GPO'S FEDERAL DIGITAL SYSTEM:

SYSTEM RELEASES AND CAPABILITIES

VERSION 5.0

FDSYS REFERENCE DOCUMENT

DECEMBER 19, 2007

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
12/19/2007	*Releases and Capabilities v5.0*	Liz Pruszko	Revision to Complement RD 3.2

1.0 Introduction

1.1 *Document Purpose*

The U.S. Government Printing Office (GPO) has created this document to provide an overview of the capabilities as they are implemented in releases of the Federal Digital System (FDsys). As a crucial document that serves as a high level overview of the system functions according to a planned implementation schedule, the document will be revised periodically to reflect program changes and updates. This revision was updated in anticipation of the next release, Release 1C, which is the first public release of FDsys. In describing R1C and the sub-releases, system functionality has been defined by features and features have been aggregated into feature sets. This approach is new to this revision and should provide a more complete overview of expected system functionality in R1C. Additional information on FDsys, including detailed documentation and presentations can be found by going to http://www.gpo.gov/projects/fdsys.htm.

1.2 *System Purpose*

The U.S. Government Printing Office (GPO) Federal 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 will automate many lifecycle processes for digital content and make it easier to deliver content in formats suited to customers' needs.

1.3 *System Scope*

FDsys will include all known Federal Government documents within the scope of GPO's Federal Depository Library Program (FDLP) and other information dissemination programs. 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.

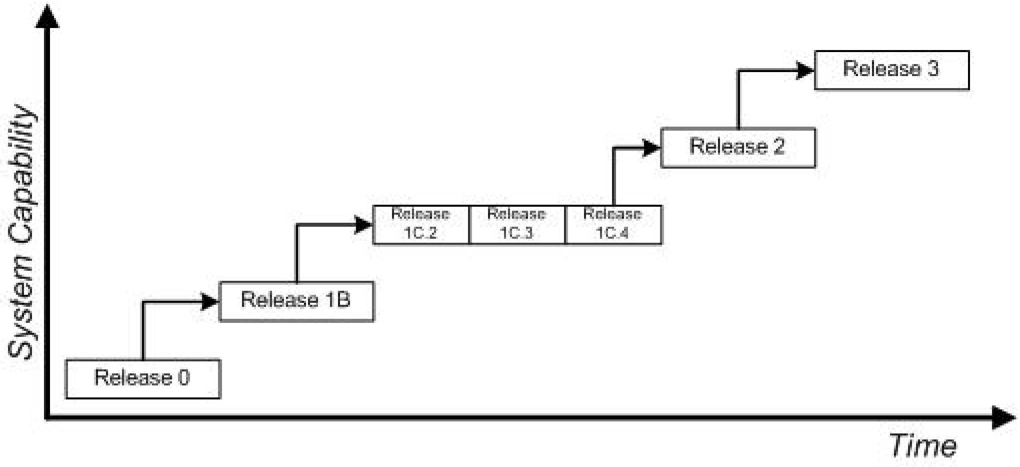
1.4 *System Overview*

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).

1.5 *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.

1.6 *Schedule of Releases*



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

Release 1A
Initial functionality *

Release 1B – Completed September 2007
Internal Prototype/Proof of Concept for Beta Testing

Release 1C.2 - Target Date November 2008
Basic Features of First Public Release

Release 1C.3 - Target Date Spring 2009
Additional Features of First Public Release

Release 1C.4 - Target Date Fall 2009
Final Features of First Public Release

Release 2 - Target Date TBD
Enhanced Access and Capabilities

Release 3 – Target Date TBD
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 will 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 will be able to do the following:

- 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.

3.0 Release 1B

3.1 *Overview*

Completed in September 2007, Release 1B is an internal release of FDsys as a proof-of-concept to test the core functionality that future releases will build upon. It supports FDsys formal and informal beta testing. The FDsys beta version was released to limited internal and external audiences to ensure that the foundation for FDsys is operating as expected.

3.2 *Capabilities*

Development was centered on internal workflows, building the external search infrastructure with advanced search technology and content and job submission. The core public user search feature was built on the concept of simple search / advanced results. The concept uses a single search box to submit queries and navigational elements to help users hone in on desired documents within the search results. Release 1B also focused on deposited content, which is born digital content submitted by publishing agencies. The content in Release 1B comes from three sources: the 2006 Federal Register, Congressional Bills of the 109th Congress and select agency publications.

Other characteristics include:

- Limited metadata recording
- Submission Information Packages (SIPs), Access Content Packages (ACPs), and Dissemination Information Packages (DIPs) created
- Limited Pre-Ingest Processing
- Limited Ingest Processing
- Unique IDs assigned to all digital objects
- Basic storage architecture
- Basic security enabled
- One-way ILS integration by ESB
- Unique ID and password user authentication
- Deposited content submission
- Simple search
- Simple help

4.0 Release 1C

4.1 *Overview*

Release 1C has been split into three consecutive sub-releases. Features in feature groups were used to analyze the functionality to be implemented in this release.

4.1.1 Release 1C.2

Release 1C.2 will build on lessons learned from Release 1B and include the key functionality of a public release. This includes scaling the system infrastructure and building a digital repository that conforms to the OAIS reference model and enables the management of content and metadata. This release will replace the familiar WAIS-based GPO Access in use since 1994, with full functionality for the top 25 collections and limited functionality for all other content.

4.1.2 Release 1C.3

Release 1C.3 will expand on 1C.2 by adding the capability for submission of Congressional material, enabling persistent names, and providing enhanced functionality.

4.1.3 Release 1C.4

Release 1C.4 will provide additional enhancements to functionality added in 1C.2 and 1C.3 including submission of Federal agency material.

4.2 *Schedule of Features*

The following table indicates when development will occur on different features in Release 1C.

