise integration field. It is standards based, depending heavily on XML, and related Extensible Stylesheet Language Transformations (XSLT), XPath and XQuery technologies. Because of its flexibility and capability to enable a highly scalable system, it has become a preferred approach to build the Service-Oriented Architecture in enterprise applications.

3.5.5 Data Mining

Data mining consists of the tools and processes for the extraction, analysis, and presentation of business process information (BPI), content metadata, and system metadata to enhance internal and external business efficiencies. BPI is administrative, non-content specific information that is used within the business process and package description to support access aids and data mining. Content metadata is descriptive, technical, structural, administrative, and preservation information about content. System metadata is data generated by the system that records jobs, processes, activities, and tasks of the system.

GPO will provide intuitive data mining capabilities, including access to selected external data repositories (e.g., Oracle). The data mining functional element will need to extract and analyze information from all GPO Systems.

FDsys 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), subject to privacy and legal restrictions. The ability to track monetary transactions will also be required.

The data mining resources of the FDsys will allow for the following:

- Extracting BPI in multiple formats from the entire collection.

- Normalizing data based on administrator defined parameters (e.g., identify missing values or metadata, data formats, types and discrepancies, anomalies).
- Performing multirelational analyses on BPI (e.g., cross tabulations, categorization, clusterization, regression analysis, data patterns and relationships).
- Presenting BPI according to user preferences and GPO business rules (e.g., views based on access levels, exporting of results, linking of results to data).
- Mining BPI within the system at multiple levels of aggregation and granularity (e.g., Service Provider performance history, customer agency billing information, ordering habits, preferences of customers and users).
- Predicting future trends (visualization capability) in order to adjust workflow or anticipate demand.

3.6 *CONTENT SUBMISSION*

Content submission accepts digital content and creates compliant SIPs for ingest into the system. Digital content includes:

- Deposited content: content intentionally submitted to GPO by Content Originators
- Harvested content: content within the scope of dissemination programs that is gathered from Federal agency websites
- Converted content: digital content created from a tangible product

Content submission includes toolsets for creating, collaborating on, and approving content. These are referred to as style tools.

Content submission also includes an order function. Content Originators may submit content, order and re-order content, and specify delivery of content and GPO services through Content Originator ordering.

3.6.1 Deposited Content

Deposited content is content intentionally submitted to GPO by Content Originators. The Submission Information Package (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 will identify and employ best practices for preparing and submitting deposited content, including metadata to capture all the customers' requirements. FDsys must be able to accept all content submitted by Content Originators, including content furnished in proprietary formats. FDsys must be able to assemble content into a compliant SIP for ingest into the system.

3.6.2 Converted Content

Converted content is digital content created from a tangible product. Tangible publications are defined for products such as ink-on-paper, microforms, CD-ROM, or DVDs, characterized by content recorded or encoded on a physical substrate. The digital collection created from this process will be made available for permanent public access through GPO's dissemination programs. In addition to GPO's efforts, the agency will continue to work with various user communities including Federal agencies, the Library of Congress, National Archives and Records Administration (NARA) and the library community on digitizing a comprehensive collection of legacy materials.

In addition to traditional scanning, other techniques of digitization currently exist and could evolve in the future. There may also be instances in which a successful conversion and/or Optical Character Recognition (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).

GPO recognizes that non-text based formats also exist in the legacy collection. These formats include analog audio and video. Specifications will be developed on a case-by-case basis for the creation of these files.

The desired outcome of the conversion process will be to produce a Submission Information Package (SIP) that includes the electronic preservation master files and submission level metadata that will be ingested into FDsys.

3.6.3 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 harvest in-scope content, and will collectively be referred to as the "harvester" in this document.

The harvester will consist of discovery, assessment, and harvesting tools. The discovery tools will locate electronic content from targeted 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. The assessment tool also identifies the applicable relationships between versions. The harvesting tool gathers content and available metadata.

3.6.4 Style Tools

Style tools will allow Content Originators to prepare content in pre-ingest processing. The goal of style tools is to move GPO upstream in the content origination process. Style tools accept content and provide composition, collaboration, and approval tools.

