

**U.S. Government Publishing Office
govinfo
System Design Document**

**Volume XLI:
Data Management Definition (DMD)**

Education Reports from ERIC

R1C2 Edition

Prepared by: govinfo Program

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

April 16, 2010

Revision History

| Rev. | Date | Description |
|------|------------|--|
| 0.1 | 05/10/2009 | Initial Version |
| 0.2 | 06/01/2009 | Initial Draft |
| 0.3 | 06/15/2009 | Search Technologies Technical Review |
| 0.4 | 11/05/2009 | Updates based on PDF-to-Text conversion |
| 0.5 | 12/21/2009 | Search Technologies Architect Review |
| 0.6 | 03/31/2010 | PST Review |
| 0.7 | 04/05/2010 | Implement PST comments and added {PCS/isFallbackTitle} |
| 0.8 | 04/06/2010 | Implement Peer Review Comments |
| 0.9 | 04/15/2010 | Changed value for {PM/Publisher} constant |
| 0.10 | 04/16/2010 | Incorporated final Peer Review Comments |
| 0.11 | 09/25/2010 | <ol style="list-style-type: none"> 1. Arity of various PM elements updated 2. Section 6.2, index profile fields belong to package only. Updated various fields from upper-case to lower-case 3. 7.1 Updated {publishdate} formatting in line 2 4. 7.3 FQL updated for search field identifiers 5. Section 7.3.1 updated |
| 0.12 | 01/16/2013 | Changed {PM/topic} and {PM/geographicSubject} to {PM/subject/topic} and {PM/subject/geographic} respectively in accordance with PTR 728 |
| 2.0 | 05/27/2014 | Minor correction, updated to as-built, edited for grammar and consistency, and accepted all changes. |
| 2.1 | 05/27/2014 | Search Engine Replacement Project (SERP) Update |
| 2.3 | 02/24/2015 | GPO name change |
| 2.4 | 03/15/2015 | Added Preservation Planning chapter |
| 3.0 | 01/29/2018 | Updated to reflect the govinfo interface as built |

Responsibilities

| Description | Responsible Party |
|-------------|--|
| Co-Owners: | Program Management: Lisa LaPlant Technical: Ronald Matamoros Robert van Es |

Documentation Conventions

- Strings with embedded values are indicated with curly-braces, for example:
 Compilation of Presidential Documents Volume {PCS/volume},
 Issue {PCS/issue}, {PM/dateIssued}
- References to XML entities and attributes are referenced using the XPath standard. However, to save room, common prefixes may be abbreviated (see next section).

System

The name of the system discussed in the DMD is now **govinfo**. When the system was originally designed and implemented, the name of the system was FDsys or Federal Digital System. While many of the references to FDsys have been changed in this document, this name persists in areas related to computer code, file names, schemas etc. For example, the name of the computer file containing metadata for a package continues to be "fdsys.xml".

Abbreviations

This document uses the following abbreviations for specifying elements in "fdsys.xml":

- PH Package header
 fdsysPackage/packageHdr/
- PM Package metadata
 fdsysPackage/packageHdr/descPkgMd/
- PCS Package collection specific metadata
 fdsysPackage/packageHdr/descPkgMd/collectionSpecific/
- GM Group metadata (generic granule metadata)
 fdsysPackage/mdSect/descMdGroups/descMdGroup
- GCS Group collection specific metadata (collection specific granule metadata)
 fdsysPackage/mdSect/descMdGroups/descMdGroup/collectionSpecific/

Table of Contents

| | |
|--|-----------|
| 1. Introduction..... | 6 |
| 1.1. General Description..... | 6 |
| 1.2. Document Types..... | 6 |
| 2. "fdsys.xml" Metadata Schema Elements | 7 |
| 2.1. Package Metadata | 7 |
| 2.2. Granule Metadata..... | 9 |
| 2.3. Standardized References..... | 9 |
| 3. Renditions and Input Files..... | 10 |
| 3.1. Renditions | 10 |
| 3.2. Plant Processing..... | 10 |
| 3.3. Migration Input & Packaging..... | 10 |
| 3.4. Day Forward Input | 10 |
| 4. Parsing | 11 |
| 4.1. Renditions | 11 |
| 4.2. Parsing Text..... | 11 |
| 4.2.1. Preprocessing..... | 11 |
| 4.2.2. Granules..... | 11 |
| 4.2.3. Parsing Metadata | 11 |
| 4.2.4. PM/title and PCS/isFallbackTitle | 13 |
| 4.2.5. PM/dateIssued | 13 |
| 4.2.6. PM/abstract..... | 14 |
| 4.2.7. PCS/accessId | 14 |
| 4.2.8. PCS/type | 14 |
| 4.2.9. PCS/ericNumber..... | 15 |
| 4.3. Validation Heuristics | 15 |
| 4.3.1. Validating Granules and Packages | 15 |
| 5. govinfo Processing..... | 16 |
| 5.1. Special Manual Interventions Required..... | 16 |
| 5.2. Text Creation | 16 |
| 5.3. PDF Processing..... | 16 |
| 5.3.1. Renaming PDF files..... | 16 |
| 5.4. HTML Processing..... | 16 |
| 6. Content Publishing and Indexing..... | 17 |
| 6.1. Indexing Granularity | 17 |
| 6.2. Index-profile field mapping..... | 17 |
| 6.3. Computing "treesort" | 20 |
| 6.4. Solr Fields..... | 20 |
| 7. Search and Browse..... | 24 |
| 7.1. Search Results Presentation | 24 |