Feature Group	**Feature**	**Release 1C**		
		R1C.2	**R1C.3**	**R1C.4**
Metadata Management	Authenticity Metadata	X		X
	Content Metadata	X		X
	Descriptive Metadata	X		X
	ILS Integration	X		X
	Mandatory AIP Metadata	X		X
	Mandatory SIP Metadata	X		
	Schema Registry	X		
	Technical Metadata	X		X
	Version Control	X		X
	Rights Metadata			X
OAIS Compliance	ACP Creation	X		X
	Delete Packages	X		X
	Digital Time Stamping	X		X
	Format Identification	X		X
	Ingest	X		X
	Package Structure	X		X
	Unique ID	X		
	Package Management			X
GPO Access	508 Compliant User Interfaces	X		

	Browse Content	X	X	
	Checking/Reformatting Content for 508 Compliance		X	
	Create Persistent Links	X		
	Deliver Content and/or Metadata	X	X	
	Delivery By Email	X		
	Delivery By FTP	X		
	Format Transformation	X		
	Maintain *GPO Access* Capabilities	X	X	
	Maintain PDF Features	X		
	Manage Public Access	X		
	PURL/Getdoc/Getpage/Getcfr Resolution	X		
	Search	X	X	
	Search Results	X		
	Support Granularity	X		
	User Help	X	X	
	User Notifications	X		
	User Interface	X	X	
Infrastructure	Authenticate User	X		X
	COOP	X		X
	Enterprise Service Bus	X		X
	Event and Audit Logging	X		X
	Group Enforcement	X		
	Manage Security	X		X
	Network Security	X		
	Security Administration	X		X
	Security Monitoring	X		
	Storage Management	X		X
	Storage Replication	X		
	System Availability	X		
	System Backup/Restore	X		X
	System Flexibility	X		
	System Monitoring	X		
	System Performance	X		
	User Registration	X		
	User Roles	X		
	Workflow Management	X		
	Data Mining	X		X
Congressional Submission	Authorized User Search	X	X	
	Automatic Scope Determination		X	
	Batch Submission of Content and Metadata		X	
	Congressional CO Submit Content and Metadata	X	X	X
	Duplicate Detection	X	X	
	Internal Service Provider Submit Content and Metadata	X	X	X
	Manage BPI		X	
	Manage Work in Progress		X	
	Pre-Ingest Processing		X	
	Render Job BPI on GPO Forms		X	

	Submit Congressional Jobs		X	X
	Track Job		X	
	Deliver Content and/or Metadata to Authorized Users	X	X	X
	508 Compliant User Interfaces for Authorized Users	X	X	
Content Submission	Submit Jobs			X
	Submit Content and Metadata			X
	Enter BPI and Metadata			X
	Duplicate Detection			X
	Pre-Ingest Processing			X
	Render BPI on GPO Forms			X
	Deliver Content and/or Metadata to Authorized Users			X
	Scope Determination			X
Persistent Names	Persistent Name Assignment		X	X
	Persistent Name Resolution		X	
Access	508 Compliant User Interfaces			X
	Browse Content			X
	Create Persistent Links			X
	Checking/Reformatting Content for 508 Compliance			X
	Delivery by RSS			X
	Deliver Content and/or Metadata			X
	Delivery By Email			X
	Delivery By FTP			X
	Follow Relationships to Other Documents			X
	Format Transformation			X
	Maintain *GPO Access* Capabilities			X
	Maintain PDF Features			X
	Manage Public Access			X
	Partnerships			X
	Save Search Query			X
	Search			X
	Search Interface for External Systems			X
	Search Results			X
	Support Granularity			X
	User Help			X
	User Notifications			X

4.3 *Metadata Management Feature Group*

Actions or processes in the Federal 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.

4.3.1 Authenticity Metadata

The system must be able to record provenance information that is sufficient to ensure the authenticity of an object to users and to aid in the preservation of that object. This means that the system must record in metadata information about content submitted to the system and its source, ingest to the system, file transformations, and other changes to content or metadata. This includes providing a record in metadata for all validation steps, including virus check, metadata validation, message digest calculation, ingestion, derivative creation, and fixity check. Some amount of provenance information must be available to end users, although authorized users may have differing levels of access to the information. The system will be able to determine whether deposited content is "official" and communicate that status to users. The system will also record whether the content is digitally signed at submission. This feature includes the recording of intended use information and the maintenance of digital signatures and their display to end users.

4.3.2 Content Metadata

This feature gives users the ability to add, modify, or delete content metadata. Those modifications, if made to the ACP need to be mirrored in the AIP, and vice versa. All metadata changes or additions will be made through graphical user interface (GUI) forms in this release.

4.3.3 Descriptive Metadata

Descriptive metadata, including bibliographic metadata, describes, classifies, and characterizes the identity and context of the content. Descriptive metadata will be accumulated by capture from processes, by discernment or assignment by processes or Service Specialists, or by acquisition from an external source. Users will be able to view descriptive metadata. Descriptive metadata for every publication in a *GPO Access* collection, as defined in the RD, must be expressed in Metadata Object Description Schema (MODS) for indexing in the ACP cache.

4.3.4 ILS Integration

Full integration with the ILS's Aleph product is defined as the bidirectional exchange of data between the two systems. Bibliographic metadata in MARC format is required for one category of access. The ILS is the tool for creating those records and maintaining the database, including holdings information. When new bibliographic records are created, updated, or deleted in the ILS, the descriptive metadata is updated in the AIP, ACP, and ACP cache.

4.3.5 Mandatory AIP Metadata

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

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.

This feature requires the AIP to contain a descriptive metadata file for the package in MODS format with some minimum metadata elements. These elements are mandatory for the Submission Information Package (SIP) to be ingested by FDsys. This information will indicate the metadata source, the title of the publication, the agency publisher, the date of ingest, date of creation, the file size, the MIME type, the content package unique ID and whether the content is in scope for GPO's dissemination programs. For each

rendition in a content package there will be a corresponding metadata file specifying technical parameters of the content file.
For tangible items, the location of the tangible item and the source from which content was converted are also mandatory. And for harvested content the URL and date of capture are mandatory.

4.3.6 Mandatory SIP Metadata

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

This feature requires the SIP to contain a descriptive metadata file for the package in MODS format with some minimum metadata elements. These elements are mandatory for the SIP to be ingested by FDsys. This information will indicate the metadata source, the date of content creation, the location of the tangible item (if it exists) and the URL and date captured for harvested content.

4.3.7 Schema Registry

The schema registry records attributes of standards and schemas/DTDs in use in the system for the purpose of verifying that metadata is well-formed and valid with respect to its standard. In 1C the supported input standard is MARC. The supported schemas are METS, MODS, MARCXML, and PREMIS.

4.3.8 Technical Metadata

The system will create, accept, and manage technical metadata sufficient for preservation actions and for archival management. This includes format-related metadata ascertained during the format identification process. The system shall record information about relationships, agents, events, and the attributes of digital objects, as defined in the PREMIS data dictionary.

4.3.9 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 used to analyze Content Packages and assign the appropriate version identifier, consistent with requirements to record version information and chain of custody. This will also apply to content that is part of a serial or series.

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 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.

4.3.10 Rights Metadata

In some cases, FDsys will need to record rights information in order to properly preserve and provide access to digital objects. Rights metadata will be recorded in PREMIS (Preservation Metadata: Implementation Strategies). The PREMIS standard contains a rights entity that allows the association of rights with specific digital preservation actions. Works that are in the public domain can be preserved without obtaining permission, and there are no limits to the types of modifications or transformations that can be made of the item.

This feature provides that FDsys will employ metadata which relates the rights information of a target digital object(s) and its associated content package, including copyright information. Authorized users will be allowed to modify access rights to content based on copyright information provided by Content Originators.

