Updated December 1, 2023

Congressionally Mandated Reports Submission Guidelines

These submission guidelines are designed to assist Federal agencies with submitting congressionally mandated reports that comply with the Access to Congressionally Mandated Reports Act.

Submission Checklist:

- ❑ Transmittal letter and all parts are combined into one file per format.
- ❑ Files have been reviewed for non-public information and redacted as appropriate.
- ❑ PDF is saved as PDF/A-1 or PDF/A-2.
- ❑ PDF is searchable or OCRed and 508 accessible.
- ❑ PDF is not password protected, locked, or digitally signed.
- ❑ PDF metadata under Properties is populated for Title and Author.
- ❑ Optional XML file is submitted with XSD, DTD, CSS, XSL, and JPG files.
- ❑ Optional Excel file includes column and row header definitions in a separate tab.
- ❑ Optional Text and XML files are UTF-8 encoded.

File Formats:

GPO supports the submission of congressionally mandated reports in open, machine-readable file formats. All submissions require, at minimum, a file containing the searchable text of the report and transmittal letter(s). The file should conform to the PDF/A-1 or PDF/A-2 standards. Pursuant to the May 9, 2013, Open Data Policy Memorandum (OMB M-13-13), Project Open Data, and the Foundations for Evidence-Based Policymaking Memorandum (OMB M-19-23), executive agencies should produce congressionally mandated reports and the information therein in formats and information modeling practices consistent with their agency's public access plan. These plans typically include agency and domain-specific guidelines and requirements for Data Management Plans and Peer-reviewed Publications, and likewise contain recommendations for the publication of Federal information via mediums which are findable, accessible, interruptible, and reusable (FAIR Data Principles endorsed by NIST). See below for optional formats that may be submitted in addition to PDF/A files.

GPO will not accept files of any format or their content if any aspect of the data has the following characteristics:

- Platform-dependent
- Proprietary
- Compressed or 'lossy'
- Contains embedded files, including audio, video, executables, or scripts;
- Contains text that is not searchable or that does not have optical character recognition (OCR)
- Contains encryption
- Contains password protection
- Contains a digital signature
- Contains personally identifiable information (PII) such as Social Security numbers, Government-issued identification numbers, or personal financial account numbers

Federal agencies may be contacted to resubmit reports if the reports do not meet the criteria listed above. In addition, reports submitted to the GPO submission portal should not contain PII; during the submission process, Federal agencies will be required to verify that the report being submitted does not contain PII.

How to Organize Information:

Combine the transmittal letter, multiple parts, appendices, executive summaries, and other ancillary materials into one file per file format. For example, one required PDF/A file, one optional XLSX file, one optional Text file, and one optional XML file plus associated schema, stylesheet, and images. File names should be concise and should not contain spaces, punctuation, or special characters except for dashes and underscores (e.g., 2023-Report_to_Congress-USGPO.pdf).

Required File Format:

PDF

- PDF files should be saved as PDF/A-1 or PDF/A-2. For information about how to create a PDF/A file, see the "About PDF/A-1 and PDF/A-2" section below.
- PDF files should be accessible and conform to the Revised 508 Standards (https://www.access-board.gov/ict/). For more information on accessibility guidelines, see https://www.section508.gov/create/pdfs.
- PDF metadata should be populated with relevant information for Title and Author. Note: if saving as PDF/A from Microsoft Word, add Title and Author to the file in Microsoft Word before saving as PDF/A. If converting from a non-PDF/A PDF in Adobe Acrobat, PDF metadata can be found under File > Properties.
- File extension should be .pdf.