| | | |
|------------|---|-----------|
| 7.2. | <i>Navigators</i> | 25 |
| 7.2.1. | Navigator Examples | 25 |
| 7.3. | <i>Search Fields</i> | 26 |
| 7.3.1. | Search Form Query Completion | 27 |
| 7.4. | <i>Collection Browsing</i> | 27 |
| 7.4.1. | Front Page..... | 27 |
| 7.4.2. | Browse Hierarchy | 27 |
| 7.4.3. | Presentation of articles..... | 28 |
| 7.4.4. | Complete Collection Browsing Example..... | 28 |
| 7.4.5. | Related Resources..... | 28 |
| 7.5. | <i>Citation Search</i> | 29 |
| 8. | Content Details | 30 |
| 8.1. | <i>Downloads</i> | 31 |
| 8.2. | <i>Actions</i> | 31 |
| 8.3. | <i>Header</i> | 31 |
| 8.4. | <i>Summary Fields to Display</i> | 31 |
| 9. | "mods.xml" Mapping | 33 |
| 9.1. | <i>"mods.xml" Structures</i> | 33 |
| 9.2. | <i>"mods.xml" Components</i> | 33 |
| 10. | Preservation Planning | 35 |

1. Introduction

1.1. General Description

The ERIC Collection contains records to education related materials and journals such as:

- Books
- Research syntheses
- Conference papers
- Technical reports
- Policy papers

The data is used primarily by people interested in education policy, instructors and students in teaching programs. The data comes from a myriad of different sources and prefaced by an abstract and metadata on the first page of the PDF file. The GPO collection is static with data going from 1995 to 2004.

GPO Access contains reports on federally funded education research topics from the U.S. Department of Education's Educational Resources Information Center. Reports on *GPO Access* begin with those received in October 2002. Prior reports are available from select [Federal depository libraries](#) nationwide in microfiche. A larger selection of ERIC reports is available from the [ERIC program](#). Files are available in Adobe Portable Document Format (PDF) only.

| PM/collectionCode | Display Name |
|-------------------|-----------------------------|
| ERIC | Education Reports from ERIC |

1.2. Document Types

There is only a single type of file for ERIC, the PDF file that contains the educational resource.

| PCS/docClass | User-readable (full) | User-readable (abbreviated) | Description |
|--------------|--|-----------------------------|---|
| ERIC | Education Resources Information Center | ERIC | ERIC indexes education journals, the majority of which are peer-reviewed. Most of these journals are indexed comprehensively - that is, a record for every article in each issue is included in ERIC. Some journals are indexed selectively - that is, only those articles that are education-related are included. |

2. "fdsys.xml" Metadata Schema Elements

2.1. Package Metadata

Notes:

- See above for XML entity abbreviations (PM and PCS)
- Items which are <blank> do not need to be included in the "fdsys.xml" file.
- All of the data values below are assumed to be strings unless specified otherwise.

| XML Entity | Description or "constant value" | Source | Arity* |
|-------------------------------|--|------------|--------|
| Generic Metadata Fields | | | |
| PM/collectionCode | "ERIC" | constant | 1 |
| PM/quality | <i>Filled out by parser (see section 4.3.1)</i> | parser | 0-1 |
| PM/scope | "fdlp" | constant | 1 |
| PM/governmentAuthor1 | "Department of Education" | constant | 1 |
| PM/governmentAuthor2 | "Education Resources Information Center" | constant | 1 |
| PM/governmentAuthor3 | <blank> | n/a | 0 |
| PM/starprintNumber | <blank> | n/a | 0 |
| PM/category | "Executive Agency Publications " | constant | 1 |
| PM/title | <i>The title of the report as parsed from the document</i> | parser | 1 |
| PM/title/@info | "from-parsing" | parser | 1 |
| PM/sourceContentType | "deposited" | constant | 1 |
| PM/packageDigitalOrigin | "born digital" | constant | 1 |
| PM/personalAuthor | <blank> | n/a | 0-1 |
| PM/branch | "executive" | constant | 1 |
| PM/typeOfResource | "text" | constant | 1 |
| PM/genre[@authority="marcgt"] | "government publication" | constant | 1 |
| PM/geographicLocation | <blank> | n/a | 0-1 |
| PM/dateIssued | <i>The date of the document as parsed from its summary metadata.</i> | parser | 1 |
| PM/dateCreated | <blank> | n/a | 0-1 |
| PM/dateCopyrighted | <blank> | n/a | 0-1 |
| PM/dateValid | <blank> | n/a | 0-1 |
| PM/dateModified | <blank> | n/a | 0-1 |
| PM/dateIngested | <i>Date ingested into govinfo</i> | submission | 1 |
| PM/edition | <blank> | n/a | 0-1 |

| XML Entity | Description or "constant value" | Source | Arity* |
|--|--|------------|--------|
| PM/issuance | "monographic" | constant | 1 |
| PM/language | "eng" | constant | 1 |
| PM/abstract | <i>The abstract parsed from the document metadata.</i> | parser | 0-1 |
| PM/tableOfContents | <blank> | n/a | 0-1 |
| PM/subject/topic | <blank> | n/a | 0-1 |
| PM/subject/geographic | <blank> | n/a | 0-1 |
| PM/temporalSubject | <blank> | n/a | 0-1 |
| PM/classification/ [authority="sudocs"] | "ED 1.615:" | constant | 1 |
| PM/otherIdentifier [@idStandard= 'ils-system-id'] | "000573142" | n/a | 1 |
| PM/otherIdentifier [@idStandard= 'migrated-doc-id'] | <blank> | n/a | 0-1 |
| PM/otherIdentifier [@idStandard= 'stock-number'] | <blank> | n/a | 0-1 |
| PM/otherIdentifier/ [@idStandard= 'sudoc-item-number'] | <blank> | n/a | 0-1 |
| PM/otherIdentifier [@idStandard='isbn'] | <i>From document metadata.</i> | parser | 0-1 |
| PM/otherIdentifier [@idStandard='issn'] | <blank> | n/a | 0-1 |
| PM/part | <blank> | n/a | 0-1 |
| PM/waisDatabaseName | <blank> | n/a | 0-1 |
| PM/notes | <blank> | n/a | 0-1 |
| PM/recordSource | "DGPO" | constant | 1 |
| PM/recordCreationDate | <i>Date that the "mods.xml" is created</i> | processing | 1 |
| PM/recordChangeDate | <i>Date that the "mods.xml" is modified</i> | processing | 1 |
| PM/recordOrigin | "machine generated" | constant | 1 |
| PM/subTitle | <blank> | n/a | 0-1 |
| PM/creator | <blank> | n/a | 0-1 |
| PM/publisher | "U.S. Department of Education" | constant | 1 |
| PM/pageCount | <i>Computed based on the total page count of the pages in the PDF rendition.</i> | processing | 1 |
| PM/frequency | <blank> | n/a | 0-1 |