4.4 *OAIS Compliance Feature Group*

The Open Archival Information System (OAIS) reference model 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. Using the OAIS as a reference model, the system should provide OAIS foundation services 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.

This feature group encompasses functionality needed to fulfill the high-level design concept of FDsys as a preservation repository which is used to enable current and permanent access to content. Implicit is the use of the information packages and rigorously protected archival storage. GPO's adoption of the OAIS reference model in the FDsys design is intended to provide the technological environment for a trusted digital repository.

4.4.1 ACP Creation

Access Content Packages (ACPs) are 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 Access Content Storage and the ACP cache, high availability / high access storage, to enable timely search and retrieval while protecting the Archival Information Packages in a preservation repository instance. 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.

This feature governs how ACPs are created and managed, including the addition of renditions to ACPs. At creation, an ACP contains all renditions from AIP, except renditions created to aid in preservation (e.g., XML). ACPs shall contain all metadata from the AIP and can replicate AIP structure for ephemeral content and screen, press, or print optimized renditions will not have a corresponding AIP.

4.4.2 Delete Packages

Because OAIS employs a package based approach, the system must be able to delete packages when necessary. This includes the deletion of renditions from an AIP, entire AIPs, renditions in an ACP and entire ACPs.

4.4.3 Digital Time Stamping

Digital time stamps establish authenticity and fixity of digital objects entering and being managed within the system. They consist of a checksum of all the content in the AIP and the date and time of ingest. This means that the system will apply a digital time stamp to content when received, ingested, new renditions are created and all renditions are deleted. The system time used in digital time stamping will be controlled by the network time.

4.4.4 Format Identification

Format identification is the process of determining the format to which a digital object conforms. Format validation is the process of assessing whether a digital object conforms to the format's specification. Validation records anomalies (e.g., invalid page mode in a PDF) and inhibitors (e.g., encryption, password protection). The system will evaluate whether digital objects are well-formed (i.e., whether it meets the purely syntactic requirements for its format). Format characterization is the process of determining the format-specific significant properties of an object of a given format. In this feature, the system will determine file type without using file extension, reference an external file format registry, validate file formats and identify basic characterization of file formats, including usage inhibitors.

4.4.5 Ingest

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.

Ingest, from an OAIS perspective, receives submissions, validates the transfer, performs verification and validation on the content, generates an AIP, generates descriptive metadata, and transfers the AIP to storage. During this phase, a series of validation steps occur, including a virus check, duplicate detection, mandatory metadata evaluation, metadata file validation, format identification, and scope determination. Ingest will include the ability to:

- Validate mandatory metadata files
- Create a SIP
- Create an AIP from an in-scope SIP
- Retain metadata elements from an SIP in the AIP
- Reject SIPs that haven’t been ingested after a configurable period of time
- Verify all metadata files are valid
- Verify that at least one content file is present
- Allow users to reject a SIP
- Check SIPs for viruses or malicious code
- Detect duplicate content
- Determination if content is in scope

4.4.6 Package Structure

METS records list of content files, list of metadata files, order of content files, folder structure, and relationships between content files and metadata files. Subdirectories in a package mirror subdirectories as submitted or as manually changed by a Service Specialist. Each content file will be related to one or more metadata files and each metadata file shall be related to one or more content files. Each package will contain at least one content file and at least one metadata file.

4.4.7 Unique ID

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.

Unique IDs will be assigned to each SIP, content file, granule, metadata file and job. Unique IDs and recorded in the metadata and link corresponding packages. The UID is unique, does not contain I and O, is human readable, is alphanumeric, and is expressible in XML ID.

4.4.8 Package Management

This feature helps to control the management of packages, including the SIP, AIP, ACP and DIP. SIPs for harvested content will contain one or more rendition consisting of the original harvested digital objects. While the SIP for Converted Content is required to contain renditions produced by the conversion process and include information describing the conversion process for that rendition. SIPs can also be aggregated into a ZIP file.

The system will provide the capability for authorized users to add renditions of a publication to an AIP. The ACP will include copies of renditions from its corresponding AIP based on business rules. Additionally, the system shall provide the capability to include metadata files as required supporting access and delivery. FDsys will be able to produce DIPs which are interoperable with other OAIS-based repositories.

4.5 *GPO Access Feature Group (Releases 1C.2 and 1C.3)*

Under legal authority of Title 44, Chapters 17, 19, and 41 of the United States Code (U.S.C.), GPO's Library Services & Content Management 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), Cataloging and Indexing (C&I), 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.

4.5.1 508 Compliant User Interfaces

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.

This feature requires FDsys GUIs to be section 508 compliant with Section 508 Web-based intranet and internet information standards according to 36 CFR Part 1194.22.

4.5.2 Browse Content

Browse content requirements provide the capability for external End Users to browse content. For 1C.2, this includes browse by collection, table of contents, and date issued. For 1C.3, it includes browse by collection specific metadata elements.

4.5.3 Checking/Reformatting Content for 508 Compliance

For Release 1C.2, the system will ingest content as it is currently available on *GPO Access*. For 1C.3, the system will provide the capability for Service Specialists to manually check and correct content to enable Section 508 compliant content to be delivered from the system. The system will allow a Service Specialist to record that a 508 complaint rendition has been created.

4.5.4 Create Persistent Links

For 1C.2, this feature provides for persistent, predictable links migrated and day forward *GPO Access* content in FDsys, including granules as necessary. A link to a rendition or granule should not stop working after a session (e.g., close browser).

This feature also provides for internal linking of publications at all levels of granularity to support current *GPO Access* functionality, including the Federal Register Table of Contents. The system will provide the capability to link citations in digital objects to the appropriate rendition of a version of a publication. Internal linking refers to the ability to link a piece of text within a document to another digital object.

4.5.5 Deliver Content and/or Metadata

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.

For 1C.2, this feature will support delivery of migrated and day forward *GPO Access* content. The system will deliver content and metadata separately and as packaged DIPs via links on the public user interface. The system will support delivery to Mac and Windows platforms. The primary End User delivery formats for 1C.2 are HTML/text, PDF, and XML when it has been submitted on behalf of the Content Originator and has been approved for public release. In addition, some migrated and day forward *GPO Access* content will be available in Excel and JPEG formats. Metadata will be available in PREMIS, MARC, and MODS. Release 1C.3 will include delivery in XHTML format, delivery to non-GPO storage devices, and the capability to maintain interactive content functionality.

4.5.6 Delivery by Email

In Release 1C.2, public users will have the capability to email a link to a publication.

4.5.7 Delivery by FTP

This feature provides for the delivery of DIPs to authorized users via FTP get and put.

