



# Digital Content Contributors: Increasing access to digitized content

David Walls  
Suzanne Ebanues  
Jessica Tieman



# Increasing Historic Content in govinfo

## CONGRESSIONAL RECORD:

Official Register of the United States, 1930 7

### GOVERNMENT PRINTING OFFICE

NAME	OFFICIAL TITLE	Legal residence		Compensation
		State	Cong. dist.	
George H. Carter	Public Printer	Iowa	9th	\$10,000
John Greene	Deputy Public Printer	Mass.	5th	7,500
Mary A. Tate	Assistant to the Public Printer	Tenn.	2d	4,200
Henry H. Wright	Chief clerk	N. Y.	28th	4,400
Edward J. Wilver	Disbursing clerk	Pa.	16th	4,400
William A. Smith	Congressional Record clerk	D. C.		4,000
Daniel P. Bush	Medical and sanitary director	Nebr.	1st	4,400
James K. Wallace	Superintendent of accounts and budget officer	Ohio	1st	5,200
Ernest E. Emerson	Purchasing agent	Md.	5th	4,800
Byron L. Wehmhoff	Technical director	Wash.	3d	5,200
Burr G. Williams	Chief instructor of apprentices	Iowa	10th	3,600
Ellwood S. Moorhead	Production manager	Pa.	6th	5,600
Edward A. Huse	Night assistant production manager	Mass.	6th	5,000
Hermann B. Barnhart	Superintendent of printing	Ind.	9th	4,600
Bert E. Bair	Superintendent of presswork	Mich.	6th	4,600
Martin R. Spielman	Superintendent of binding	Mo.	4th	4,600
Edward G. Whall	Superintendent of platemaking	Mass.	12th	4,600
William A. Mitchell	Superintendent of planning	N. C.	5th	4,600
Alfred E. Hanson	Superintendent of construction and maintenance	Mass.	14th	5,200
Alton P. Tisdell	Superintendent of documents	Ohio	22d	5,000
William H. Kervin	Storekeeper and traffic manager	N. Y.	39th	4,400

## PROCEEDINGS AND DEBATES

FORTY-THIRD CONGRESS.

SPECIAL SESSION

OF THE SENATE OF THE UNITED STATES.

MARCH 4 TO 26, 1873.

VOLUME I.

federal register

TUESDAY, FEBRUARY 5, 1974  
WASHINGTON, D.C.

Volume 39 ■ Number 25

Pages 4551-4649

PART I



### HIGHLIGHTS OF THIS ISSUE

This listing does not affect the legal status of any document published in this issue. Detailed table of contents appears inside.

**MANDATORY PETROLEUM ALLOCATION**—Federal Energy Office proposes to amend certain distillate fuel regulations; comments by 2-10-74. 4592

**PHASE IV**—CLC issues general retail trade exemptions; effective 2-1-74. 4557

**ANTIBIOTIC DRUGS**—  
FDA changes moisture determination method for certain topical powder preparations; effective 3-7-74. 4570  
FDA certifies additional potency for a suspension of erythromycin; effective 3-7-74. 4570

**MOTOR VEHICLE DEFECT REPORTS**—DoT requires inclusion of identification numbers; effective 5-6-74. 4578

**NUCLEAR REACTOR CONSTRUCTION**—AEC proposes to permit certain pre-construction site preparation activities; comments by 3-7-74. 4582

**EMERGENCY LIVESTOCK FEED PROGRAM**—USDA revises eligibility requirements and certain other provisions; effective 2-5-74. 4567

**TOBACCO**—USDA marketing quotas for various types and grades (4 documents); effective 1-31-74. 4558, 4560, 4563, 4565

**COMMODITY LOANS**—USDA amends certain certification requirements; effective 2-5-74. 4566

**SAVINGS AND LOANS**—FHLBB proposal on acquisition of newly converted stock insured institutions; comments by 2-20-74. 4594

(Continued inside)

#### PART II:

**ENVIRONMENT**—EPA proposes textile industry effluent limitation guidelines; comments by 3-7-74. 4627

#### PART III:

**GRAIN STANDARDS**—USDA proposes rules of practice; comments by 3-18-74. 4639

# govinfo Priorities

- Collection Development Librarian and Collection Development Working Group
- [GPO's System of Online Access Collection Development Plan](#)

U.S. GOVERNMENT PUBLISHING OFFICE (Comprehensive level)