| XML Entity | Description or "constant value" | Source | Arity* |
|-------------------------------------|---|----------|--------|
| Collection-Specific Metadata Fields | | | |
| PCS/docClass | "ERIC" | constant | 1 |
| PCS/accessId | <i>The access identifier for this package, used to uniquely identify this granule to the public. Example "ERIC-ED463724"</i> | parser | 1 |
| PCS/type | <i>The type of education resource. Example, "Creative Works"</i> | parser | 1-n |
| PCS/institution | <i>The "INSTITUTION" is the originating institution for of the resource. Example "Arizona Univ., Tucson. Coll. of Education."</i> | parser | 0-1 |
| PCS/sponsorAgency | <i>The sponsoring agency of the education resource. Example "Special Education Programs (ED/OSERS), Washington, DC."</i> | parser | 0-n |
| PCS/subject | <i>The "DESCRIPTORS" attribute found in the metadata section of the PDF file</i> | parser | 0-n |
| PCS/identifiers | <i>The "IDENTIFIERS" attribute found in the metadata section of the PDF file</i> | parser | 0-n |
| PCS/ericNumber | <i>[KEY]The ERIC # for the report. This the file name without the extension. Example "ED464427"</i> | parser | 1 |
| PCS/personalAuthor | <i>The author(s) of the ERIC resource as parsed from the document metadata. Example, "Kaput, James J."</i> | parser | 0-n |
| PCS/isFallbackTitle | <i>Will be "true" if a fallback title was used for this package, or "false" if it is a normal, descriptive title.</i> | parser | 1 |

*Arity is with respect to the parent element. For example, a parent element with an arity of "0-1" is optional and single-valued, while a child element with arity of "1" is mandatory, but only if the parent element is present.

2.2. Granule Metadata

There are no granules in this collection.

2.3. Standardized References

The ERIC DMD uses no standard references.

3. Renditions and Input Files

3.1. Renditions

| Format Name | Mime Type | Classification | From | Public | Gran-ules | Description |
|---------------|-----------------|---------------------|---------------|--------|-----------|--|
| pdf-submitted | application/pdf | production | Plant | no | no | The original PDF files |
| pdf | application/pdf | web-optimized | pdf-submitted | yes | no | The PDF rendition, for public consumption. This is separated from "pdf-submitted" since the files will be renamed and to allow for future digital signing. |
| text | text/plain | production, derived | pdf-submitted | no | no | Produced from the submitted PDF by processing. Note: This is an ACP derived rendition. It is not stored in the AIP. |

Notes:

1. The format name will be used for the rendition folder name, which will in turn be used as a component of the URL for access.
2. If the rendition is public, then it will also have classification="public-access".

3.2. Plant Processing

This is a static collection. There will be no further GPO Plant processing required.

3.3. Migration Input & Packaging

Detailed packaging rules were computed as each directory of files was migrated from the former GPO Access system. The following is a rough guide for how to determine, for each file name, how to determine the package ID to which it will be assigned.

| Rendition | Sample File Names | Instructions for determining package ID |
|---------------|-------------------|--|
| pdf-submitted | ed464338.pdf | Remove extension from file name. |
| text | ed464338.txt | This rendition is generated from the "pdf-submitted rendition". It should be generated following the output format of the "pdftotext" tool. The name of the file is the same as the PDF rendition but with a ".txt" extension instead. For example, "ed464338.pdf" yields "ed464338.txt". The file is set on the "text" rendition folder of the package. |

3.4. Day Forward Input

This is a static collection. There will be no day-forward processing required.

4. Parsing

4.1. Renditions

This collection will be parsed using the text rendition produced from the submitted PDF. The DMD assumes the text output resembles the Adobe "Save as Text" format. Dates for all renditions should be converted to YYYY-MM-DD format. Also, only dates should be stored. There are no date-time formats in this collection.

4.2. Parsing Text

All the metadata for the document is found on the first page.

4.2.1. Preprocessing

Reduce the parsing scope to the first page by excluding all other pages. The document will have a top header separating the pages. The second header instance in the document marks the end of the first page. Below is an example of the header pattern to match:

```
7 organizations that offer help to families, caregivers, and teachers and an
annotated list of 11 print resources.) (CR)
```

```
Reproductions supplied by EDRS are the best that can be made
from the original document.
```

```
This is the background image for an Adobe Acrobat Capture OCR page with image
plus hidden text.
```

```
Perspectiva General sobre la Sordo-Ceguera. DB-LINK.
Diciembre de 1995
Por
Barbara Miles
```

4.2.2. Granules

There are no granules associated with this collection.