4.5.8 Format Transformation

For 1C.2 an XML preservation rendition will be created for in scope content at ingest. In addition, the system will transform PDF files to HTML/text. It will transform of XML digital objects into other registered XML digital objects, XML metadata into other registered XML metadata, and system metadata to other registered XML metadata, as necessary. Transformations will be performed without deleting the source content, and quality metrics in the form of a “pass/fail” report will be provided for transformation performed by the system. For all other transformation in 1C.2, authorized users will have the ability to manually transform renditions of ACPs.

1C.3 will build upon 1C.2 by providing a PostScript to PDF transformations. At this time, the system will support the transformation of PDFs to PDFs with lower optimization (e.g., press to screen) and PDFs with searchable text (e.g., OCRed text). In addition the system will provide other screen optimized renditions, as necessary and specified for inclusion in the DIP.

4.5.9 Maintain *GPO Access* Capabilities

For 1C.2, the system will ingest content currently available on *GPO Access*. Access to those items will not be restricted or otherwise diminished. This includes continuing to provide the capability for users to print and download content. Files will be segmented into smaller files as necessary to support existing business processes. The system will ingest new day forward *GPO Access* content. The system will migrate existing *GPO Access* GUIs by replacing, updating, or providing new functionality.

In addition for 1C.2, the system will interface with the GPO Automated PDF Signing System (APS) for the application of digital signatures on publicly available PDF files. A non-digitally signed PDF rendition will also be available in the ACP and the AIP. The signed rendition will appear in the public end user search results.

4.5.10 Maintain PDF Features

In 1C.2, the purpose of this feature is to ensure that PDF features such as bookmarks, thumbnails, links, and comments that are added to a PDF prior to submission are not removed by FDsys processes. In addition, PDF post-it note comments will be indexed and searchable via FAST.

4.5.11 Manage Public Access

For 1C.2, this feature provides the capability for GPO to manage access to ACPs. The system will provide public end user access to in scope final published versions of content. It will limit access to content that is out of scope, embargoed, or not approved for public release. This feature provides the capability for authorized users to add new renditions to the ACP. It also synchronizes content and metadata between in the AIP and ACPs.

4.5.12 PURL/GetDoc/GetPage/GetCFR Resolution

This feature ensures that existing links created by customers using PURLs, GetDoc, GetPage, and GetCFR will continue to work once FDsys goes live.

4.5.13 Search

In 1C.2, content in the system will primarily consist of migrated and day forward *GPO Access* files, and the Search feature will include core functionality to enable public users to locate and ultimately retrieve content and metadata from the system. Metadata will be parsed from content in order to provide an enhanced search experience. Users will have the capability to select content collections to search and choose from available search complexity levels (e.g., simple, advanced/fielded, citation, expert Boolean). Users will be able to direct their search against content, metadata, or content and metadata together. Users will also have the ability to search for content granules (e.g., Resume of Congressional Activity in the Congressional Record). Booleans, stemming, proximity searching, synonym searching, and search qualifiers will be supported in 1C.2. The system will also recognize alternate spellings of search terms and suggest corrected spellings for terms. Authorized GPO users will be able to monitor and refine search functionality through a COTS business manager interface. The system will also provide the capability for internet search engines to index publicly available content. 1C.3 will build upon 1C.2 by providing an editable list of stop words and idioms.

4.5.14 Search Results

For Release 1C.2 this feature provides for the configuration, display, and navigation of search results. Users will have the ability to bi directionally sort, filter, and navigate through search results. Users will be able to limit the number of search results displayed per page. Different renditions will be presented under the same search result and results will be provided at the lowest level of granularity supported by the content package. Search results will also include a document summary. Requirements in this feature include provisions for GPO to determine which metadata elements are visible and define search filters (i.e. navigators), and the system will provide for the uniform display of search filters. Release 1C.3 will include the capability to display the total number of results in the result set returned by the search.

4.5.15 Support Granularity

FDsys granularity requirements are necessary to support a large number of *GPO Access* publications that include granular content. Granular content can be described as content that is broken into smaller content units such as chapters, parts, or sections. Stakeholders expect to have the ability to retrieve both the entire publication and a single granule based on the natural boundaries of the publication. They also require the capability to associate granules with the entire publication. In the current state, this is most often accomplished via a HTML table of contents page that includes links to granules. Stakeholders require the ability to retrieve granular content in PDF and HTML/text. PDF granules should be delivered at the page or page range level of granularity and may contain remnants of other granules at the beginning or end of the granule. Release 1C.3 will provide granular content in XML as it is available and approved for public dissemination.

4.5.16 User Help

For 1C.2, FDsys user help requirements are necessary to provide users with assistance on how to use the system. This is accomplished through the use of a help menu and clickable contextual help icons on the GUIs. Additional user help includes providing a web form for users to contact GPO for assistance. Release 1C.3 will build upon 1C.2 by providing context specific help that is displayed when a user points their mouse over an item on the user interface and GPO will have the ability to manage this context specific help.

4.5.17 User Notifications

For 1C.2, FDsys will maintain the existing *GPO Access* listserv notification for the Federal Register. The system will email the daily Federal Register Table of Contents as HTML with content links enabled, and HTML attachment with content links enables, and as text. Public users will continue to be able to sign up to receive the daily Federal Register Table of Content in their email inbox.

4.5.18 User Interface

For 1C.2 standards-based Web pages will be created to support FDsys features. Web pages will allow public users to perform simple, advanced, citation, and expert Boolean search for content and metadata, access specific collections, view search results, view detailed information about content, download DIPs, learn about publications, receive help, and contact GPO, Public users will have the ability to access the system without registering. In addition, users will be able to bookmark search results, content detail pages, and individual collection pages. In order to make XML metadata files more usable, formatting will be applied when they are displayed in a browser.

The intention is for GPO to migrate *GPO Access* to be supported by the functionality that will be provided by FDsys. Web pages will be designed using standards and templates, and authorized GPO users will have the ability to manage static content, images, CSS, and hyperlinks in editable areas on Web-based GUIs via an HTML editor. For Release 1C.2, Web pages will be tested in Internet Explorer 6 for Windows and Firefox 2.0 on Macintosh and Windows.

For 1C.3, FDsys will display lists of publications and browse testing will be expanded to include Safari 2.x on Macintosh; Firefox 1.5.x on Macintosh, Windows, and Linux; Microsoft Internet Explorer 7.x. on Windows; Netscape Navigator 7.x on Windows; and Konqueror 2.x on Linux.

4.6 *Infrastructure Feature Group*

4.6.1 Authenticate User

In Release 1C, the system will provide the ability for users to login and their credentials must be authenticated by a unique User ID and Password. PKI certificate authentication will be utilized in Release 1C.4. The system will able to verify the identity and authority of users and permit them to perform functions granted to them through their user role to content groups to which they have been granted access. Security administrators will be able to create roles.