TITLE	SuDoc #	DATES	PRIORITY
Annual Report of the Superintendent of Documents	GP 3.1	1895 -	2
GPO Style Manual	GP 1.23/4	Prior to 2000	2
Monthly Catalog of U.S. Government Publications	GP 3.8/8		1
Monthly Catalog of U.S. Government Publications Periodical/Serials Supplement	GP 3.8/5		1
U.S. Congressional Serial Set Supplement (annual)	GP 3.8/6		
Checklist U.S. Public Documents, 1789-1909		1911	3*
Catalogue of the Public Documents of the ... Congress, ... Session		1895 - 1907	1
Index to the Reports and Documents of the ... Congress, ... Session ... With Numerical lists and schedule of volumes	GP 3.7:no.		1
Numerical lists and schedule of volumes of reports and documents of Congress (annual)	GP 3.7/2	1933/34 -	1



# Digital Content Contributors

- Institutions that digitize content within scope of the FDL P
- Digitization meets GPO specifications
- Content is provided to GPO for ingest into **govinfo**
- GPO provides ongoing access and preservation of the digital content

# Partner Content in govinfo



<b><u>Branch</u></b>	Executive
<b><u>Category</u></b>	Executive Agency Publications
<b><u>SuDoc Class Number</u></b>	CZ 1.9:
<b><u>Government Author</u></b>	Canal Zone Government.
<b><u>Publication Title</u></b>	Annual Report of Insurance Business Transacted in the Canal Zone During the Calendar Year 1965, Including Laws Concerning Licensing of Insurance Companies
<b><u>Date Issued</u></b>	December 31, 1965
<b><u>GPO Partnership</u></b>	This content was digitized by the University of Florida Libraries and is made publicly available through a partnership between the U.S. Government Publishing Office and the University of Florida.
<b><u>Rights Statement</u></b>	This document is a work of the United States Government and is not subject to copyright pursuant to 17 USC 105.
<b><u>Topic</u></b>	Panama Canal



# Steps to Become a Digital Content Contributor

- Discuss your digitized content with GPO
- Submit sample master and access files
- Sign a Memorandum of Agreement
- Transfer content to GPO via SFTP



# Digitization Guidelines

# Where to find Guidelines

Go to [FDLP.gov](http://FDLP.gov)  
 Go to “Partnerships”  
 Under “Digital Content  
 Contributors”

<https://www.fdlp.gov/file-repository/outreach/fdlp-partnerships/2924-digitization-guidance-final>

August 31, 2017

## Guidance for Contributing Digital Content to FDsys/govinfo

The following guidance for digital reformatting is based on current best practices. The results will provide the best image capture for the range of information products to be ingested into FDsys/govinfo.

The guiding principle for digitization rests in GPO’s definition of preservation copy of record for digital content:

The preservation copy of record for digital content is the preservation master file stored in a trustworthy repository. Derivatives of the preservation master copy are made available for access. *The digital copy of record should be produced to specifications that will allow the creation of a printed facsimile version, should one be needed.*

### PREScribed TECHNICAL SPECIFICATIONS

Specifications for preservation master, access derivative files, and metadata are described as “Preferred” and “Acceptable.” The Preferred specifications follow the latest best practices and guidance for digital preservation and support GPO’s operation of FDsys/govinfo as a digital repository. The Acceptable specifications are given to accommodate a variety of stakeholders and their technical capability.

Digitize publications using the preferred JPEG 2000, or the acceptable TIFF master file format following the prescribed sampling rate of 300 to 600 ppi to accurately capture the original content. Each page scanned will result in a separate digital file in the prescribed master format. An access file in pdf format will be derived from the master image files. Optical character recognition software will be run to embed machine readable text into the pdf.

	PREFERRED	ACCEPTABLE
MASTER FILE	Uncompressed JPEG 2000, 300-600 ppi, 24 bit RGB color conforming to ISO/IEC 15444-1 standard for JP2000	Uncompressed TIFF 6.0 300-600 ppi, 24 bit RGB color
ACCESS FILE	PDF/A 2-b with embedded Optical Character Recognition	PDF/A with embedded Optical Character Recognition
TECHNICAL METADATA	MIX XML*	
BIBLIOGRAPHIC METADATA	MARC XML	MARC XML

\* NISO z39.87 defines a set of metadata elements for raster images to enable users to develop, exchange, and interpret digital image files. These elements, such as information about the compression, color profile, resolution, scanner or digital camera make and model, can be recorded and preserved as technical information for still images. When possible, this information should be recorded as a set of XML elements conforming to the NISO metadata for images in xml Schema (MDQ).