4.2.3. Parsing Metadata

All the metadata about the content is found on the first page of the document and will be parsed in a similar fashion. The few exceptions will be noted in subsequent sections. Below is a sample of the metadata from the text rendition.

DOCUMENT RESUME

```
AUTHOR      Kleiner, Anne; Lewis, Laurie
TITLE       Internet Access in U.S. Public Schools and Classrooms: 1994-
            2002. E.D. Tabs.
INSTITUTION National Center for Education Statistics (ED), Washington,
            DC.; Westat, Inc., Rockville, MD.
REPORT NO   NCES-2004-011
PUB DATE    2004-00-00
```

| | |
|----------------|---|
| NOTE | 85p.; Project Officer, Bernard Greene. For the 1994-200'1 edition, see ED 472 678. |
| AVAILABLE FROM | For full text: http://nces.ed.gov/pubs2004/2004011.pdf . |
| PUB TYPE | Numerical/Quantitative Data (110) -- Reports - Research (143) -- Tests/Questionnaires (160) . . |
| EDRS PRICE | EDRS Price MF01/PC04 Plus Postage. |
| DESCRIPTORS | *Classroom Environment; Educational Equipment; Information Dissemination; Information Technology; *Internet; Public Education; *Public Schools |
| ABSTRACT | <p>This report presents data on Internet access in U.S. public schools from 1994 to 2002 by school characteristics. It provides trend analysis on the progress of public schools and classrooms in connecting to the Internet and on the ratio of students to instructional computers with Internet access. For the year 2002, this report also presents data on the types of Internet connections used; student access to the Internet outside of regular school hours; laptop computer loans; hand-held computers for students and teachers; and school Web sites. It also contains information on computer hardware, software, and Internet support and Web site support at the school; teacher professional development on how to integrate the use of the Internet into the curriculum; and technologies and procedures to prevent student access to inappropriate material on the Internet. Appended are the Methodology and Technical Notes; and Questionnaire. (Contains 43 tables and 4 figures.) (Author)</p> <p>Reproductions supplied by EDRS are the best that can be made from the original document.</p> |

All documents start with the text "DOCUMENT RESUME". But the first metadata does not appear until the line starting with "AUTHOR". The author metadata element is typically the first element to be found, but "TITLE" has also shown up where there is no author element. All metadata element names are all in capital letters. The other important factor to notice is that values for the metadata elements start always from a particular character spacing in document. In the previous example the metadata values start at character 19. Elements with multiple values have the value delineated by a ";" (semi-colon).

Common actions:

- Remove any carriage returns or end of line codes.
- Do not include the last period as part of the value.
- Set the {PM/title/@info} metadata attribute to "from-parsing".
- Remove any asterisks (i.e. "*") from the values
- Remove any semi-colons (i.e. ";") from the values

- Remove the extra "ISBN" from the ISBN values only.

Table of mappings from element name in the text to specific metadata values:

| Element from Text Rendition | Metadata Name |
|-----------------------------|--|
| TITLE | PM/title |
| AUTHOR | PCS/personalAuthor |
| PUB DATE | PM/dateIssued |
| ISBN | PM/otherIdentifier[@idStandard='isbn'] |
| PUB TYPE | PCS/type |
| ABSTRACT | PM/abstract |
| INSTITUTION | PCS/institution |
| SPONS AGENCY | PCS/sponsorAgency |
| DESCRIPTORS | PCS/subject |
| IDENTIFIERS | PCS/identifiers |

4.2.4. PM/title and PCS/isFallbackTitle

If no title can be parsed from the document, set {PM/title} to the following fallback title:

"Education Report {PCS/ericNumber}"

Where {PCS/ericNumber} is formatted with intervening spaces as follows:

{alphaprefix} {ddd} {ddd}

And "ddd" segments are taken in order from the numeric suffix.

For example, ED464761 is formatted as "ED 464 761".

If the fallback title is used, be sure to set {PCS/isFallbackTitle} to "true". Otherwise, set it to "false".

4.2.5. PM/dateIssued

The {PM/dateIssued} value is found on the line for the "PUB DATE". For example:

| | |
|--------------|--|
| SPONS AGENCY | Special Education Programs (ED/OSERS), Washington, DC. |
| PUB DATE | 1995-12-00 |
| NOTE | lop.; For the English version, see ED 436 056. |

Here the {PM/dateIssued} is "1995-12-01".

Note:

- The format is interpreted as YYYY-MM-DD
- The "PUB DATE" can set the month and date value to "00".
- When the month or date is set as "00", default the value to "01".

4.2.6. PM/abstract

The {PM/abstract} data can be found at the end of the metadata section. It is preceded by the uppercase "ABSTRACT" markup. It is terminated by a pattern "Reproductions supplied by EDRS are the best that can be made". It is important to note that the value for {PM/abstract} does not start on the same line as the "ABSTRACT" element. It starts on the next line and frequently does not use the same character starting position.

For example:

| | |
|-------------|---|
| DESCRIPTORS | *Federal Aid; Higher Education |
| ABSTRACT | This guide explains student financial aid programs the U.S. Department of Education's Federal Student Aid (FSA) office administers. The first three pages are a quick reference; the rest of the publication provides more of what you need to know about the financial aid programs offered. (AMT) |
| | Reproductions supplied by EDRS are the best that can be made |

Please note the abstract should be truncated to the "." prior to the terminating pattern. As with other metadata elements remove any carriage returns in the metadata, but not the trailing period.

4.2.7. PCS/accessId

The formula for deriving the access identifier:

{PM/collectionCode}-{PCS/ericNumber}

Example:

ERIC-ED463948

4.2.8. PCS/type

The {PCS/type} is found after the "PUB TYPE" tag on the metadata page in the document. There can more than one identifier per document and they should all be indexed. Unlike other metadata values, the values for this element are separated by "--".