4.6.2 COOP

The Continuity of Operations feature set intends to describe the different requirements for fulfilling a complete COOP plan in the event of localized problems with a single instance of the system. The requirements define certain conditions that must be met in order to maintain functionality, and reference existing GPO and Federal guidelines that the system must adhere to in reference to COOP.

4.6.3 Enterprise Service Bus

The system will consist of many internal individual functional elements (i.e. services), each specializing in a business functional area. The system will 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 will 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 will provide the capability to add, replace or remove service components declaratively via configurations in XML. The system will 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.

The Enterprise Service Bus will allow interoperability with other services across GPO through business process orchestration, transaction management, and the ability to deploy and manage services. The ESB will utilize common standards for communication, and will be administered via GUIs.

4.6.4 Event and Audit Logging

In Release 1C, the system will keep an audit log of all transactions in the system. These audit logs will contain information about the event that occurred including: date, time, outcome, source, description, and user. Audit logs will be created for all system functions and will be accessed by authorized users. These users will not able to edit any of the data stored with the log and they will be stored and managed according to GPO Publication 825.33.

4.6.5 Group Enforcement

In Release 1C, the system will authenticate users and permit them to perform functions assigned to them through their user role. They will be able to perform these functions to content, Business Process Information (BPI) and metadata that is associated with the group(s) that they have been assigned to. Security administrators will be able to create groups while both security administrators and delegated authorities will be able to assign users to groups. The access to specific group content (WIP, AIP, ACP, BPI, SIP) will be protected from unauthorized user alteration.

4.6.6 Manage Security

In Release 1C, various GPO, federal and industry standards will be used to manage and enforce security. These include the use of LDAP protocols, PKI standards for authentication, as well as hash and digital signature standards. In addition, the system will maintain user privacy, and ensure that private data is kept confidential.

4.6.7 Network Security

In Release 1C, the system shall be able to be operated at a hosting vendor site at GPO's discretion, and the connection between other instances and GPO offices shall be encrypted and mutually authenticated.

4.6.8 Security Administration

In Release 1C, security administrators and delegated authorities will have the capability to create and assign users to roles and groups. User accounts will have the capability to be created, activated, managed, expired, and suspended if necessary. Users will also have the capability to manage specific preferences within FDsys. An interface will exist for all security administrators to manage functions and enforce security policy.

4.6.9 Security Monitoring

In Release 1C, the system will have the capability for authorized security administrators to monitor system security policy settings and enforcement.

4.6.10 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.

Specialized Storage Types

- Failover Storage - Separate storage location replicated from primary storage 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.
- 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 (AIP)
- Access Content Package Storage (ACP)
- Business Process Information Storage (BPI)

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	AIP	ACP	BPI
Response Time	<= 2sec	<= 2sec	<= 2 sec	<= 2sec
Failover Storage	Yes	Yes	Yes	Yes
Back-Up Retrieval Media Storage*	Yes	Yes	Yes	Yes
Long-Term Permanent Archival Storage	No	Yes	No	No

* If architecture requires it

This feature set refers to the overall definition of various storage levels and standards. Archive-related requirements define the repository for the system, and specify preservation integrity for data. Multiple storage classes are also called out, including separate AIP storage locations, as well as failover and offline storage. The set also refers to different storage standards and compatible architectures.

4.6.11 Storage Replication

Storage Replication further defines the failover storage for the system, and specifies the backup and integrity requirements necessary for failover. Failover requirements include the ability to switch over to redundant components at a disaster recovery location in order to survive a localized disaster affecting a single instance of the system. Storage replication also specifies that all system data, content packages and BPI be replicated at the disaster recovery site.

4.6.12 System Availability

The system shall support an average peak time availability of 99.7%.

4.6.13 System Backup/Restore

System Backup/Restore refers to the ability to backup and restore data, applications, and workflow in the event of a failure. This set also defines backup retrieval media storage, and specifies the components necessary for backup availability.

4.6.14 System Flexibility

This high-level feature set refers to the systems ability to accommodate changes to technology, policy, hardware, software, personnel, and location without requiring a major system redesign. System Flexibility also provides open interfaces for the system, and open access to data in an effort to use non-proprietary components wherever possible.

4.6.15 System Monitoring

The System Monitoring feature set defines the monitoring standards that will be implemented for the system, including sensor monitors, alerts/notifications, and the general health of the system.

4.6.16 System Performance

The System Performance feature set refers to the requirements that must be met in order for the system to return requests to the end user in a reasonable timeframe. This feature set also refers to the sizing ability of the system.

4.6.17 User Registration

In Release 1C, users have the capability to create and manage a unique user account to perform functions with FDsys. All users will have the capability to self-register or have their account created from them by a security administrator or delegated authority. All user accounts must be assigned a specific role and group by a security administrator or delegated authority.

4.6.18 User Roles

In Release 1C, the system will able to verify the identity and authority of users and permit them to perform functions granted to them through their user role to groups to which they have been granted access. Users can have multiple roles but will only have a single role per group. The user account will be provisioned and managed using the Oracle Identity & Access Management system and the structure of the account, role, and group will be in an LDAP schema. Internal users will have their LDAP data stored in Active Directory while external users will be stored in an FDsys instance of Oracle Internet Directory (OID). Non-standard user data will be stored in a separate BPI database. FDsys will also have to integrate with the enterprise directory service application (Critical Path Meta Directory) to communicate to the GPO Active Directory.

4.6.19 Workflow Management

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.

The Workflow Management feature set defines the different terminology associated with Workflow, including activities, jobs, priorities, and workflow instances. The feature set ensures that the workflow tool uses an established schema and allows users to implement business rules when defining the workflow tool. Within the workflow tool, users can control the execution of workflow, assign priorities to jobs, and monitor the workflow.

4.6.20 Data Mining

In this release, the system will only maintain current reporting functionality and utilize existing reporting functionality from implemented COTS applications.

4.7 *Congressional Submission Feature Group*

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 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.

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.

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.

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.

4.7.1 Authorized User Search

This feature includes requirements for core search functionality for authorized users, including Content Originators, Service Specialists, and Service Providers. The feature includes queries on content, metadata, and business process information (BPI), access to which is determined by an authorized user's role and group. FDsys authorized 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, metadata, and BPI both simultaneously and separately across multiple internal repositories and databases. 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.

4.7.2 Automatic Scope Determination

FDsys will be able to perform automatic scope determinations on content based on its content, metadata, and BPI. The scope determination can also be made by the access rights and privileges assigned to the user submitting the content. For example, all Congressional Bills will automatically be considered in-scope based on either the metadata or the permissions given to the Congressional CO submitting the Bills. Content such as letterhead and envelopes will be considered out of scope.

