Digital Content

Contributors: Increasing access to digitized content

David Walls
Suzanne Ebanues
Jessica Tieman
govinfo Priorities

- Collection Development Librarian and Collection Development Working Group

- **GPO’s System of Online Access Collection Development Plan**
Digital Content Contributors

- Institutions that digitize content within scope of the FDLP
- 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

**Branch**
Executive

**Category**
Executive Agency Publications

**SuDoc Class Number**
CZ 1.9:

**Government Author**
Canal Zone Government

**Publication Title**
Annual Report of Insurance Business Transacted in the Canal Zone During the Calendar Year 1965, Including Laws Concerning Licensing of Insurance Companies

**Date Issued**
December 31, 1965

**GPO Partnership**
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.

**Rights Statement**
This document is a work of the United States Government and is not subject to copyright pursuant to 17 USC 105.

**Topic**
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
Go to “Partnerships”
Under “Digital Content Contributors”

## Technical Specifications

<table>
<thead>
<tr>
<th></th>
<th>PREFERRED</th>
<th>ACCEPTABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MASTER FILE</strong></td>
<td>Uncompressed JPEG 2000, 300-600 ppi, 24 bit RGB color conforming to the</td>
<td>Uncompressed TIFF 6.0, 300-600 ppi, 24 bit RGB color</td>
</tr>
<tr>
<td></td>
<td>ISO/IEC 15444-1 standard for JP2000</td>
<td></td>
</tr>
<tr>
<td><strong>ACCESS FILE</strong></td>
<td>PDF/A 2-b with embedded Optical Character Recognition</td>
<td>PDF/A with embedded Optical Character Recognition</td>
</tr>
<tr>
<td><strong>TECHNICAL METADATA</strong></td>
<td>MIX XML*</td>
<td></td>
</tr>
<tr>
<td><strong>BIBLIOGRAPHIC METADATA</strong></td>
<td>MARC XML</td>
<td>MARC XML</td>
</tr>
</tbody>
</table>
# Other Specifications

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Skew</strong></td>
<td>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.</td>
</tr>
<tr>
<td><strong>Gutter Margin</strong></td>
<td>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.</td>
</tr>
<tr>
<td><strong>Pixels Per Inch</strong></td>
<td>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.</td>
</tr>
</tbody>
</table>
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
Go to “Preservation”
Go to “Guidelines & Best Practices”

Packaging Basics

PDF/A

JPEG2000 or TIFF

XML

Thumbnail JPEG

Metadata
pdf-submitted

tiff-submitted

thumbnails-submitted

Complete package
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”
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:

```xml
<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

Content Details
Annual Report, Fiscal Year Ended September 30, 1990

Summary
Branch: Executive
Category: Executive Agency Publications
SuDoc Class Number: Y 3.P 192:1/
Government Author: Panama Canal Commission.
Publication Title: Annual Report, Fiscal Year Ended September 30, 1990
Date Issued: September 30, 1990
GPO Partnership: 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.
Rights Statement: This document is a work of the United States Government and is not subject to copyright pursuant to 17 USC 105.
Topic: Panama Canal
Examples from Panama Canal

Example:
Examples from Panama Canal

Example:

- jpeg2000-submitted
- metadata-submitted
- pdf-submitted
- thumbnails-submitted
Examples from Panama Canal

Example:

```xml
<?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