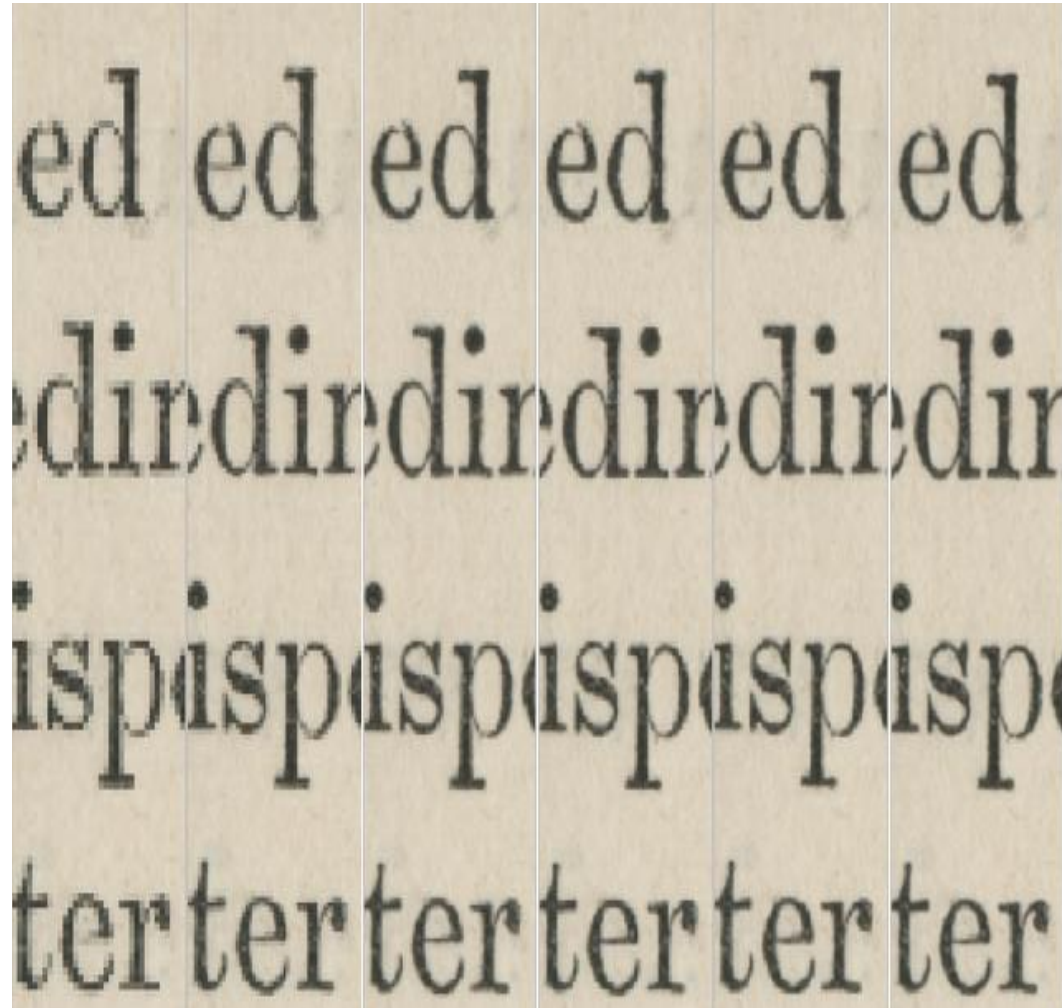


# Technical Specifications

	PREFERRED	ACCEPTABLE
MASTER FILE	Uncompressed JPEG 2000, 300-600 ppi, 24 bit RGB color conforming to the ISO/IEC 15444-1 standard for JPEG 2000	Uncompressed TIFF 6.0, 300-600 ppi, 24 bit RGB color
ACCESS FILE	PDF/A 2-b with embedded Optical Character Recognition	PDF/A with embedded Optical Character Recognition
TECHNICAL METADATA	MIX XML*	
BIBLIOGRAPHIC METADATA	MARC XML	MARC XML

# Other Specifications

Skew	No software de-skew should be used. The skew variation should be +/- 1 degree at the vertical axis at the left side of the page.
Gutter Margin	If publications are digitized from bound volumes, sufficient gutter margin has to exist to avoid page curvature that may interfere with image clarity, text readability, and OCR accuracy.
Pixels Per Inch	Test different ppi to be sure all text is readable. Higher pixel rates, especially beyond 600 ppi, do not create a clearer image, just larger files due to the number of pixels per inch captured.