For Example:

| | |
|----------------|--|
| AVAILABLE FROM | For full text: http://nces.ed.gov/pubs2004/2004011.pdf . |
| PUB TYPE | Numerical/Quantitative Data (110) -- Reports - Research (143) -- Tests/Questionnaires (160) . . |
| EDRS PRICE | EDRS Price MF01/PC04 Plus Postage. |

The {PCS/type} values for this example are "Numerical/Quantitative Data", "Reports – Research", "Tests/Questionnaires".

- As with other metadata elements remove any carriage returns in the metadata.
- Additionally, remove any text in parentheses encountered and the "--" characters.

- Frequently values have dash in the data such as "Reports -Evaluative (142)". The dash should have a single space on either side, so the value would appear as "Reports - Evaluative".
- Set each value into a separate {PCS/type} metadata field.
- Note: If unable to parse out a {PCS/type} set it to "Other".

4.2.9. PCS/ericNumber

The {PCS/ericNumber} is the file name of the document, dropping the extension. Convert the file name to uppercase before assigning the value to {PCS/ericNumber}.

For Example:

```
Directory of \data\2000\pdf-submitted
05/05/2009 01:40 PM <DIR> .
05/05/2009 01:40 PM <DIR> ..
04/02/2009 02:55 PM 4,604,128 ed463510.pdf
04/02/2009 02:55 PM 5,670,425 ed463572.pdf
04/02/2009 02:55 PM 4,020,631 ed463617.pdf
04/02/2009 02:55 PM 1,501,243 ed463618.pdf
.
.
.
```

The files highlighted would have the {PCS/ericNumber} of ED463510, ED463572, ED463617, and ED463618 respectively.

4.3. Validation Heuristics

The following validation heuristics should be checked by the parser and added to the element in the "fdsys.xml" value as a "quality=" attribute or the <quality> element.

4.3.1. Validating Granules and Packages

There are two quality elements for flagging quality issues encountered when parsing packages and granules.

PM/quality (and PM/quality/@quality)

GM/quality (and GM/quality/@quality)

In general, quality/@quality should specify either "error", "low", or "medium". The text inside the <quality> element should provide some descriptive text as to why the quality was tagged as such.

| Situation | Metadata | @quality | Descriptive Text |
|-------------------------------|------------|----------|--|
| Unrecognized file name format | PM/quality | error | "Unrecognized file name format" |
| Incorrect format | PM/quality | error | "Text file format appears to contain locator data" |

5. govinfo Processing

This section describes in detail all of the steps required to process the collection files inside of **govinfo**. This includes creating renditions, creating the table of contents, mapping metadata, expected manual edits, etc.

5.1. Special Manual Interventions Required

There are no special manual interventions required for this collection.

5.2. Text Creation

The PDF files will need to be converted to text using the "pdftotext" tool.

Text files will:

- Be stored in the "text" rendition
- Have the same file name as the PDF file, except...
- Will have a ".txt" extension instead a ".pdf"

For example: "ed453610.pdf" will be run through the "pdftotext" tool and will result in an "ed453610.txt" file which will be stored in the text rendition.

5.3. PDF Processing

5.3.1. Renaming PDF files

The submitted PDF file for the entire package will need to be copied to the "pdf" rendition and renamed. The original file name is of the form "ed{digits}.pdf" and should be renamed to a file name of the form "{PCS/accessId}.pdf". For example, the file "ed483019.pdf" should be copied to the "pdf" rendition and renamed "ERIC-ED483019.pdf".

5.4. HTML Processing

There will be no HTML rendition for this collection. Only a PDF rendition will be made public. The text version of the files will not be made public in any way. They are used only for parsing.

6. Content Publishing and Indexing

6.1. Indexing Granularity

There will be an entry in the public web site search engine indexes for:

1. The package as a whole
 - This entry will be used for the "Browse by Date" feature on **govinfo** as well as for simple and advanced searches.
 - The PDF content file for the package will be indexed with this entry, as well as the metadata

6.2. Index-profile field mapping

Notes:

1. {pdfobj} is computed as follows:


```
/fdsysPackage/contentSect/rendition[label/mime='application/pdf' and
isPublic='true']/digitalObject/techMdGroup
```
2. There are no granules in ERIC. All the fields defined below belong to the package.

| Index-profile Field | From "fdsys.xml" Entity(ies) and <i>Special Instructions</i> | Purpose |
|---|---|---------------------|
| Standard ESP Fields | | |
| title | {PM/title} | results, sorting |
| getpath | file://{package-directory-location}/{pdfobj}/filePath <i>Note: "getpath" is a special field for indexing document content by file name into the "body" and "content" fields.</i> | full content search |
| teaser | {PM/abstract} | results |
| contenttype | "text/html" | indexing control |
| language | {PM/language} | indexing control |
| charset | "utf-8" | indexing control |
| url | https://www.govinfo.gov/content/pkg/{PCS/accessId}/ {pdfobj}/filePath | admin & testing |
| Standard Document Types and Identifiers | | |
| accode | {PM/collectionCode} | navigator, results |
| packageid | {PCS/accessId} | results |
| granuleid | <blank> | n/a |
| docClass | {PCS/docClass} | navigator, results |
| granuleClass | <blank> | n/a |
| processingcode | {PM/collectionCode} | results |
| Granule and Collection Hierarchy Fields | | |

