
**U.S. Government Printing Office
Federal Digital System
User Guide Document
Federal Register XML Rendition**

Prepared by: Program Management Office

**Office of the Chief Information Officer
U.S. Government Printing Office**

September 21, 2009

Revision History

Revision	Date	Description
1.0	September 21, 2009	Version 1.0

Contact Information

For any questions regarding this document contact pmo@gpo.gov.

Table of Contents

1. Introduction.....	1
1.1. Purpose	1
1.2. Legal Status & Authenticity of Federal Register files via Data.gov.....	1
2. Schema Description	5
2.1. Sections Available in XML.....	7
2.2. Sections, Parts, and Articles	7
2.2.1. XPath Examples.....	9
2.3. NEW PART Section.....	9
2.4. The Contents of Article	10
2.5. PREAMBLE Tag	11
2.6. SUPLINF Tag	12
2.7. Presidential Documents.....	12
3. Resources Directory.....	15

1. Introduction

The U.S. Government Printing Office (GPO) and the National Archives' Office of the Federal Register (OFR) partnership is offering bulk data downloads of Federal Register files to the general public via Data.gov and FDsys. This effort began when the President challenged Federal agencies to create a more open and transparent government, promote accountability, and provide information to citizens about what their Government is doing (see 74 FR 4685, January 26, 2009 at <http://www.gpo.gov/fdsys/pkg/FR-2009-01-26/pdf/E9-1777.pdf>). The Public Printer's letter of March 9, 2009 pledged to provide trusted information in whatever form is required to meet the President's objectives http://www.gpo.gov/pdfs/news-media/letter_030909.pdf.

In addition, the Office of the Federal Register coordinates with the Office of Science Technology Policy to ensure that the OFR/GPO partnership meets customer expectations. To follow through on our commitment, we are expanding and accelerating the development of FDsys to provide XML-structured content as rendered output. This will give users access to masses of data to reconfigure and redistribute as they wish to meet the specialized needs of their constituencies.

1.1. Purpose

The purpose of this document is to provide an overview of Federal Register XML files and associated schema. The FDsys Bulk Data repository at <http://www.gpo.gov/fdsys/bulkdata/> contains the Federal Register in XML from 2000 to the present. Please see FDsys at www.fdsys.gov for access to the Federal Register in PDF and HTML formats.

1.2. Legal Status & Authenticity of Federal Register files via Data.gov

Q. What is the data set available for the Federal Register in XML?

A. Federal Register files in XML have been converted and simplified from the SGML rendition used for the printed publication. The Federal Register is available in XML starting with the year 2000, when GPO began composing Federal Register material in SGML. OFR/GPO plan to convert the remaining electronic data set for the Federal Register (1994-1999) at a future date.

Q. Are bulk data downloads of XML files offered on FDsys via Data.gov part of the official version of the Federal Register?

A. No, the XML-structured files offered for bulk download are not part of the official on-line format of the *Federal Register*. While GPO's XML files are based on the original source data submitted by Federal agencies, OFR markup, and GPO typesetting and composition markup in SGML, only the PDF and Text versions of *Federal Register* content on GPO Access and FDsys have legal status as parts of the official online format of the *Federal Register*. Additional development will be required before OFR/GPO can specify that XML files are a part of the official online edition of the *Federal Register*.

Q. Do OFR/GPO plan to include XML files as part of the official format of the Federal Register?

A. Yes. The current set of XML-structured material (minus graphics) is not yet characterized as part of the official online *Federal Register* format because the underlying data used to create a viewable rendition of the material must be thoroughly scrutinized and tested. The XML-structured files are derived from SGML-tagged data and printing codes, which may produce anomalies in display. For example, complex tabular material in XML files may not display as correctly composed objects equivalent to the tables that appear in the Text and PDF files. Tabular material that displays in an ambiguous or distorted manner could affect the substantive meaning of regulations and other documents.

Our goal is to clean up the XML data and develop associated style sheets to the point that the XML rendition can be characterized as a display format of the official online edition. We are also developing the means to embed graphics in the files, so that an XML version may be deemed both official and complete. The OFR will issue a *Federal Register Bulletin* to alert users when the XML-structured files are ready to be included as part of the official online edition.