# Selecting Publications

- Review representative sample of the titles.
- Vacuum publications covered in dust and dirt.
- Remove dirt, marginal notes, or pencil underlining with plastic art eraser.
- Use replacement pages if underlining and marginal notes in pen cannot be removed.
- Look for missing and damaged pages. Use replacement pages instead.
- Don't use publications with mold or tight gutter margins.



# Best Practices

- Preserve the look and feel of the original publication.
- Capture all information content from the original.
- Develop and document your digitization specifications.
- Conduct a pilot test and evaluate the results.
- Assess level of staff available and be knowledgeable about your equipment.
- Conduct a brief test each day to test that the equipment is working properly.



# Packaging & Metadata Guidelines

# Where to find Guidelines

Go to [FDLP.gov](http://FDLP.gov)  
Go to “Preservation”  
Go to “Guidelines & Best Practices” >

<https://www.fdlp.gov/file-repository/about-the-fdlp/miscellaneous-informational/2992-packaging-guidance-09252017-final>

## Guidance for Packaging Digital Content for Submission to FDsys/govinfo

The following guidance is for Digital Content Contributor Partners to utilize when packaging or preparing digital content for transfer of the digital content to the U.S. Government Publishing Office (GPO) for inclusion in govinfo. In order for GPO to submit content to the preservation repository, packages must be properly formed, and files must be named appropriately. Following the packaging guidance will assist GPO to submit content into govinfo efficiently.

### govinfo Packaging Structure

Content in govinfo must be packaged in a specific structure. A package must include:

- tiff-submitted or jpeg2000-submitted folder: for the master preservation files (TIFF or JPEG2000) if master files are being ingested
- pdf-submitted folder: access derivative files (PDF or PDF/A)
- metadata-submitted folder: contains the marc.xml metadata and template.xml (XML) files (if provided)
- thumbnails-submitted (80x129 pixels jpeg) (if provided)

Other renditions can also be included in a package but are less common; additional file formats may be approved or requested by GPO.