| Index-profile Field | From "fdsys.xml" Entity(ies) and Special Instructions | Purpose |
|--------------------------------------|--|--------------------------|
| nodeclass | "simple;package;browse;" | search control |
| treesort | <i>(See next section below)</i> | sorting |
| ancestors | <blank> | n/a |
| thisnode | <blank> | n/a |
| Publishing Specific Fields | | |
| publishdate | {PM/dateIssued} | sorting |
| publishdatehier | {PM/dateIssued} <i>Formatted as "YYYY; YYYY/MM; YYYY/MM/DD"</i> | hierarchical navigator |
| publishyear | {PM/dateIssued} <i>Formatted as "YYYY"</i> | navigator |
| publishmonth | {PM/dateIssued} <i>Formatted as "MM"</i> | navigator |
| publishweek | {PM/dateIssued} <i>Formatted as "MM-DD/W" where the MM-DD/W is the month and day of the Friday on or after the {PM/dateIssued} OR the last of the month, whichever is earlier.</i> | navigator |
| publishday | {PM/dateIssued} <i>Formatted as "MM-DD/W" where "W" is the numeric day of the week where 1=Sunday and 7=Saturday</i> | navigator |
| publishmonthyear | {PM/dateIssued} <i>Formatted as "YYYY-MM"</i> | navigator for browse |
| firstpage | <blank> | results, citation search |
| lastpage | <blank> | results, citation search |
| pageprefix | <blank> | |
| governmentauthor | {PM/governmentAuthor1}; {PM/governmentAuthor2}; | navigator |
| xmldata | <i>(a copy of the "mods.xml")</i> | advanced search |
| Fields for Relevancy Ranking Control | | |
| grank1 | {PCS/ericNumber}; {ericnum-formatted} <i>Note: {ericnum-formatted} is {PCS/ericNumber} formatted as "{alphprefix} {ddd} {ddd}. Where "ddd" are successive digits from the numeric suffix. For example: "ED463412" becomes "ED 463 412".</i> | relevancy ranking |
| grank2 | {PM/title}; "Education Reports from ERIC" | relevancy ranking |
| grank3 | {PCS/institution}; {PCS/sponsorAgency}; {PCS/personalAuthor}... <i>Note: Include all occurrences of {PCS/personalAuthor} fields separated by semi-colons.</i> | relevancy ranking |

| Index-profile Field | From "fdsys.xml" Entity(ies) and <i>Special Instructions</i> | Purpose |
|-----------------------------------|---|-------------------|
| grank4 | {PM/abstract}; {PCS/subject}; ... {PCS/identifier}; ... <i>Note: Include all {PCS/subject} and {PCS/identifier} values separated by semi-colons.</i> | relevancy ranking |
| grank5 | <blank> | relevancy ranking |
| grank6 | <blank> | relevancy ranking |
| File Access Fields | | |
| pdffile | {pdfobj}/filePath | content delivery |
| pdfsize | {pdfobj}/fileSize | content delivery |
| htmlfile | <blank> | n/a |
| htmlsize | <blank> | n/a |
| other1file | <blank> | n/a |
| other1size | <blank> | n/a |
| other1mime | <blank> | n/a |
| other2file | <blank> | n/a |
| other2size | <blank> | n/a |
| other2mime | <blank> | n/a |
| Common govinfo Metadata | | |
| branch | PM/branch | navigator |
| chamber | <blank> | navigator |
| category | PM/category | navigator |
| Standard Navigator Add-Ins | | |
| fdsys_orgs | {PCS/institution}; {PCS/sponsorAgency} | navigator |
| fdsys_people | {PCS/personalAuthor}; ... <i>Note: There may be multiple {PCS/personalAuthor} values. Include them all separated by semi-colons.</i> | navigator |
| fdsys_locations | <blank> | navigator |
| fdsys_concepts | {PCS/type}; ... {PCS/subject}; ... {PCS/identifier}; ... <i>Note: There may be multiple values for the three metadata fields. Include them all separated by semi-colons.</i> | navigator |
| Collection-Specific Fields | | |
| generic1val | {PCS/type}; ... <i>Note: Include all values separated by semi-colons.</i> | navigator |
| agencies | {PCS/sponsorAgency} | navigator |

| Index-profile Field | From "fdsys.xml" Entity(ies) and <i>Special Instructions</i> | Purpose |
|---------------------|---|------------------------|
| categoryhier | Subject/{PCS/subject}; ... Identifier/{PCS/identifier}; ... Type/{PCS/type}; ... <i>Notes:</i> <ol style="list-style-type: none"> There can be multiple subjects, identifiers and types. Include all of them, separated by semicolons. For each entry set the appropriate prefix constant. <i>For example:</i> Subject/Children; Subject/Communication Skills; Subject/Communication (Thought Transfer); Subject/Deaf Blind; Type/Guides; Type/Non-Classroom | hierarchical navigator |
| r.ericnumber | {PCS/ericNumber} | results |
| r.fericnumber | {PCS/ericNumber} <i>Note: Format ERIC number as "{alphaprefix} {ddd} {ddd}".</i> <i>Where {ddd} are digits from the numeric suffix. For example, "ED 463 412"</i> | results |
| r.isfallbacktitle | {PCS/isFallbackTitle} | results |

Note: Fields prefixed with "r:" are stored in the "resultsbundle" index-profile field, as a nested list of name/value pairs. These pairs are unbundled for display as needed by the search API layer.