4.7.3 Batch Submission of Content and Metadata

The system will be able to accept batch input of multiple files and metadata by Content Originators, Service Specialists and Internal Service Providers. These files could constitute a single publication or multiple publications.

4.7.4 Congressional CO Submit Content and Metadata

FDsys will provide Congressional Content Originators with an interface to electronically submit Congressional Bills, Letterhead, Envelopes, and other Ephemeral material to GPO, along with content metadata related to these publications. Congressional Bills will require its own specialized interface, while all other publications/content will be submitted to GPO through a generic GUI. The generic GUI must be able to allow Congressional COs to enter information that is currently transferred to GPO on SF-1 and pink, white, blue, and red requisitions forms.

4.7.5 Duplicate Detection

The system must be able to detect content that is an exact duplicate to content already within the system. The system will need to detect duplicate content based on the content itself and any associated metadata, including version information, bibliographic information, and authentication information.

4.7.6 Internal Service Provider Submit Content and Metadata

Internal Service Providers will need to have access to all content and job information that is originally submitted to FDsys by a Congressional CO (e.g., Congressional Bills and non-legislative content). Plant Operations will also need to be able to submit new renditions of this content and any additional job information.

Plant Operations will also need to be able to submit all Congressional and OFR content to FDsys that was not originally submitted to FDsys by a Congressional CO. Plant Operations will also need to have access to all congressional and OFR content within FDsys.

4.7.7 Manage BPI

This feature includes requirements that ensure that BPI entered through the system is stored in a database with proper access controls to ensure data is not inadvertently changed.

4.7.8 Manage Work in Progress

This feature includes requirements that enable authorized users (Content Originators, Service Specialists, and Service Providers) to work in a collaborative environment to manage work in progress jobs and content. This includes the capability for users to check in and check out content and jobs, and provide alerts and notifications to users that a jobs and/or content have been checked out.

4.7.9 Pre-Ingest Processing

The system will be able to create and deliver pre-ingest bundles (PIBs) and Dissemination Information Packages (DIPs) to Internal Service Providers (GPO Plant), who will also have the capability to deposit final published versions of content after final approval from the Congressional Content Originators.

4.7.10 Render Job BPI on GPO Forms

The system will need to be able to output jobs onto printed standard forms. For ephemeral material, the system will need to be able to output a printed Pink Requisition form complete with the data that has been entered for each field. For Congressional Bill open requisitions, the system will need to be able to output a printed SF-1 complete with the data that has been entered for each field.

4.7.11 Submit Congressional Jobs

The system will allow Congressional Content Originators and Service Specialists to submit jobs for Congressional Bills and Ephemeral materials. These jobs will be entered through GUIs which will capture BPI and metadata on standard GPO forms. The system will alert and notify users of job activities based upon workflow. Users will be able to add notes and details specific to jobs through the GUIs.

4.7.12 Track Job

The system will need to have the capability to allow users to track jobs they have sent to GPO. GPO users will need to have the ability to manage status information about specific jobs, and Content Originators will need to be able to view this information. Content Originators will need to be able to track their jobs using a Content Originator supplied tracking number.

4.7.13 Deliver Content and/or Metadata to Authorized Users

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 authorized users (CO, Service Specialist, Service Provider) requests and 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.

4.7.14 508 Compliant User Interfaces for Authorized Users

This feature includes requirements that ensure that user interfaces for Content Originators, Service Specialists, and Service Providers conform to Section 508.

4.8 *Content Submission Feature Group*

This feature group includes functionality similar to the Congressional Submission feature group, adapting the features to agency and converted content.

4.8.1 Submit Jobs

The system will allow Congressional Content Originators and Service Specialists to submit jobs for Congressional Bills and Ephemeral materials. These jobs will be entered through GUIs which will capture BPI and metadata on standard GPO forms. The system will alert and notify users of job activities based upon workflow. Users will be able to add notes and details specific to jobs through the GUIs.

4.8.2 Submit Content and Metadata

This feature includes the capability for authorized users to submit content and metadata. This includes the capability for agencies to submit content as well as the submittal of converted and harvested content from Service Specialists or Service Providers.

4.8.3 Enter BPI and Metadata

This feature includes the capability for authorized users to enter BPI and metadata that originally existed on tangible forms, such as the SF-1, 2511, 952.

4.8.4 Duplicate Detection

The system will be able to detect content that is an exact duplicate to content already within the system. This function will happen at Ingest. The system will need to detect duplicate content based on the content itself and any associated metadata, including version information, bibliographic information, and authentication information.

4.8.5 Pre-Ingest Processing

The system will be able to create and deliver pre-ingest bundles (PIBs) and Dissemination Information Packages (DIPs) to Internal Service Providers (GPO Plant), who will also have the capability to deposit final published versions of content after final approval from the Congressional Content Originators.

4.8.6 Render BPI on GPO Forms

The system will need to be able to output jobs onto printed standard forms. For ephemeral material, the system will need to be able to output a printed Pink Requisition form complete with the data that has been entered for each field. For Congressional Bill open requisitions, the system will need to be able to output a printed SF-1 complete with the data that has been entered for each field.

4.8.7 Deliver Content and/or Metadata to Authorized Users

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 authorized users (Content Originator, Service Specialist, Service Provider) requests and 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.

4.8.8 Scope Determination

FDsys will be able to perform automatic scope determinations on content based on its content, metadata, and BPI. The scope determination can also be made by the access rights and privileges assigned to the user submitting the content. For example, all Congressional Bills will automatically be considered in-scope based on either the metadata or the permissions given to the Congressional CO submitting the Bills. Content such as letterhead and envelopes will be considered out of scope.

4.9 *Persistent Name Feature Group (Releases 1C.3 and 1C.4)*

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.

4.9.1 Persistent Name Assignment

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

4.9.2 Persistent Name Resolution

A resolution system will locate and provide access to content and metadata associated with assigned persistent names.

4.10 *Access Feature Group (Release 1C.4)*

The Access feature group provides enhanced functionality to features delivered in earlier releases as part of the *GPO Access* feature group.

4.10.1 508 Compliant User Interfaces

This feature provides for 508 compliance as necessary for multimedia, client side image maps, and server side image maps.

4.10.2 Browse Content

Browse content requirements provide the capability for external End Users to browse content. 1C.4, builds upon earlier releases by providing browse by descriptive metadata elements and topics.

4.10.3 Create Persistent Links

This feature includes requirements for citation linking. The system will provide the capability to link citations in digital objects to the appropriate rendition of a version of a publication. Internal linking refers to the ability to link a piece of text within a document to another digital object. Citation linking will be enabled for select *GPO Access* publication in FDsys as outlined in the Requirements Document.

4.10.4 Checking/Reformatting Content for 508 Compliance

For Release 1C.4, the system will provide the capability for authorized users to flag content for a manual section 508 accessibility check.