- Information about PDF/A-1 is available at ISO 19005-1:2005 Document management -- Electronic document file format for long-term preservation -- Part 1: Use of PDF 1.4 (PDF/A-1): (http://www.iso.org/iso/catalogue_detail?csnumber=38920)
- Information about PDF/A-2 is available at ISO 19005-2:2011 Document management -- Electronic document file format for long-term preservation -- Part 2: Use of ISO 32000-1 (PDF/A-2): (http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=50655)

Optional File Formats:

XML

- XML should be well formed and valid. Special characters should be escaped or enclosed in CDATA.
- Relevant XSD, DTD, CSS, XSL, or JPG files that are required for validation, display, and understandability should be submitted along with the XML file.
- References and links should be relative (e.g., <img src="image-name.jpg">).
- Encoding should be UTF-8 without BOM (byte order mark), and UTF-8 should be declared in the prolog along with the XML version.
- File extension should be .xml.
- Information about Extensible Markup Language 1.1 (XML) Second Edition is available from https://www.w3.org/TR/2006/REC-xml11-20060816.

Text

- Encoding should be UTF-8. For information about how to save as Text with UTF-8 encoding, see the "[Text Encoded as UTF-8](#)" section below.
- File extension should be .txt.
- Information about RFC 3629: UTF-8, A Transformation Format of ISO 10646 is available from http://tools.ietf.org/html/rfc3629.

Excel

- Definitions of column and row headers should be provided in a separate tab, and the meaning of tab names should also be defined.
- As applicable, domain-specific information should be included in the file to aid in the long-term understandability of the data.
- Macros should be disabled and external data sources should be transferred into the spreadsheet so the file is "self-contained." If the Excel file contains formulas,

charts, graphs, or custom styling please contact GPO for a file assessment prior to submission.

- Excel files should not be password protected or locked.
- File extension should be .xlsx, and file should be an Open Office XML Workbook (i.e., Excel Workbook).
- Information about Office XML Open File Formats is available from https://www.ecma-international.org/publications-and-standards/standards/ecma-376/.

About PDF/A-1 and PDF/A-2:

What is PDF/A?

PDF/A (https://pdfa.org/resource/iso-19005-pdfa/) is a standardized file format which ensures that visual appearance of electronic documents is preserved over time without the need of additional software or tools to render and view them. These features of the PDF/A standard provide a greater assurance that the information within text-based documents can reliably be viewed well into the future and mitigate digital obsolescence as technologies change over time. More about PDF/A standards can be found from the PDF Association (https://pdfa.org/resource/pdfa-in-a-nutshell-2-0/).

Why PDF/A-1a, PDF/A-1b and PDF/A-2a, PDF/A-2b?

PDF/A-1 and PDF/A-2 are conformance levels described within the PDF/A standards. The "B" and "A" conformance differentiate between B-basic, and A-accessible requirements. Basic conformance defines a minimal set of requirements for the reliable reproduction of a document's visual appearance; accessible conformance extends Level B to include additional features to improve a document's accessibility and reliable text extraction.

**How do I create PDF/A-1b, PDF/A-1b and PDF/A-2a, PDF/A-2b?**

Creation of PDF/A-1 and PDF/A-2 documents begins at the point at which the document is created, digitally imaged, or converted, and requires commercially available (e.g., Microsoft Word https://www.microsoft.com/en-us/microsoft-365, Adobe Acrobat Products https://www.adobe.com/products/catalog.html?id=deptAcrobatFamily) or free PDF creator software (e.g., Simpo PDF Creator Lite http://www.simpopdf.com/pdf-creator-lite.html or PDF Creator http://www.pdfforge.org).

Saving Documents as PDF/A in Adobe Acrobat

Within Adobe Acrobat, PDFs can be converted or created by selecting "Save > Save as type 'PDF/A'" which will then provide a Preflight pop-up box with various settings. PDF/A-2b is

the most flexible of the PDF/A conformance standards and is preferred to other conformance standards.