6.3. Computing "treesort"

The "treesort" field will be used to sort the documents when being displayed for collection browsing. Tree sort will be made up of the following components, each separated by a forward slash:

| Order | "fdsys.xml" Field | Formatting and <i>Special Instructions</i> |
|-------|-------------------|--|
| 1 | PM/dateIssued | Format to YYYY |
| 2 | PCS/ericNumber | |

For example:

1996/ED438032

1996/ED439239

6.4. Solr Fields

For compatibility and convenience purposes, the index fields defined in the Index Profile, i.e. the index transform file named "index.xsl", remain exactly as described in the previous sections. However, upon Publishing, these fields are mapped to Solr fields and additional fields are also defined.

| Index-Profile Field | Action | Solr Field Name | Type | Notes |
|--|-----------------|-----------------|------------------|---|
| Standard Fields | | | | |
| title | copy | title | text tokens | The 2 nd field is used for sorting. |
| | | title_exact | string | |
| getpath | file extraction | <none> | | The file referenced here is used to populate "body" and then discarded. |
| | | body | text tokens | |
| teaser | flag teaser | teaser | string | Stored, not indexed! Generated from "body" for granules or from file access fields for package. |
| contenttype | copy | contenttype | text tokens | |
| language | copy | language | string | Default is "eng" |
| charset | ignore | <none> | | Not required for Solr |
| url | copy | url | string | |
| Standard Document Types and Identifiers | | | | |
| accode | multi-valued | accode | lowercase tokens | |
| packageid | copy | packageid | lowercase | Unique Solr Doc Id – only use packageid when not a granule |
| | | id | string | |
| docclass | copy | docclass | lowercase tokens | |
| processingcode | copy | processingcode | lowercase | |
| Granule and Collection Hierarchy Fields | | | | |
| nodeclass | multi-valued | nodeclass | lowercase | |
| treesort | copy | treesort | string | |
| Publishing Specific Fields | | | | |
| publishdate | date/time | publishdate | date | |
| publishdatehier | multi-valued | publishdatehier | string | |
| publishyear | copy | publishyear | string | |
| publishmonth | copy | publishmonth | string | |
| publishweek | copy | publishweek | string | |
| publishday | copy | publishday | string | |

| Index-Profile Field | Action | Solr Field Name | Type | Notes |
|--|----------------|-------------------|------------------|--|
| publishmonthyear | copy | publishmonthyear | string | |
| governmentauthor | multi-valued | governmentauthor | text tokens | |
| xmldata | xml | xml | xml tokens | MODS data |
| Fields for Relevancy Ranking Control | | | | |
| grank1 | copy | grank1 | text tokens | |
| grank2 | copy | grank2 | text tokens | |
| grank3 | copy | grank3 | text tokens | |
| grank4 | copy | grank4 | text tokens | |
| File Access Fields (Stored, not indexed!) | | | | |
| pdffile | package teaser | pdffile | string | Stored, not indexed! Generate package teaser from this file if not yet set. |
| pdfsize | copy | pdfsize | integer | Stored, not indexed! |
| Common govinfo Metadata | | | | |
| branch | copy | branch | lowercase tokens | |
| category | multi-valued | category | lowercase tokens | |
| Standard Navigator Add-Ins | | | | |
| fdsys_orgs | multi-valued | companies | lowercase tokens | |
| fdsys_people | multi-valued | personnames | lowercase tokens | |
| fdsys_concepts | multi-valued | concepts | lowercase tokens | |
| Collection-Specific Fields | | | | |
| categoryhier | multi-valued | categoryhier | string | |
| agencies | multi-valued | agencies | lowercase tokens | |
| generic1val | copy | generic1val | lowercase tokens | |
| Results Bundle (copied to individual fields that are stored but not indexed) | | | | |
| r.ericnumber | copy | r_ericnumber | string | Stored, not indexed! |
| r.fericnumber | copy | r_fericnumber | string | Stored, not indexed! |
| r:isfallbacktitle | copy | r_isfallbacktitle | string | Stored, not indexed! |
| Navigator Fields (See Section 7.2) | | | | |

Key to Action Column Values:

| Action | Explanation |
|-----------------|--|
| copy | Use the Index Profile Field as is |
| ignore | Discard |
| multi-valued | Break the value up into separate values (usually separated by semi-colons) that each has the same Solr field name |
| date/time | Convert to standard date/time format, appending "T12:00:00Z" if no time is specified |
| file extraction | The value of the field is a file. Store the contents of the file and discard the file name. |
| flag teaser | For granules. If "teaser" is not already set, generate it from file contents stored in "body" |
| package teaser | For packages. If "teaser" is not already set, generate it from pdf file, html file, or xml file contents, whichever comes first. |
| xml | This field is processed by a special MODS XML handler |

Key to Type Column Values:

| Type | Tokenized | Explanation |
|--------------------------|---------------------------|---|
| string | no | Treat text as one single token. |
| integer | no | Treat as a 32-bit integer. |
| date | no | Full standard date and time string. E.g. "1970-01-01T23:59:59Z" |
| lower case | no | As a "string" only converted to all lower case first. |
| lower case tokens | whitespace | As "lower case" but tokenized on whitespace. |
| text tokens | whitespace punctuation | lower case, lemmatized (stemming), punctuation removal, etc. |
| text tokens unstemmed | whitespace punctuation | As "text tokens", but without any lemmatization (stemming). |
| xml tokens | whitespace punctuation | As "text tokens", but recognizes and handles XML tags. |
| boundary | no | Prefix the value with "XXPREFIX " and append " XXSUFFIXXX". |

7. Search and Browse

7.1. Search Results Presentation

| Line | Pattern (using index-profile fields) and <i>Special Instructions</i> |
|------|--|
| 1 | <p>{r.fericnumber} – {title}</p> <p><i>Notes:</i></p> <ol style="list-style-type: none"> 1. The title should link to the Content Details page. 2. If {r:isfallbacktitle} is "true", use "Education Report from ERIC" instead of {title} |
| 2 | <p><i>Education Reports from ERIC. {agencies}.{publishdate}.</i></p> <p><i>Note: {publishdate} is formatted as "DayOfWeek, Month Day, Year", e.g. "Friday, May 2, 2008"</i></p> |
| 3 | {teaser} |
| 4 | A row of buttons linking to the PDF content rendition, the Content Details page, and the electronic sharing feature respectively. |

Examples:

i)

ED 481 811 - Status and Trends in the Education of Blacks

Education Reports from ERIC. Monday, September 1, 2003.