4.10.5 Delivery by RSS

FDsys will allow users to sign up to receive DIPs via RSS.

4.10.6 Deliver Content and/or Metadata

Release 1C.4 will include the delivery of additional formats for both content and metadata. The system will deliver metadata in Dublin Core and MARC XML. In addition, the system will support delivery to the LINUX platform.

4.10.7 Delivery by Email

FDsys will deliver DIPs and batches of DIPs via E-mail.

4.10.8 Delivery by FTP

Release 1C.4 builds upon 1C.2 by allowing users to request delivery based on user defined criteria.

4.10.9 Follow Relationships to Other Documents

This feature provides the capability for users to navigate from a document to a related document. Users will have the ability to access related content that is part of the Federal legislative, regulatory, and judicial opinion processes. Relationships will be established based on citations and stored in metadata. The Requirements Document specifies the relationships that will be established for this release.

4.10.10 Format Transformation

Release 1C.4 will provided the capability to create one or more access derivative renditions for an ACP if they don't already exist in the AIP. The system will convert native files such as TIFF, Locator, SGML, EPS, JPEG, Photoshop, Illustrator, Windows Metafile, Bitmap, HTML, Quark, InDesign, Word, Excel, and PowerPoint to access formats such as PDF and HTML/text. As part of the transformation process, images will be converted to descriptive text.

4.10.11 Maintain *GPO Access* Capabilities

Release 1C.4 builds upon Release 1C.2 by ingesting the remaining in-scope content on GPO's Federal Bulletin Board, Permanent Server, and *GPO Access* Web servers.

4.10.12 Maintain PDF Features

The system will maintain PDF features (e.g., bookmarks, comments, links, thumbnails) when individual PDF files are combined into a single PDF file.

4.10.13 Manage Public Access

Release 1C.4 builds upon 1C.2 by limiting access to content with re-dissemination restrictions, limited distribution content, and copyrighted content. It also provides the capability for an authorized user to prevent an ACP from being created at ingest.

4.10.14 Partnerships

Through the FDLP, GPO has established numerous content partnerships to provide electronic access to in-scope content housed at partner institutions. The partner institutions are responsible for maintaining public access to the content and have agreed to provide the content to GPO if they are no longer able to maintain it. This content has been cataloged by GPO and PURLs have been created in cataloging records to point to the content. For 1C.4, GPO will provide access to partner Web sites via links on an HTML page.

4.10.15 Save Search Query

Users will be able to save search results individually or as a batch. Full functionality will be implemented in R2.

4.10.16 Search

Release 1C.4 will build upon core search functionality enabled in earlier releases. Users will be able to perform a search for conceptually related terms. New concept relationships will be suggested at ingest and authorized GPO users will have the ability to manage concepts. This release also includes capabilities to apply one or multiple taxonomies to content. Users will also have the ability to perform a natural language search and to search inside publications.

4.10.17 Search Interface for External Systems

This feature provides for the creation of an external End User search API.

4.10.18 Search Results

Release 1C.4 will provide enhanced search functionality. The system will take users to the exact occurrence of a search term in a search result. Users will have the ability to hide document summaries so they do not display in search results. The system will also categorize and cluster results, and versions will be grouped into one entry in a results list. Authorized GPO users will have the ability to modify relevancy ranking factors so more relevant results are returned to users.

4.10.19 Support Granularity

FDsys will support granularity down to the level of any individual graphic.

4.10.20 User Help

Release 1C.4 will provide the capability for users to opt out of user support features such as context specific help that is displayed when a user pointer their mouse over an item on the user interface and clickable help icons.

4.10.21 User Interface

This release will provide GUIs for enhanced features such as selecting files to be packaged as a DIP, accessing publications on partner Web sites, and signing up to receive an email notification when new publications are available that match their search query or are added to a collection. Authorized users will also have the ability to manually create new Web pages.

4.10.22 User Notifications

Release 1C.4 will build upon Release 1C.2 by providing additional email and RSS notifications. Users will be able to sign up to receive notifications for system events, business events, and job processing events. This includes receiving a notification when now content is added to a collection or when there is a match to a user defined string.

5.0 Release 2

5.1 *Table of Capabilities*

The following table indicates which capabilities will be developed in Release 2.

System Capability		Release 2 (Overview)
Content Processing	Pre-Ingest Processing	Enhanced Functionality
	Ingest Processing	Enhanced Functionality
	Preservation Processing	Additional Capability
	Version Control	Enhanced Version Relationships
	Duplicate Content	Near Duplicate Detection
Infrastructure	Storage Management	Expanded as required
	Security	Enhanced as required
	Data Mining	Additional Capability
Content Submission	Harvested Content	Additional Capability
Content Access	Contact Management	GPO Users Manage Contacts
	Checking/Reformatting Content for 508 Compliance	Automated 508 assessment, transformation, and validation
	Deliver via RSS	Configure New RSS Feeds
	Deliver Content and/or Metadata	Batch delivery of multiple renditions and associated metadata from one or more publications
	Delivery by FTP	Delivery Based on Business Rules
	Follow Relationships to Other Document	Enhanced Legislative and Regulatory Relationships
	Format Transformation	Additional Transformations to Support Preservation Processes
	Knowledge Base	User Support Knowledge Base
	Partnerships	Additional Partnerships
	Reference Tools	Basic Reference Tools
	Save Search Query	Save Searches and Receive Notification with New Results
	Save Search Results	Save Search Results
	Search	Enhanced Functionality
	Search Results	Enhanced Functionality
	Support Granularity	Enhanced Functionality
	Training and Events	Interactive Training
	User Help	Enhanced Functionality
	User Helpdesk	Submit, Manage, and Access Helpdesk Inquiries.
	User Interface	Customizable User Interfaces and Mobile Support
	User Notification	Personalized Notification
Bulk Signing	PDF Signing Using Bulk Signing System	Apply Digital Signatures or Other Integrity Marks to Content
Persistent Name	Persistent Name Assignment	Enhanced Capabilities
	Persistent Name Resolution	Enhanced Capabilities

5.2 *Additional Capabilities*

5.2.1 Preservation Processing

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

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.

5.2.2 Harvested Content and Harvester

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.

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.

5.2.3 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 multi-relational 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.

Users will be able to extract, input, and store data and perform complex data analyses using a number of different statistical methods. These data analyses can include standard statistical displays (charts, tables, graphs, etc.). The resulting reports can be but into templates in real-time or delivered on a scheduled basis.

5.2.4 Contact Management

FDsys will provide the capability for GPO users to manage contacts.

5.2.5 Checking/Reformatting Content for 508 Compliance

FDsys will accept 508 requirement and implementation guidance from Content Originators. It will also automatically transform content to create Section 508 compliant content renditions and validate that content is compliant with Section 508 standards.