### govinfo File Naming Conventions

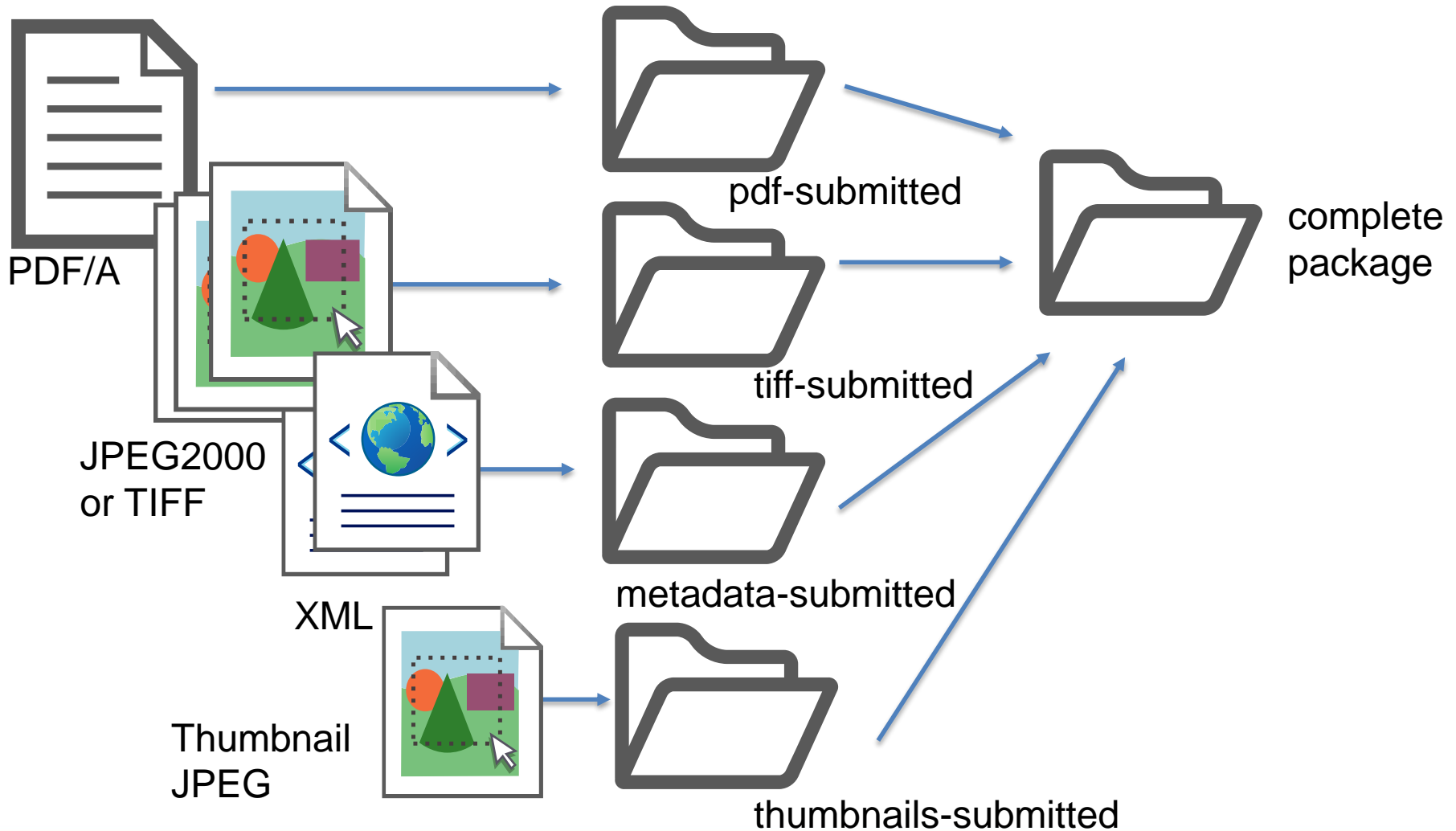
All files within the same package should be named with the same “stem” name. For example, if a given package represents a bill called “HR 1,” all files within that package should be named similarly:

hr1.pdf  
hr1.txt  
hr1.html

Using these naming conventions will ensure that each package makes its way into the proper rendition and that the system can recognize that all files belong to the same package.

All packages and files within packages should be named with the following convention:

# Packaging Basics

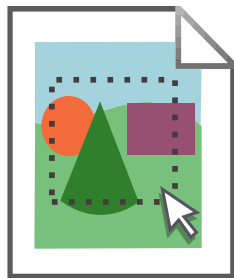


# Content descriptions

A single PDF document of the item



JPEG20000 or TIFF images for all pages of the item



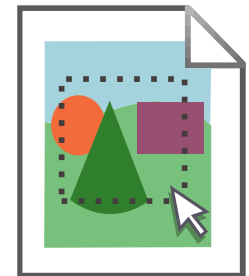
MARC.XML



template.xml



80x129 pixel image of book cover or first page of item





# File naming conventions

All packages, and pdf and image files within packages, should be named with the following convention:

“GovernmentAuthor-TitleofWork-DesignationofWork”

For example: “CDC-AnnualReport-Volume1”

When possible, use common abbreviations for authors and titles in the file names to make the file names as short in character length as possible. The maximum character length is 30 characters, not including dashes. Other examples of properly named package folders (or files within package folders) include:

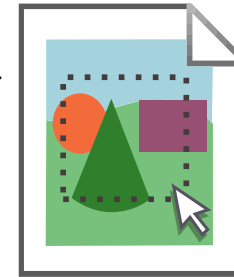
“NOAA-ActionPlanFishRelease-Vol161”

“NLM-NativeVoices-Vol2”

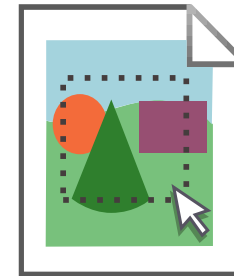
“DOT-BeyondTraffic-Vol3”