3.6.5 Content Originator Ordering

Content Originator ordering is a system interface to FDsys that allows Content Originators to submit content, order and re-order content, specify content delivery, and request other service options. It will provide the capability to create, capture, augment, and store agency processing requirements specific to ordering functions, preservation needs, version, and job specifications (e.g., SF1, 952, 2511, 3868). In addition, Content Originator ordering will allow users to discover the cost of job and fulfillment options, select fulfillment choices, and discover payment/billing status when applicable. Service Providers will use the interface to interact, deliver, and report upon order status. Service Specialists will use the interface to manage the ordering process. In addition, the system shall support the ability for Service Specialists or Content Originators to add additional copies (riders) to a request or order. Content Originator ordering will pass content to pre-ingest processing, notify Content Evaluators when job are placed, and integrate with GPO's. Context specific help and support will be accessible through the interface.

3.7 *CONTENT ACCESS AND PROCESSING*

Content access and processing provides the services and functions that allow users to determine the existence, description, location and availability of content, and request delivery of content and metadata. In addition, content access and processing allows for the management of Access Content Packages and user interaction with the system.
Content access and processing includes information about:

- Search – Performing queries on content and metadata so that content can be retrieved from storage and delivered to users
- Request - Processing no-fee and fee based content delivery requests
- Cataloging - Creating descriptive metadata that conform to accepted standards and support access and delivery of standard bibliographic records
- Reference tools - Creating lists and resources that assist users in locating and accessing content
- User interface - Developing and managing user interactions with the system
- User support - Providing answers to user inquiries and directing users to content and services
- Accessibility – Providing content and system accessibility for persons with disabilities

Under legal authority of Title 44, Chapters 17, 19, and 41 of the United States Code (U.S.C.), 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), International Exchange Service, GPO Sales Program, By-Law programs, and the GPO Access public Web site. The FDLP distributes electronic and tangible publications to a network of over 1,250 Federal Depository libraries across the country. GPO is able to provide these publications to depository libraries for no-fee through a congressional appropriation. Select publications are also available for sale to the public via the GPO Sales Program, including through the U.S. Government Bookstore.

3.7.1 Accessibility

FDsys will provide the capability to create, assess, and validate content packages for compliance with Section 508 technical standards. In addition, FDsys components and technologies shall comply with Section 508 technical standards. Section 508 refers to a statutory section in the Rehabilitation Act of 1973, which is codified in 29 U.S.C. 794d. In 1998, President Clinton signed the Workforce Investment Act of 1998, which amended Section 508 of the Rehabilitation Act of 1973 to provide access to and use of Federal executive agencies' electronic and information technology (EIT) by individuals with disabilities. Section 508 also requires Federal executive departments and agencies that develop, procure, maintain, or use electronic and information technology to ensure that Federal employees and members of the public with disabilities have access to and use of information and data, comparable to that of the employees and members of the public without disabilities, unless it is an undue burden to do so.

3.7.2 Search

Search executes queries on content and metadata so that content can be retrieved from storage, processes, and delivered to users. FDsys search tools should meet or exceed industry standards for search and retrieval technology. As necessary, more than one search tool may be used to meet the needs of all user classes who will be searching the system. The FDsys search tools must handle user searches of content and metadata both simultaneously and separately across multiple internal repositories. Search must have the ability to search multiple media, file formats, and levels of granularity. Search should produce a highly relevant, organized, usable, and detailed results list that provides the location and description of content. Search tools should provide innovative methods for users to access information related to their query. Search must include accessible and customizable graphical user interfaces that allow all users to submit and refine queries, filter results, and export results sets.

3.7.3 Request

Request will allow users to request delivery of content and metadata from FDsys. Request must have the capability to handle no-fee and fee-based delivery requests. An example of a no-fee request for delivery is a Public End User downloading a PDF document that is within scope of the Federal Depository Library Program. An example of a fee-based request for delivery is a Public End User using a shopping cart function to order publications from an e-commerce Web site. For fee-based content, request must provide the capability for End Users to discover the cost of delivery, choose delivery options, and submit payment for delivery. In addition, request must provide the capability for GPO and external Service Providers to request delivery of content packages for the purpose of content processing and delivery. Request must ensure that customer transactions can be conducted in a secure environment. Request must have the ability to interact with GPO systems or other Authorized Representatives for a variety of services, including but not limited to financial and inventory control systems. Request must provide the capability for users to manage and securely store information in user accounts such as order histories, preferences for delivery options, and preferred payment methods.