Q. What is the legal basis for making determinations about official status?

A. The Federal Register Act established the Administrative Committee of the Federal Register (Administrative Committee or ACFR) as the regulatory body with authority to determine the format(s) of the official serial publication known as the *Federal Register* (44 U.S.C. Ch. 15). Under 1 CFR 5.10, the official formats approved by the ACFR include a paper edition, a microfiche edition, and an online edition. OFR and GPO carry out ACFR regulations by developing appropriate means to display Federal Register material in the official formats.

Q. When did the Federal Register first appear online?

A. The official online edition dates to 1994 when Congress authorized an online edition in Public Law 103-40 (the Government Printing Office Electronic Information Enhancement Act of 1993; codified at 44 U.S.C. 4101).

Q. Are Federal Register XML bulk download files digitally signed?

No, XML files available for download are not digitally signed. They can be manipulated and enriched to operate in the various applications that users may devise. GPO is evaluating technology that could be used to digitally sign XML files for future official editions posted on FDsys. Adding signed non-PDF files to FDsys would be an enhancement for FDsys users, but would not be used to restrict or adversely affect the XML bulk data downloads available to our customers.

Q. What does the term “digitally signed” mean?

A. Currently, GPO uses digital signature technology on PDF documents to add a visible Seal of Authenticity (a graphic of an eagle) to authenticated and certified documents. The technology allows GPO to secure data integrity, and provide users with assurance that the content is unchanged since it was disseminated by GPO. A signed and certified document also displays a blue ribbon icon to the left of the Seal of Authenticity and in the Signatures tab within Adobe Acrobat or Reader. When users print a document that has been signed and certified by GPO, the Seal of Authenticity will automatically print on the document, but the blue ribbon will not print.

Q. How reliable is the metadata and the underlying tagging in bulk data files?

A. The document markup and metadata found within bulk data files are generally reliable and complete. However, there are variations in this underlying data due to inconsistencies in the composition and typesetting process. As a result, some data-mining applications and user aids may not produce 100 per cent accurate results.

Q. What is the legal status of Federal Register user aids?

A.: Federal Register user aids, including, finding aids, indexes, search tools, metadata associations, and tagging schemes are not part of the legal text of the *Federal Register*. These ancillary features help users explore and extract data, but the official legal text stands on its own. No person should form absolute legal conclusions based on search results, finding aids, metadata associations, extractions of data, and the like. Ultimately, only the official text of the *Federal Register* may be relied upon as evidence in a court of law.

Q. What is the authenticity of Federal Register bulk data files after they have been downloaded to another site?

A. We cannot vouch for the authenticity of data that is not under OFR/GPO control. OFR and GPO are providing free access to Federal Register data via XML for display in various applications and mash-ups outside the FDsys domain. The OFR/GPO partnership does not endorse third party applications, and does not evaluate how our original legal content is displayed on other sites. Consumers should form their own conclusions as to whether the downloaded data can be relied upon within an application or mash-up. An application may link to the official *Federal Register* on FDsys to provide users with additional assurance.

Q. Do OFR and GPO assert any control over downstream uses of bulk data?

A. In general, there are no restrictions on re-use of information in Federal Register documents because U.S. Government works are not subject to copyright. OFR and GPO do not restrict downstream uses of Federal Register data, except that independent providers should be aware that only the OFR and GPO are entitled to represent that they are the providers of the official versions of the *Federal Register* and related Federal Register publications.

Q. How can re-publishers indicate the source of Federal Register data?

A. Re-publishers of Federal Register data may cite FDsys and OFR/GPO as the source of their data, and they are free to characterize the quality of data as it appears on their site. But private sector re-publishers are prohibited from using the seal of the National Archives and Records Administration (NARA) or stylized Federal Register logos identified in NARA regulations (36 CFR part 1200) on their products because that would unlawfully misrepresent the legal status of the material, and could falsely identify private organizations as entities of the Federal Government.

