_____

Guidance for Packaging Digital Content for Submission to FDsys/govinfo

The following guidance is for Digital Content Contributer 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** effiencently.

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:
"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:
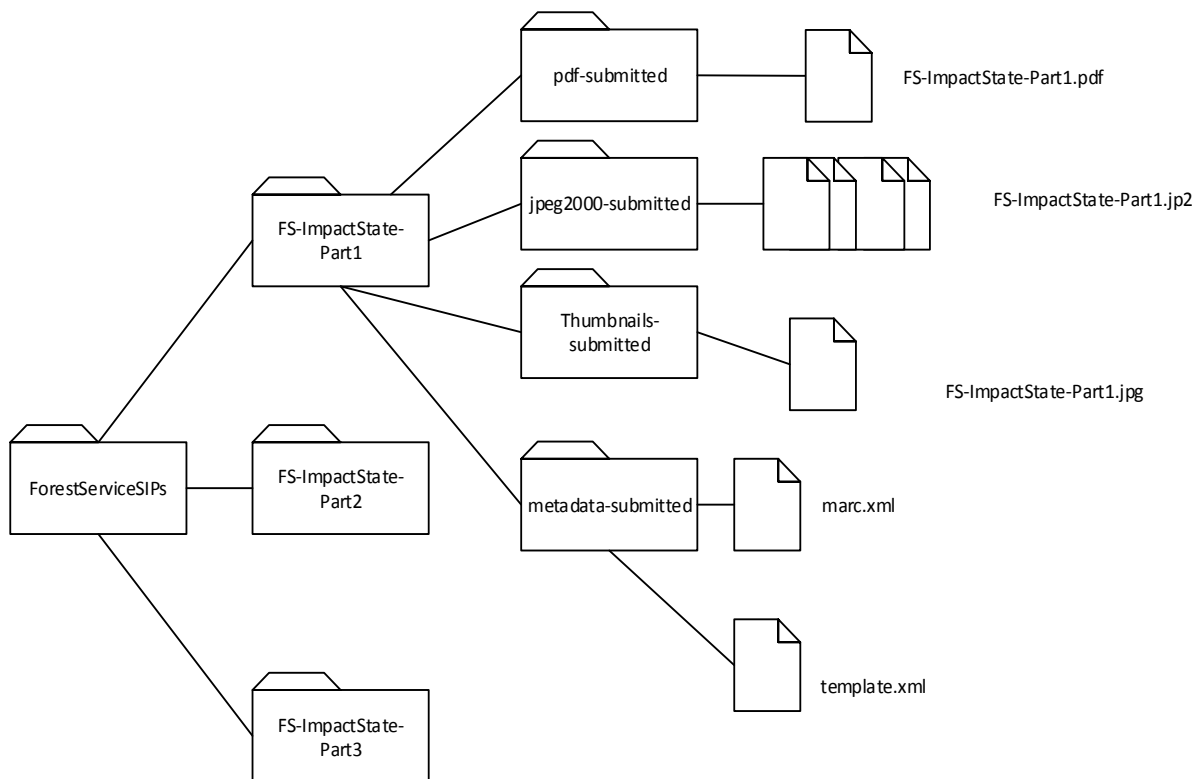
"NOAA-ActionPlanFishRelease-Vol161"

"NLM-NativeVoices-Vol2"

"DOT-BeyondTraffic-Vol3"

Creating Packages

- Each title (e.g. indidivdual serial within a series, part within a multi-part, volume, or monograph) will have its own top-level folder.
- Each top-level folder will be uniquely named with the proper file-naming convention listed above.
- Withn each top-level folder for each title, the following sub-folders will be created and named as such:
  - 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)
- The following files within each sub-folder will be named as such:
  - tiff-submitted, jpeg2000-submitted, pdf-submitted, and thumbnails-submitted folder files: title files with the same name as the top-level folder (e.g. RailroadRetirementBoard-ThePrivacyactandtheRRBoverview.jp2)
  - metadata-submitted folder files: all marc xml files will be titled "marc.xml" and all template xml files will be named "template.xml"
- All packages must contain a marc.xml and a template.xml file within the metadata-submitted folder. If not provided by the partner, GPO will create these files.
- The content producer may choose to place all "top-level" folders for each title within a folder for the purposes of submission, see below as illustrated:

A folder-structure diagram showing: ForestServiceSIPs containing FS-ImpactState-Part1, FS-ImpactState-Part2, and FS-ImpactState-Part3. FS-ImpactState-Part1 contains pdf-submitted (FS-ImpactState-Part1.pdf), jpeg2000-submitted (FS-ImpactState-Part1.jp2), Thumbnails-submitted (FS-ImpactState-Part1.jpg), and metadata-submitted (marc.xml, template.xml).