5.2.6 Delivery via RSS

FDsys will provide the capability for GPO to create new RSS feeds.

5.2.7 Deliver Content and/or Metadata

This release will build upon earlier releases to include delivery of audio and video content and metadata in ONIX format. The system will provide batch delivery of content and metadata from a single publication. It will also provide batch delivery of content and metadata from multiple publications (e.g., custom composition). The release also includes e-commerce features to support the GPO Sales Program.

5.2.8 Delivery by FTP

The system will build upon earlier releases by providing the capability to determine if delivery via FTP is possible based on business rules.

5.2.9 Follow Relationships to Other Documents

FDsys will provide enhanced legislative and regulatory content relationships in Release 2.

5.2.10 Format Transformation

Additional transformations as specified in the Requirements Documents to support preservation processes.

5.2.11 Knowledge Base

The system will provide the capability for users to manage and search the knowledge base.

5.2.12 Partnerships

The system will provide access to additional partnerships as necessary. It will also provide the capability for users to search cataloging records in order to provide access to select external repositories with which GPO has formal partnership agreements.

5.2.13 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.

In order to leverage library community knowledge of Government publications, the system will provide the capability for users to customize reference tools, and GPO will have the ability to manage any reference tools that are created.

5.2.14 Save Search Query

The system will provide the capability for public end users to establish an account with the system to store search queries. The system will execute save searches on a schedule defined by the user and notify users when the executed search returns results that were not included in the original search results (e.g., email me on a weekly basis when new documents about “railroad retirement” are added to the system).

5.2.15 Save Search Query

The system will provide the capability for public end users to establish an account with the system to store results sets and portions of results sets. This is similar to a “my folder” or “shopping cart” concept. Users will be able to save search results individually or as a batch.

5.2.16 Search

Release 2 will provided enhanced functionality. Users will be able to search by all fields in schema registered in the system. It will also support intelligent search (e.g., “notices” as a query terms directs the search towards the Federal Register).

5.2.17 Search Results

Release 2 will provided enhanced functionality. The system will provide recommendations to users for content and services based on preferences and queries of users and groups of similar users. The system will provide the capability to analyze search results, display those results graphically, display inline image thumbnails, highlight query terms in a document summary, and highlight query terms in the document. In addition, the system will provide the capability for a result set to equal the size of all content in all indexes.

5.2.18 Support Granularity

Release 2 will provided enhanced support for granularity by providing granularity at the paragraph level and down to the level of embedded graphical elements in publications. The system will provide the capability for users to access select agency publications at a level of granularity that is less than a publication. In addition the system will provide for the creation of new granules by aggregating and decomposing existing granules (e.g., custom composition).

5.2.19 Training and Events

The system will provide the capability for users to register for and participate in online training and events.

5.2.20 User Help

The system will provide user support phone numbers, email addresses, mailing addresses, real-time text chat, and fax numbers based on the function they are performing in the system. The system will also provide the capability for GPO to manage information that is displayed as a result of clicking on a help icon on the GUI.

5.2.21 User Helpdesk

The system will provide the capability for users to submit, manage, and access helpdesk inquiries.

5.2.22 User Interface

In release 2.0 FDsys will provide customizable user interfaces for public users. Uses will have the ability to create an account with FDsys. Through this account they will be able to specify delivery options, alert services, help features, preferred contact methods, frequently accessed tools, and search preferences. They will also be able to access custom lists and reference tools.

Public users will have the ability to add, remove, hide, change the placement of, and modify the size of elements on the user interface. They will also have the ability to change the text size and color scheme from available options. Customizations will be saved across user sessions and users will have the ability to revert the default display.

Release 2 includes the ability to display FDsys on mobile devices and support for WML. In addition, GUI will be designed so they could support non-English language text.

5.2.23 User Notification

Release 2 will provide enhanced user notifications including support for Sales notification. The system will notify users when their subscriptions are about to end (e.g., renewal notices), deliver personalized offers based on individual user history or users with similar histories. (e.g., "you may also be interested in…"), and provide notification of deliver fulfillment.

5.2.24 PDF Signing Using Bulk Signing System

FDsys authentication features, including requirements in the Bulking Signing Feature Group, 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. Integrity marks, such as digital signatures and watermarks, will be used to convey authentication information to users.

Requirements in this feature provide for the application of integrity marks on content, including digital signatures on PDF documents. The feature also includes requirements for validating integrity marks.

6.0 Release 3

6.1 *Table of Capabilities*

The following table indicates which capabilities will be developed in Release 3.

System Capability		Release 3 (Overview)
Infrastructure	Storage Management	Expanded as required
	Security	Enhanced as required
	ESB	Integration with Oracle
Content Submission	Style Tools	Additional Capability
	Content Originator Ordering	Full Functionality
	Pre-Ingest Collaboration	Additional Capability
Access	Section 508 Compliant User Interfaces	508 Compliant Software Applications and Operating Systems
	Deliver Content and/or Metadata	Access content that has been published in non-English and non-Roman character sets
	Delivery by FTP	Secure FTP and Performance Requirements
	Reference Tools	Dynamically Generated and Interoperable Reference Tools
	Support Granularity	Granularity for Audio and Video
	User Help	Dynamically Generated Context Specific User Help
Persistent Name	Persistent Name Assignment	Enhanced Capabilities
	Persistent Name Resolution	Enhanced Capabilities

6.2 *Additional Capabilities*

6.2.1 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.

6.2.2 Pre-Ingest Collaboration

The Pre-Ingest processing area will support a collaborative environment that enables Congressional Content Originators and the GPO Plant to collaborate on works in progress and jobs. The system will allow Congressional Content Originators to have multiple individuals view, edit, and submit job information and content. These users will be able to collaborate on individual jobs in Pre-Ingest Processing, and an authorized user (e.g., the Clerk of the House) must have the capability to approve all jobs and content submitted before they are ingested by FDsys. "Check in and check out" capabilities for work-in-progress content will be provided, and the system will also track versions of work in progress content.

6.2.3 Section 508 Compliant User Interfaces

Requirements build upon compliance in Release 1C to provide FDsys software applications and operating systems that are 508 compliant according to 36 CFR Part 1194.21, as necessary.

6.2.4 Section 508 Compliant User Interfaces

Access content that has been published in non-English and non-Roman character sets. Deliver descriptive metadata in COSATI.

6.2.5 Delivery by FTP

The system will push DIPs to user via Secure File Transfer Protocol (SFTP). This release also includes performance requirements.

6.2.6 Reference Tools

The system will provide dynamically generated reference tools and be interoperable with third party reference tools (e.g., search catalogs of other libraries).

6.2.7 Support Granularity

The system will provide for granularity in audio and video content.

6.2.8 User Help

The system will provide dynamically generated context specific user help.