2. Schema Description

The schema chosen to represent the Federal Register is a simplified version of the SGML schema that is used as part of the print production process, with some presentation and print specific tags removed or collapsed, and then converted to well-formed XML. This schema was chosen for the following reasons:

1. It is a complete and faithful representation of the Federal Register, which matches most closely to the author's original intent.
2. It describes the data using semantic tags in a way that is appropriate to the Federal Register Domain. For example, <RULE>, <NOTICE>, and <AGENCY> are all tags in this schema.
3. It fully describes the structure of the Federal Register, including the large structure (parts, articles, corrections, table of contents, etc.), the document structure (titles, paragraphs, sections, etc.), and semantic structure (CFR references, agency names, contact information, amendment text, etc.)

Since the schema is not an authoring schema and the SGML to XML conversion process maintains the order of the tags, this allows the XML schema to be more permissive than if it were used for checking authored content.

The schema being produced for this effort describes the data as it actually occurs from the OFR. Documents are not being cleaned up because they do not match the schema; instead, the schema was selectively relaxed. Such an approach maintains 100% fidelity to the original data, and eliminates any errors that might occur in schema interpretation or further data manipulation.

The following table lists the SGML tags that were removed or collapsed into a source attribute in the Federal Register XML.

Purposed Fields to be Removed	Fields to be Removed or Collapsed	Description
EDITOR	Removed	Used to indicate content editor
FNC	Removed	Used to generate a new column
FNEP	Removed	Used to generate a new even page
FNOP	Removed	Used to generate a new odd page.
FNP	Removed	Force new page.
Q	Removed	Inserts vertical spaces
NPAR	Collapse to P	Used to generate a new paragraph where the <P> tag would create a run in entry.
P	Collapse to P	Normally used for paragraph. A paragraph here has the first line indented.
P1	Collapse to P	Paragraph, indented one em on left.
P-1	Collapse to P	Paragraph, turnovers indented one extra em on left.

P1-3	Collapse to P	Paragraph, indented one em on left, turnovers indented three ems on left.
P2	Collapse to P	Paragraph, indented two ems on left.
P-2	Collapse to P	Paragraph, turnovers indented two ems on left. Same as FP1-2, which should not be used.
P2-4	Collapse to P	Paragraph, indented two ems on left, turnovers indented four ems on left.
P-3	Collapse to P	Paragraph, turnovers indented three ems on left.
P-DASH	Collapse to P	Paragraph, the last line of which fills with low-line dashes
OLNOTE1	Collapse to OLNOTE1	Sets footnote for first overlay note
OLNOTE2	Collapse to OLNOTE1	Sets footnote for second overlay note.
OLNOTE3	Collapse to OLNOTE1	Sets footnote for third overlay note.
OLNOTE4	Collapse to OLNOTE1	Sets footnote for fourth overlay note.
OLNOTE5	Collapse to OLNOTE1	Sets footnote for fifth overlay note.
OLNOTE6	Collapse to OLNOTE1	Sets footnote for sixth overlay note.
FP	Collapse to FP	Flush paragraph
FP1	Collapse to FP	Flush paragraph, all lines indented one em.
FP-1	Collapse to FP	Flush paragraph, turnovers indented one em
FP1-2	Collapse to FP	Paragraph, first line indented one em and turnovers indented two ems
FP2	Collapse to FP	Flush paragraph, all lines indented two ems
FP-2	Collapse to FP	Flush paragraph, turnovers indented two ems
FP2-2	Collapse to FP	Flush paragraph, all lines indented two ems.
FP2-3	Collapse to FP	Paragraph, first line indented two ems and turnovers indented three ems
FP3	Collapse to FP	Flush paragraph, all lines indented three ems.
FP-DASH	Collapse to FP	Flush line that fills with low-line dashes.
FRP	Collapse to FP	Flush right material, actually held in 1 em from right margin
FRP0	Collapse to FP	True flush right material
HD	Collapse to HD	First level head in the following sections.
HD1	Collapse to HD	First level head in the following sections.
HD2	Collapse to HD	Second level head in the following sections.
HD3	Collapse to HD	Third level head in the following sections.
HD4	Collapse to HD	Fourth level head in the following sections.
HD5	Collapse to HD	Fifth level head in the following sections.
HD6	Collapse to HD	Sixth level head in the following sections.