Example of Preflight Pop-up Box:
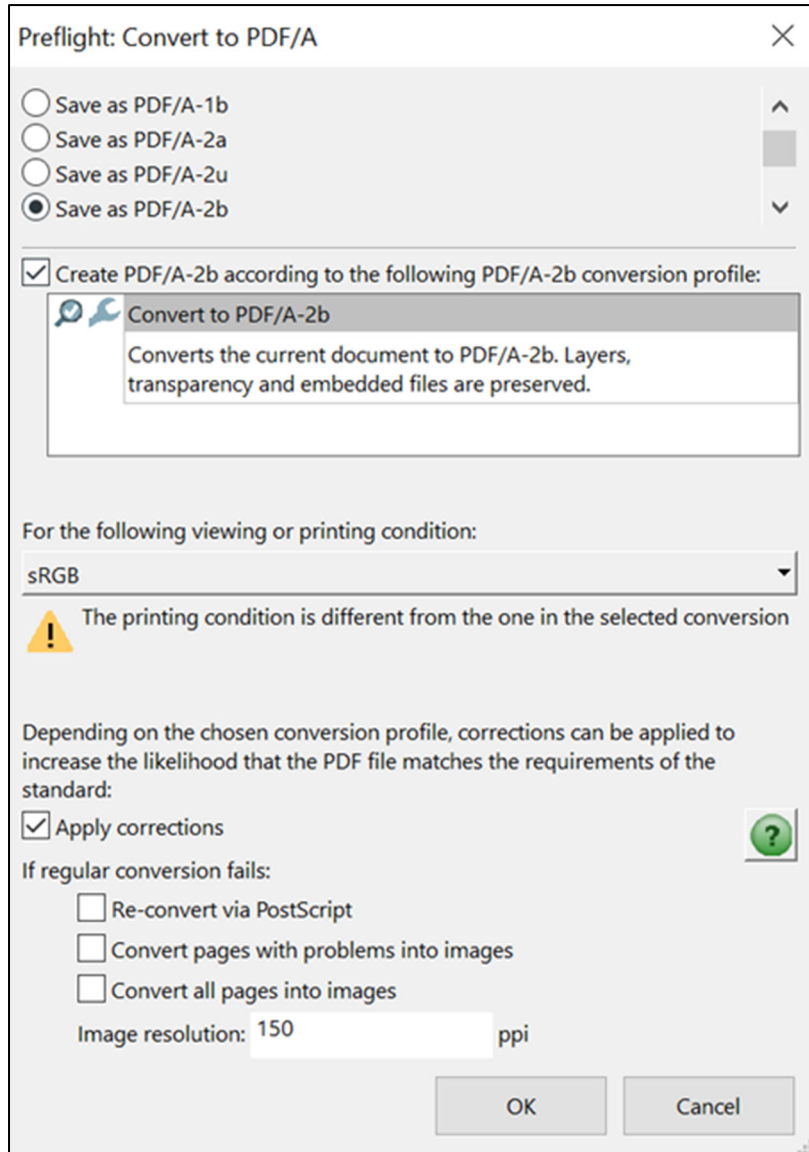
Saving documents as PDF/A in Microsoft Word

Microsoft Word software often allows the option to "Save As PDF > Options > PDF/A compliant." These are examples of the selections made when saving a document as a PDF/A in Word. Note the "Document structure tags for accessibility" and "PDF/A compliant" have been selected.