Metadata Requriments

MARC Files

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 approrpriately 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.

If partners are not able to provide a MARC.xml file or to edit a file to reflect the individual issue/volume, GPO will create/update the file prior to submission of the package into **govinfo**.

Template Files

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**.

The following values must reflect the bibliographic information of the title. Note that the "image" element value will change for each title to reflect the title of the document in the thumbnails-submitted folder of the work.

"template.xml" element	Example	Description
embargoDatetime	2011-08-22T17:00:00Z	Entering an "embargo" date and time allows users to set a specific time for the content to become publically visible after it is ingested into the system. This is an optional field. Note: Time must be specified as GMT. Provided by GPO.
governmentAuthor1	Health and Human Services Department	First level government author. Note: Must match authority file exactly, including upper and lower case. Note: This is required.
governmentAuthor2	Office of Public Health and Science	Second level government author. Note: Must match authority file exactly, including upper and lower case.
governmentAuthor3	National Institute of Environmental Health Sciences	Third level government author. Note: Must match authority file exactly, including upper and lower case.
secondGovtAuthor1	Environmental Protection Agency	If the document is created by a second Government agency, this field can be used as well. First level government author Note: Must match authority file exactly, including upper and lower case.
secondGovtAuthor2	Radiation & Indoor Air	If the document is created by a second Government agency, this field can be used as well. Second level government author. Note: Must match authority file exactly, including upper and lower case.
secondGovtAuthor3		If the document is created by a second Government agency, this field can be used as well. Third level government author. Note: Must match authority file exactly, including upper and lower case.

"template.xml" element	Example	Description
category	Legislative Agency Publications	One of the following: "Regulatory Information", "Bill and Statutes", "Legislative Agency Publications", "Budget and Presidential Materials", "Congressional Committee Materials", "Proceedings of Congress and General Congressional Publications", "Directories of Organizations and Officials", "Congressional Rules and Procedures", "Legislative Agency Publications", "Judicial Agency Publications", "Judicial Publications", "Executive Agency Publications". Note: Defaults to "Other" if not specified.
sourceContentType	converted	Use "harvested" if content was acquired through web harvesting. Use "converted" if content was generated from a digital imaging process (e.g. scanned by an FDL). Note: Defaults to "deposited" if not specified.
packageDigitalOrigin	digitized other analog	"born digital", "reformatted digital", "digitized microfilm", "digitized other analog" Use "digitized other analog" if generated from paper digitization project Note: Defaults to "born digital" if not specified.
isContentSearch	true	Possible values are "package", "true", "false", or "leafs". Always set to "true." "true" specifies that the digitalObject with @isPackageContent set to "true" should be content-indexed so that the package as a whole will be returned by full-text searches. In addition, all nested granules will be indexed as well.
isBrowse	false	Always set to false.
chamber	HOUSE	Use for legislative publications ONLY. This is the chamber of congress which was responsible for the publication. Possible values are "HOUSE", "SENATE" or "JOINT".

"template.xml" element	Example	Description
fedPubName	Featured Legislative Branch Publications	This is NOT standardized bibliographic descriptive metadata. An optional "user-friendly pseudonym" for this publication (also known as the "Mini Collection Name"). You will be instructed when to use this field. This name will be used for browse and other presentation needs to group packages that belong together. Other examples might include "NIST Publications" "Warren Commission"
field	This document is a work of the United States Government and is not subject to copyright pursuant to 17 USC 105.	Example text for a field/Rights Statement field
field/@name	Rights Statement	The name of the document-custom field.
field/@name	GPO Partnerships	The name of the document-custom field.
Field	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.	

"template.xml" element	Example	Description
description	Describes the architecture and history of the United States Capitol. Printed pursuant to H. Con. Res. 221. Also known as Serial Set 14620. 106th Congress, 2d Session.	An optional field; only use when instructed and when "abstract" field is empty in MARC values
image	thumbnails/page1 .jpb	The "cover art" image for the package. This is a reference to the image wherever it may be located in the package. It must be specified as: {rendition}/{file} for example: thumbnails/coverArt.jpg Further, the rendition specified must be a publicly available rendition.
ggrank1	9/11	Data to be added to grank1. Items searched using the terms specified here will be ranked higher in search results. Grank1 is the highest value.
ggrank2	Terrorism	Data to be added to grank2. Items searched using the terms specified here will be ranked higher in search results.
ggrank3		Data to be added to grank3. Items searched using the terms specified here will be ranked higher in search results.

Sample template.xml file:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<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>Government Printing Office</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>
```