HD8	Collapse to HD	Lowest level head in the following sections.
HED	Collapse to HD	The first head in the following sections.
HED1	Collapse to HD	Special first level head in text of Section.
THED	Collapse to HD	Head that turns sideways on the page.
TSECT	Collapse to HD	Section head that turns sideways on the page.
FNC	Removed	Used to generate a new column

2.1. Sections Available in XML

The following are currently available in Federal Register XML:

- Contents
- Rules and Regulations
- Proposed Rules
- Notices
- Corrections

The following are not currently available in Federal Register XML:

- Front Matter (e.g. cover page)
- CFR Parts Affected
- Reader Aids
- CFR Checklist
- CFR Issuances
- Table of Effective Dates
- Graphics

2.2. Sections, Parts, and Articles

This section describes the top-level structure of a Federal Register XML file.

The XML schema, being a translated reproduction of the SGML schema, contains tags and content in the same order as they appear in the printed document. Major sections are grouped appropriately (<CNTNTS>, <RULES>, <PRORULES>, <NOTICES>, <NEWPART>, <CORRECT>), and all data and tags are represented – with the exception of the reduced tags from the previous section.

The XML tags and their descriptions of the schema above are shown below:

XML Tag	Description
FEDREG	The root tag for all document types in Federal Register publications. This tag includes children such as CNTNTS, RULES, PRORULES, NOTICES, NEWPART.

VOL	Contains the volume number for the publication.
NO	Contains the issue number of the volume.
UNITNAME	Contains the display name of the unit which follows, for example, "Notices", "Proposed Rules", "Rules", and "Presidential Documents".
CNTNTS	Contains table of contents pages for the Federal Register.
RULES	Used to start Rules section of the federal register. RULES may contain at least one <RULE>.
RULE	Used to start individual Rules in the Rules section.
PRORULES	Used to start Proposed Rules section of the Federal Register. Should contain at least one <PRORULE>.
PRORULE	Used to start individual Proposed Rule of the Federal Register.
NOTICES	Starts Notices section of the Federal Register. Should contain at least one <NOTICE>.
NOTICE	Start individual Notice within Notices section.

An abbreviated example of the overall section and structure is below:

```

<FEDREG>
  <VOL>65</VOL>
  <NO>21</NO>
  <UNITNAME>Contents</UNITNAME>
  <CNTNTS>
    ... THE CONTENTS OF THE TABLE OF CONTENTS ...
  </CNTNTS>
  <DATE>Tuesday, February 1, 2000 1-3-00</DATE>
  <UNITNAME>Rules and Regulations</UNITNAME>
  <RULES>
    <RULE> ... THE CONTENTS OF THE RULE ... </RULE>
    .
    .
    .
  </RULES>
  <UNITNAME>Proposed Rules</UNITNAME>
  <PRORULES>
    <PRORULE> ... THE CONTENTS OF THE PROPOSED RULE ... </PRORULE>
    .
    .
    .
  </PRORULES>
  <UNITNAME>Notices</UNITNAME>
  <NOTICES>
    <NOTICE> ... THE CONTENTS OF THE NOTICE ... </NOTICE>
    .
    .
    .
  </NOTICES>
  <NEWPART>
    ... THE CONTENTS OF THE PART ...
  </NEWPART>
  .
  .
  .

```

```
</FEDREG>
```

2.2.1.XPath Examples

The schema allows for a wide variety XPath commands for extracting items:

//RULE → Output all rules

(//PRTPAGE/@P)[0] → Get the first page number

(//PRTPAGE/@P)[last()] → Get the last page number

//RULE[descendant::PRTPAGE/@P = '12627'] → Get rule which contains page 12627

//RULE[PREAMB/AGENCY = 'NUCLEAR REGULATORY COMMISSION']

→ Get all rules for the nuclear regulatory commission

2.3. NEW PART Section

The NEWPART section holds the contents of a part which often includes a part title and a list of notices, presidential documents, proposed rules, rules, and section names. A part number will always accompany the part title. In some cases, the specified rule will contain a preamble which can contain everything up to, but not including, the supplementary information.