Example 1 of Saving as PDF/A in Microsoft Word:
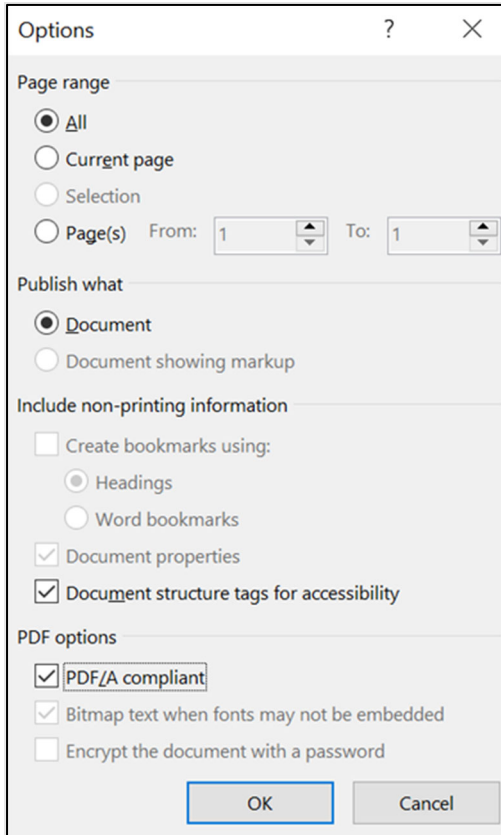
Example 2 of Saving as PDF/A in Microsoft Word with Adobe Acrobat Add-on:
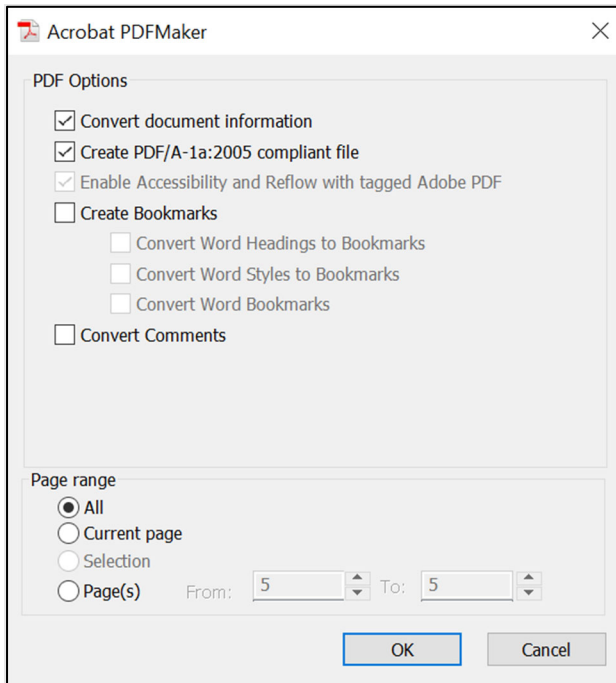
PDF/A-1 and PDF/A-2 Validation

It is not always possible to determine if a document fully meets the ISO PDF/A standard by visual appearance alone, even if the document has a blue banner indicating it was saved as a PDF/A. For an additional layer of assurance that the PDF file has been created and saved as the proper file format type and conformance level, tools are available to perform file format and standard validation. Adobe's Preflight can perform this validation function (see image as example below), but other technologies such as VeraPDF (https://verapdf.org/) can also provide this capability.
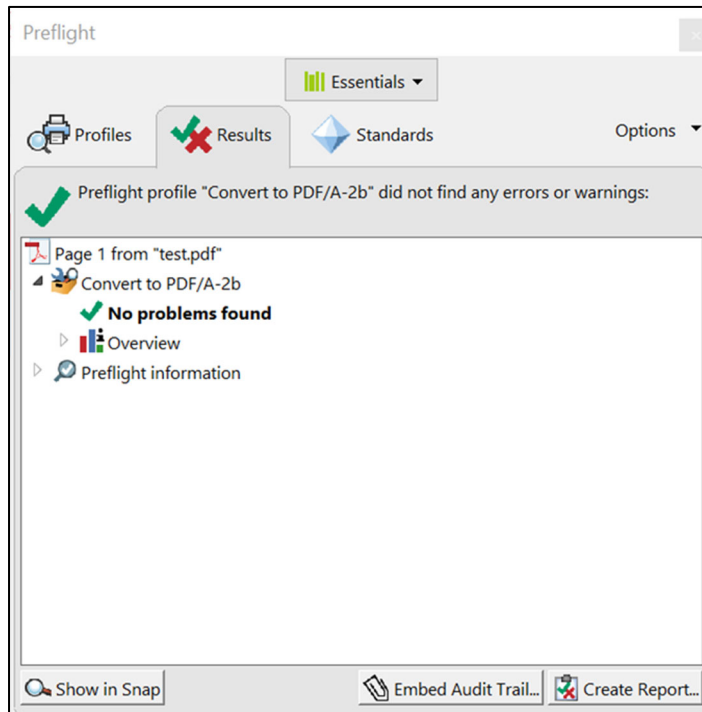
In Adobe Acrobat, select "Tools > Apply PDF Standards > Preflight > Profiles."