3.7.4 Cataloging and Reference Tools

Cataloging

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. In FDsys, cataloging tools create descriptive metadata that conform to accepted standards, and support access to and delivery of standard bibliographic records.

The cataloging process creates applicable descriptive metadata elements, including metadata 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).

Reference Tools

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

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.).

Lists, in the context of reference tools, may be static pages produced from report generation capabilities, or dynamic results lists from searches. These searches may be pre-configured (“canned”) or individually created for one-time use.

3.7.5 User Interface

The user interface functional element will allow for the management of user interactions with the system. Graphical User Interfaces (GUIs) and workbenches (sets of available tools) are key components of this functional area. A workbench will be created for each user class and GUIs will be created for each functional element as required in accordance with the release schedule. Workbenches for internal and external user classes must allow users to access toolsets and perform authorized functions. The system must have the capability to provide default workbenches that do not require users to log-in or register with the system. Users who opt to register with the system will gain the ability to customize GUIs and workbenches, and receive personalized services. The default public End User workbench must provide the capability for users to access official Federal Government information without registering with the system.

3.7.6 User Support

GPO has a strong commitment to provide superior customer service and user support. This commitment spans from assisting Content Originators at the stage of content creation to providing services that assist users in using GPO’s diverse array of tangible and electronic products. User support will provide answers to user questions and direct them to content and services. User support services include a helpdesk and knowledge base, interactive training, real-time alert services, and services that provide the capability for users to receive personalized support based on their stored preferences. User support will also be provided in conjunction with the public End User interface and will provide the capability for users to submit personal information to the system. End Users will not be required to submit personal information, however it may be needed to provide some user support features. User support will provide the capability to submit personal information to the system for all user classes. User support will be provided to all users that interact with the system. This may include answering inquiries and resolving customer complaints as well as providing any technical assistance needed for the online bookstore.

3.8 *CONTENT DELIVERY AND PROCESSING*

Content delivery encompasses the delivery of pre-ingest bundles (PIBs) and Dissemination Information Packages (DIPs). PIBs contain digital objects, business process information and metadata required for service providers to output proofs and produce end products or services. DIPs contain digital objects, business process information and metadata based to facilitate user requests.

Transformation and assembly processes will take place in delivery processing. Access Content Packages (ACP) will be transformed into DIPs and PIBs will be assembled. Archival Information Packages (AIPs) will be transformed into DIPs as necessary for preservation by other organizations. Digital objects may be adjusted based on user requests to support the delivery of hard copy, electronic presentation and digital media.

3.8.1 Hard Copy Output

Hard copy output is tangible printed content (e.g., ink on paper) produced from digital files. Hard copy output may be requested as an Access request or through the Content Originator ordering user interface. Content Originator's will include information on the desired output such as color attributes, trim sizes, binding preferences. For Content Originator ordering, hard copy output will be generated from PIBs. DIPs will be used to generate hard copy output based upon request and Content Originator re-orders.

3.8.2 Electronic Presentation

Electronic presentation output is the dynamic and temporary representation of content in digital format on End User devices, including computers and non-desktop electronic devices. Electronic presentation encompasses presenting images, text, video, audio and multimedia in electronic form.

3.8.3 Digital Media

Digital media is a content delivery mechanism consisting of data storage devices. The digital media component of FDsys includes the delivery of content for storage on the following:

Removable data storage devices (e.g., CD, DVD)

Multifunctional/handheld devices (e.g., PDA, MP3 players, e-books)

Storage at user sites (e.g., servers, personal computer)

Duplication/replication of removable digital media will be available through internal and external Service Providers. The system will determine how to deliver content to support storage on digital media.

Content may be pushed to a user's multifunctional device, or requested and pulled from the system. The system will determine how to deliver content to the user's device and offer options for delivery to those devices, when options are available.