The XML tags and their descriptions of the schema above are shown below:

XML Tag	Description
NEWPART	Used to start a new part of the Federal Register.
PART	Part group or container tag for search and retrieval purposes.
PARTNO	Sets the part number on title page of new part within volume.
PTITLE	Part title for new part within volume.

An abbreviated example of the NEWPART section and structure is below:

```
<FEDREG>
.
.
.
<NEWPART>
  <PTITLE>
    <PRTPAGE P="47239" />
    <PARTNO>Part IV</PARTNO>
    <PRES>The President</PRES>
    <PNOTICE>Notice of July 28, 2000–Continuation of Emergency</PNOTICE>
  </PTITLE>
  <PRESDOCS>
    .
    .
    .
  </PRESDOCS>
</NEWPART>
</FEDREG>
```

2.4. The Contents of Article

The contents of article tags are the ones that usually describe the majority of the data in the Federal Register. For example, XML tags such as HD and P are often used to start the head sentence in the following section while P tag is used to describe the rest of the text. The flush paragraph or FP is used to denote either large or small text depending on where it is being used. In addition, the majority of the sections are also following by Federal Register doc number as well as an agency billing code.

XML Tag	Description
HD	Used to start a new part of the Federal Register.
P	Part group or container tag for search and retrieval purposes.
DATES	Used to describe dates such as effective, applicable, and comment dates.
FP	Flush paragraph.
FRDOC	Federal Register doc number. It is used to track when entries were submitted for publication.
BILCOD	Agency billing code number.
EDNOTE	Used for editorial notes.

The contents will roughly have the same structure

```
<FEDREG>
.
.
.
<CNTNTS>
...
<AGCY>
  <HD>Agricultural Marketing Service</HD>
  ...
  <SEE>
    ...
    <P>Farm Service Agency</P>
    ...
  </SEE>
  ...
</AGCY>
...
</CNTNTS>
.
.
.
<RULES>
  <RULE>
    .
    .
    .
  <EDNOTE>
    <FP>The Farm Service has...</FP>
  </EDNOTE>
  <DATE>December 4, 2000.</DATE>
```

```

<FRDOC>[FR Doc. 00-31252]</FRDOC>
<BILCOD>File 12-5-00; 8:45 am </BILCOD>
</RULE>
</RULES>
.
.
.
</FEDREG>

```

2.5. PREAMBLE Tag

The PREAMBLE can hold contents that describe the current page of the document as well as agencies, sub agencies, actions, summaries, RIN numbers, and CFR citations. In general, a preamble tag can hold everything up to but not including the supplementary information. To view a complete list of valid fields, please see the XSD schema.

The XML tags and their descriptions of the schema above are shown below:

XML Tag	Description
PREAMB	Used to start the preamble. The preamble contains everything up to but not including the supplementary information.
PRTPAGE	Used to identify the page number in all documents in Federal Register
AGENCY	Used to identify the agency name.
CFR	Used to identify Code of Federal Regulation citation.
RIN	Used to identify regulatory information number.
SUBJECT	Used for title of subject.
ACT	Used to identify an action.
SUM	Used to identify a summary.
ADD	Used to describe contact addresses.
FURINF	Used to describe "FOR FURTHER INFORMATION CONTACT:".
SUPLINF	Used to describe supplementary information.
LSTSUB	Used to start a list of subject section.
REGTEXT	Used to designate regulatory information that will be inserted into the CFR.

An abbreviated example of the PREAMB section and structure is below:

```

<RULE>
  <PREAMB>
    <PRTPAGE P="12627"/>
    <AGENCY TYPE="F">NUCLEAR REGULATORY COMMISSION</AGENCY>
    <CFR>10 CFR Part 2</CFR>
    <RIN>RIN 3150-AI08</RIN>
    <SUBJECT>Interlocutory Review of Rulings on Requests...</SUBJECT>
    <ACT><HD SOURCE="HED">ACTION:</HD><P>Final rule.</P></ACT>
    <SUM> ... the summary of the rule ... </SUM>
  </PREAMB>
</RULE>

```

```

<DATES> .... Description of effective or applicable dates ... </DATES>
<ADD> ... contact addresses ... </ADD>
<FURINF> ... where to go for further information about rule ... </FURINF>
</PREAMB>
.
.
.

<FRDOC>[FR Doc. E8-4768 Filed 3-7-08; 8:45 am]</FRDOC>
<BILCOD>BILLING CODE 7590-01-P</BILCOD>
</RULE>

```

2.6. SUPLINF Tag

The SUPLINF tag holds supplementary information that describes Federal Register documents. This information may include the page number, list of subjects or regulatory text material from the Rules section. The list of subjects will also include a list of CFR numbers cited in the appropriate rules.