Example of Adobe Acrobat Preflight Profiles:
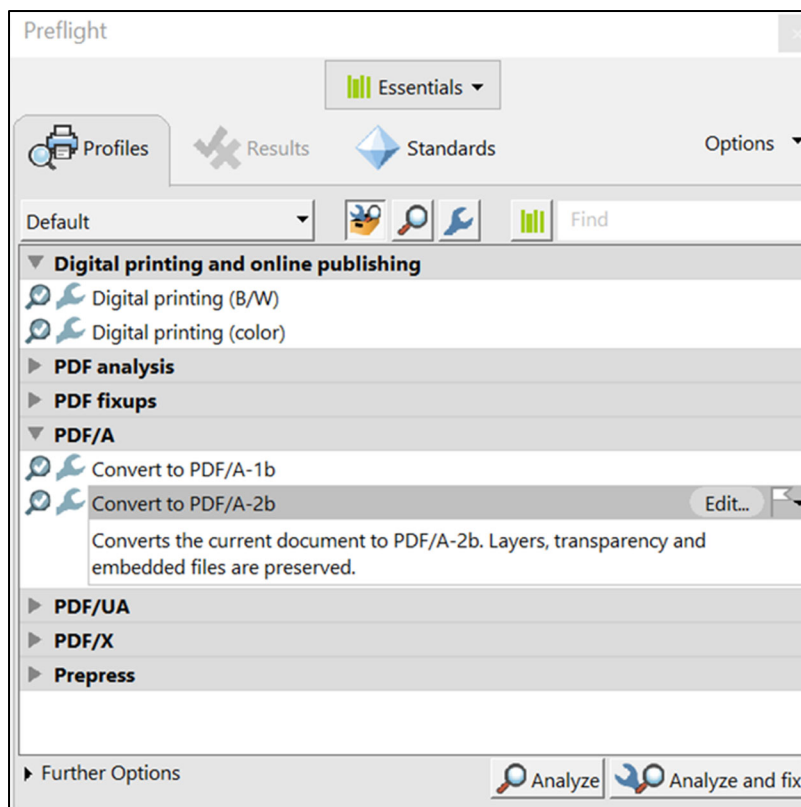
To check for PDF/A-2b Validation select "Analyze." A successfully validated PDF/A-2b document is confirmed by the message "No problems found."

Example of PDF/A-2b Validation using Adobe Acrobat Preflight:

How to Scan and Get Text from an Image with OCR (Optical Character Recognition):

Directions for how to enable OCR for a PDF document are available from Adobe: Get text from an image: How to OCR a PDF | Adobe Acrobat

How to Combine or Separate Pages within a Single PDF:

Directions for how to combine PDF documents are available from Adobe: How to merge PDFs and split a PDF easily | Adobe Acrobat

Text Encoded as UTF-8:

Microsoft Word software allows the option to save as UTF-8 Text by selecting "File > Save as type > Plain Text (*.txt) > Save > File Conversion screen: Other encoding: Unicode (UTF-8)." This is an example of the selections made when saving a document as Plain Text (*.txt) in Microsoft Word. Note "Other encoding" and "Unicode (UTF-8)" are selected.
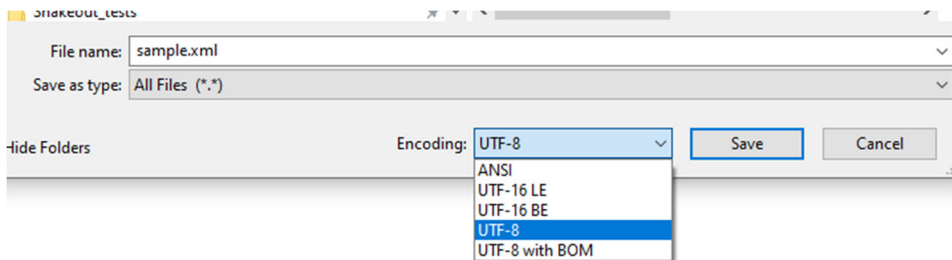
Example of Saving as UTF-8 Text in Microsoft Word:
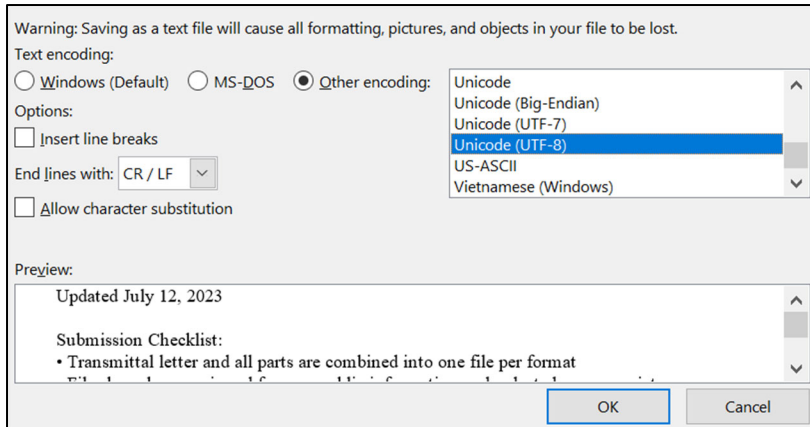
XML Encoded as UTF-8 without BOM:

In the standard Windows Notepad, save as type: All Files (*.*), include .xml in the File name, and choose "UTF-8" in the Encoding dropdown. Many other text editors offer similar options to encode XML files without a byte order mark.

Example of Saving XML in Notepad:

Rights Management:

The submitting organization grants GPO non-exclusive rights to use submitted files and metadata to maintain permanently the archival versions of content, including the right to reproduce or modify archival content and associated metadata and to create or modify new archival content with associated metadata for migration or other archival purposes; and to reproduce, display, transmit, promote, and/or otherwise distribute the access version of the electronic information. The submitting organizations acknowledge that congressionally mandated reports are works of the United States Government and are not subject to copyright pursuant to 17 USC § 105.

Preservation, Authentication, and Content Integrity:

The submitting organizations acknowledge that the congressionally mandated report will be preserved in the GovInfo ISO:2012 16363 certified Trustworthy Digital Repository, and GPO retains rights to support the integrity and authenticity of all content, manage multiple copies of content in distributed locations to mitigate risk of data loss, and migrate, transfer, or otherwise make changes to digital files to ensure long-term access to content, while striving to retain the visual and functional fidelity to the original digital content. In addition, the following applies:

- Original content files are maintained in the GovInfo preservation repository.
- Files are backed up at GPO's Continuity of Operations (COOP) and cloud sites.
- The COOP remote disaster recovery site retains near-real-time data synchronization and five-minute failover capability with full content availability and system functionality.
- GPO commits to preserving technical, descriptive, and preservation standards-based metadata for all content to ensure long-term viability and understandability of content information.
- GPO commits to regular validation checks of content to ensure bit-level integrity of content information over time.
- GPO assumes the responsibility of file format migration, transfer, conversion, or other actions to ensure the long-term access to content while striving to retain the visual and functional fidelity of the original content.
- If content is withdrawn, GPO will follow Superintendent of Documents policies.
- Documents and publications in GovInfo exist in the form of a package, which is at the top of the hierarchy of identifiable units of GovInfo content. A package consists of content files, metadata about the content, and a packaging scheme that is optimized for preservation.
- GPO maintains integrity for all file formats through generation of a SHA-256 algorithm that is stored in a PREMIS metadata file.
- GPO ensures that users can determine integrity of PDF files available on GovInfo by using digital signature technology and by adding a visible Seal of Authenticity. GPO will apply a Seal of Authenticity to PDF files.
- GPO's mechanisms to convey authenticity and integrity provides evidence to end users that published content is official, complete, and unaltered since being disseminated by GPO.
- GPO acknowledges the role of the submitting organization in the metadata for each congressionally mandated report.

Public Inquiries Regarding Congressionally Mandated Reports and Publicity:

Public inquiries received by GPO regarding the content of an individual congressionally mandated report may be forwarded to the submitting organization. GPO will address all inquiries that concern the functionality of GovInfo. GPO will publicize the availability of and promote the public use of congressionally mandated reports that are available through GovInfo. In addition, GPO will create or update bibliographic records for congressionally mandated reports in the Catalog of U.S. Government Publications (CGP) (https://catalog.gpo.gov/) with persistent links to GovInfo.