DOT-TrafficReport-Volume2-1



DOT-TrafficReport-Volume2-1



DOT-TrafficReport-Volume2-2

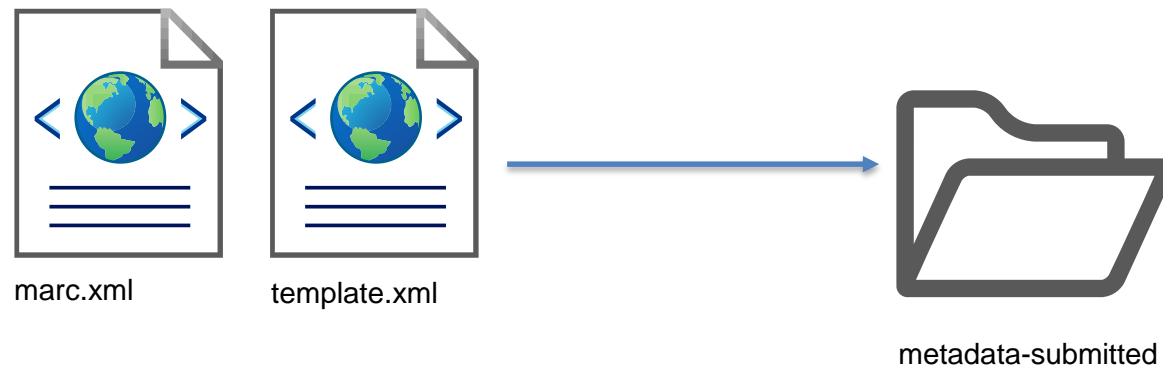


DOT-TrafficReport-Volume2

# Metadata requirements

All packages contain 1 MARC.XML file for each title

All packages contain 1 template.XML file for each title



**\*\*The naming convention for metadata files requires all MARC.xml files to be named “marc.xml” and all template files to be named “template.xml”**



# MARC.xml requirements

All MARC.xml files should include the following:

- Full SuDoc value in the 086 field
- Dates within the 260 or 264 files must be written in year-month-day format (YYYY-MM-DD); if the publication only lists a month and year, use the first day of that month. If the publication only lists a year, use the first day of that year.
- The 245 field must include subfield information to include the designation of the title if it is not a monograph.
- The 260 and 264 subfields “a,” “b,” and “c” must appropriately reflect the bibliographic information of the individual title within a series or as part of a multipart publication if it is not a monograph.
- A 500 field reading “This content was digitized by the [your institution’s name] and is made publically available through a partnership between the U.S. Government Publishing Office and [your institution’s name].
- A 710 field with your library or institution’s name. 710 \_\_ \$a Institution Name.

# Template.xml requirements

In addition to MARC.xml files, all **govinfo** packages include a template xml file. Partners may wish to provide the template xml in the digital content package as well. If a partner is not able to provide the template xml file, GPO staff will create one prior to submission of the package into **govinfo**.

Example:

```
<templateMD>
  <embargoDatetime></embargoDatetime>
  <digitalSignatureProfile>GPOSignatureProfile_LOWER_LEFT.xml
  </digitalSignatureProfile>
  <governmentAuthor1>Agriculture Department</governmentAuthor1>
  <governmentAuthor2>Forest Service</governmentAuthor2>
  <category>Executive Agency Publications</category>
  <sourceContentType>converted</sourceContentType>
  <packageDigitalOrigin>reformatted digital</packageDigitalOrigin>
  <otherIdentifier idStandard='ils-system-id'>001000611</otherIdentifier>
  <isContentSearch>true</isContentSearch>
  <fedPubName>GypsyMoth</fedPubName>
  <field name="Rights Statement"> This document is a work of the United States Government and is not subject to
  copyright pursuant to 17 USC 105.</field>
  <description>Gypsy moth -- Control -- Environmental aspects -- United States</description>
  <image>thumbnails/gypsymoth.jpg</image>
  <ggrank1>gypsy moth</ggrank1>
</templateMD>
```



# Examples from Panama Canal

govinfo [Browse](#) [About](#) [Developers](#) [Features](#) [Help](#) [Feedback](#) [SEARCH](#)

[Search](#) [Advanced](#) [Citation](#)

What are you searching for? [SEARCH](#)