The XML tags and their descriptions of the SUPLINF schema are shown below.

XML Tag	Description
SUPLINF	Used to describe supplementary information.
LSTSUB	Used to start a list of subject section.
REGTEXT	Used to designate regulatory information that will be inserted into the CFR.

An abbreviated example of the SUPLINF section and structure is below:

```

<RULE>
...
<SUPLINF>
... Supplementary information ...
<LSTSUB> ... List of CFR subjects appropriate to the rule ... </LSTSUB>
<REGTEXT PART="2" TITLE="10">
... Specific changes to the CFR (Title 10, Part 2 in this example) ...
</REGTEXT>
.
.
.
</SUPLINF>
...
</RULE>

```

2.7. Presidential Documents

The PRESDOCS tag must have one or more presidential documents which may include determinations, executive orders, memos, notices, or proclamations. A presidential notice, for example, will often have a title and presidential signature associated with it. It may also include the place of issuance which is usually “The White House.” These items are always followed by

Federal Register doc number, the date filed, and the billing code.

The XML tags and their descriptions of the schema above are shown below:

Node	Description
PRESDOCS	Used to start presidential documents section of the Federal Register.
PRESDOCU	Used to start individual Presidential document in the Federal Register.
PRESDOC	Used to start individual Presidential documents section of the Federal Register.
PRNOTICE	Notice of the Presidential documents.
DETERM	Presidential determination in Presidential documents.
EXECORD	Presidential Executive Order in Presidential documents.
PRMEMO	Presidential Memo in presidential documents.
PRORDER	Presidential Order in Presidential documents.
PNOTICE	Notice in Presidential documents.
TITLE3	CFR Title for Presidential documents.
PSIG	President associated with a Presidential document.
PLACE	Place of issuance for a Presidential document.
DATE	Date associated with the Presidential document.

An abbreviated example of the PRESDOCS section and structure is below:

```

<FEDREG>
...
  <NEWPART>
    <PTITLE>
      <PRTPAGE P="47239" />
      <PARTNO>Part IV</PARTNO>
      <PRES>The President</PRES>
      <PNOTICE>Notice of July 28, 2000–Continuation... </PNOTICE>
    </PTITLE>
    <PRESDOCS>
      <PRESDOCU>
        <PRNOTICE>
          <TITLE3>Title 3–</TITLE3>
          <PRES>The President<PRTPAGE P="47241" /></PRES>
          <PNOTICE>Notice of July 28, 2000</PNOTICE>
          <HD SOURCE="HED">Continuation of Iraqi Emergency</HD>
          <FP>On August 2, 1990, by Executive Order 12722...</FP>
          <FP>This notice shall ...<E T="04">Federal Register</E>...and...</FP>
          <PSIG>wj</PSIG>
          <PLACE>THE WHITE HOUSE,</PLACE>
          <DATE>July 28, 2000.</DATE>
          <FRDOC>[FR Doc. 00-19587</FRDOC>
          <FILED>Filed 7-31-00; 8:45 am</FILED>
          <BILCOD>Billing code 3195-01-P</BILCOD>
        </PRNOTICE>
      </PRESDOCU>
    </PRESDOCS>
  </NEWPART>

```

```
</PRESDOCS>  
</NEWPART>  
...  
</FEDREG>
```

3. Resources Directory

The resources directory in the Federal Register bulk data repository at <http://www.gpo.gov/fdsys/bulkdata/FR/resources> contains the current version of the XML schema, the XML stylesheet used to display the XML files in a browser on the FDsys website, and this user guide.