## Download

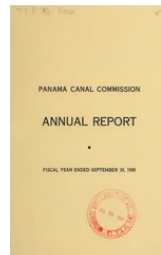
- PDF
- MODS
- PREMIS
- ZIP

## Actions

- Browse Government Publications
- CGP Record
- Share

## Content Details

### Annual Report, Fiscal Year Ended September 30, 1990



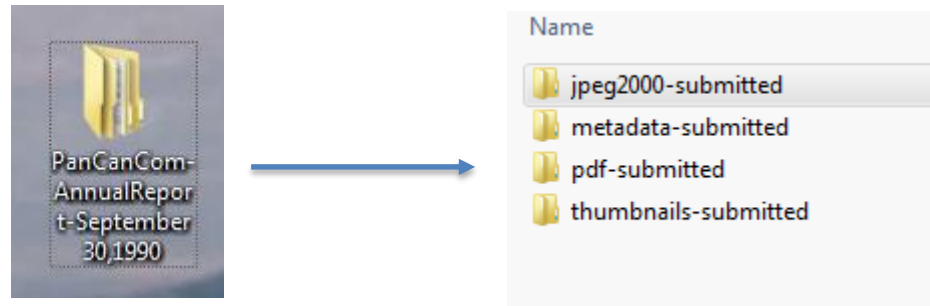
#### Summary

#### Document in Context

<b>Branch</b>	Executive
<b>Category</b>	Executive Agency Publications
<b>SuDoc Class Number</b>	Y 3.P 19/2:1/
<b>Government Author</b>	Panama Canal Commission.
<b>Publication Title</b>	Annual Report, Fiscal Year Ended September 30, 1990
<b>Date Issued</b>	September 30, 1990
<b>GPO Partnership</b>	This content was digitized by the University of Florida Libraries and is made publically available through a partnership between the U.S. Government Publishing Office and the University of Florida.
<b>Rights Statement</b>	This document is a work of the United States Government and is not subject to copyright pursuant to 17 USC 105.
<b>Topic</b>	Panama Canal

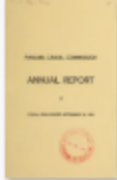
# Examples from Panama Canal

Example:



# Examples from Panama Canal

Example:

jpeg2000-submitted	metadata-submitted	pdf-submitted	thumbnails-submitted
<ul style="list-style-type: none"><li>annualreport1990unit_Page_001.jp2</li><li>annualreport1990unit_Page_002.jp2</li><li>annualreport1990unit_Page_003.jp2</li><li>annualreport1990unit_Page_004.jp2</li><li>annualreport1990unit_Page_005.jp2</li><li>annualreport1990unit_Page_006.jp2</li><li>annualreport1990unit_Page_007.jp2</li><li>annualreport1990unit_Page_008.jp2</li><li>annualreport1990unit_Page_009.jp2</li><li>annualreport1990unit_Page_010.jp2</li><li>annualreport1990unit_Page_011.jp2</li><li>annualreport1990unit_Page_012.jp2</li><li>annualreport1990unit_Page_013.jp2</li><li>annualreport1990unit_Page_014.jp2</li><li>annualreport1990unit Page 015.jp2</li></ul>	<ul style="list-style-type: none"><li>marc.xml</li><li>template.xml</li></ul>	<ul style="list-style-type: none"><li>annualreport1990unit.pdf</li></ul>	 <p>annualreport1990 unit_Page_001th m.jpg</p>

# Examples from Panama Canal

## Example:

```
<?xml version="1.0" encoding="UTF-8"?>
<templateMD>
  <embargoDatetime>2016-04-25T06:00:00Z</embargoDatetime>
  <digitalSignatureProfile>GPOSignatureProfile_LOWER_LEFT.xml</digitalSignatureProfile>
  <governmentAuthor1>Panama Canal Commission</governmentAuthor1>
  <category>Executive Agency Publications</category>
  <sourceContentType>converted</sourceContentType>
  <packageDigitalOrigin>reformatted digital</packageDigitalOrigin>
  <otherIdentifier idStandard="ils-system-id">000894041</otherIdentifier>
  <isContentSearch>true</isContentSearch>
  <fedPubName>Panama Canal</fedPubName>
  <field name="GPO Partnership">This content was digitized by the University of Florida Libraries and is made publically available through a partnership between the U.S. Government Publishing Office and the University of Florida.</field>
  <field name="Rights Statement">This document is a work of the United States Government and is not subject to copyright pursuant to 17 USC 105.</field>
  <field name="Topic">Panama Canal</field>
  <image>thumbnails/annualrepo1980unit_0001thm.jpg</image>
  <ggrank1>panama canal commission annual report </ggrank1>
</templateMD>
```





# Talk with the Partnership Team

[Partnership inquiry form](#)

Email: [PreserveFedInfo@gpo.gov](mailto:PreserveFedInfo@gpo.gov)