... DOCUMENT RESUME AUTHOR TITLE INSTITUTION REPORT NO PUB DATE NOTE AVAILABLE FROM PUB TYPE EDRS PRICE DESCRIPTORS IDENTIFIERS ABSTRACT black children Hoffman, Kathryn; Llagas, Charmaine; Snyder, Thomas D. Status and Trends in the Education of Blacks. National Center for Education Statistics (ED), Washington, DC. NCES-2003-034 2003-09-00 188p. ED...

[PDF](#) [DETAILS](#) [SHARE](#)

ii)

ED 472 451 - Including Special-Needs Students in the NAEP 1998 Reading Assessment. Part I: Comparison of Overall Results with and without Accommodations. A Report on 1998 NAEP Research Activities. Statistical Analysis Report

Education Reports from ERIC. Saturday, February 1, 2003.

... DOCUMENT RESUME ED 472 451 EC 309 663 AUTHOR Lutkus, Anthony D. ; Mazzeo, John TITLE INSTITUTION REPORT NO PUB DATE NOTE AVAILABLE FROM PUB TYPE EDRS PRICE DESCRIPTORS IDENTIFIERS Including Special-Needs Students in the NAEP 1998 Reading Assessment. Part I: Comparison of Overall Results with and without Accommodations. A Report on 1998 NAEP...

[PDF](#) [DETAILS](#) [SHARE](#)

iii)

ED 472 649 - Education Report from ERIC

Education Reports from ERIC. Friday, November 1, 2002.

... ED 472 649 DOCUMENT RESUME EA 032 306 AUTHOR TITLE' INSTITUTION REPORT NO PUB DATE NOTE AVAILABLE FROM PUB TYPE EDRS PRICE DESCRIPTORS IDENTIFIERS McDowell, Lena M.; Sietsema, John P. Directory of Public Elementary and Secondary Education Agencies, 2000-2001. National Center for Education Statistics (ED), Washington, DC . NCES-2003-310 2002-11-0...

[PDF](#) [DETAILS](#) [SHARE](#)

7.2. Navigators

Collection-specific navigators are listed here.

| Display Name | Navigator | Index-profile Field | Type | Purpose |
|-------------------|-----------------|---------------------|----------------------|-----------------------|
| Document Category | categoryhiernav | categoryhier | string, hierarchical | results, hierarchical |

7.2.1. Navigator Examples

7.2.1.1. Document Category

Note: The quality of this navigator will be evaluated in development.

{categoryhier}

Navigator values are already formatted at the index level; no need for post-processing.

- Identifier (4) >
- Subject (4) >
- Type (4) v
 - ERIC Publications - - ERIC Digests in Full Text (2)
 - Reports - Descriptive (1)

[See More](#)

7.3. Search Fields

Notes:

1. All of the standard fields, even though they are not listed below, must be available for this collection.
2. The advanced search form will be made up of all of the fields with the "Search Form" column equal to "Y".
3. Developers are encouraged to use a pull-down box or query completion for elements with pre-defined values.
4. When two or more collections are selected by a user on the search form, the metadata in the pull-down should be a join of the common elements.

| Display Name | Search field: all one word | Search QPL | Search Form | Type | Allowed Values | Help Text |
|--------------------|----------------------------|---|-------------|--------|----------------|---|
| ERIC Number | ericnumber | xml:extension:ericnumberformatted:({s}) | Y | string | | Search for an ERIC resource based on its number. Example ERIC Numbers are "ED466380", or "ED 483 022", (either format is allowed). |
| Subject | ericsubject | xml:extension:subject:({s}) | Y | string | | Search for an ERIC resource based on its subject. Examples subjects are "Cognitive Psychology", "Elementary Education" and "Story Reading" |
| Identifiers | ericidentifier | xml:extension:identifiers:({s}) | Y | string | | The topical area covered by the ERIC resource. Example Identifiers are "Family Activities; *Read Along; Team Learning". |
| Sponsoring Agency | sponsor agency | xml:extension:sponsoragency:({s}) | Y | string | | Search for an ERIC resource based on the Agency that funded the creation of the resource. An example agency is "Office of Educational Research and Improvement (ED), Washington, DC." |
| Source Institution | institution | xml:extension:institution:({s}) | Y | string | | The institution that created the ERIC resource, for example "Family Learning Association, Bloomington, IN." |

7.3.1. Search Form Query Completion

The following fields will be present in the query completion server.

| Advanced Search Field | QC Value Prefix | Data comes from "fdsys.xml" fields |
|-------------------------|-----------------|---|
| Government Author* | GAU | {PM/governmentAuthor1 }, {PM/governmentAuthor2}, {PM/governmentAuthor3} |
| SuDoc Class Number* | GSU | {PM/classification[@authority='sudocs']} |
| Type Education Resource | ERR | {PCS/type} |
| Institution | ERI | {PCS/institution} |
| Sponsor Agency | ERA | {PCS/sponsorAgency} |
| Subject | ERS | {PCS/subject} |
| Identifiers | ERD | {PCS/identifiers} |

This is not implemented in release and is designed to be implemented in future releases.

7.4. Collection Browsing

7.4.1. Front Page

The following is the description of the collection to be presented to public users:

About Education Reports from ERIC

Find reports on federally funded education research topics from the U.S. Department of Education's Educational Resources Information Center. Reports on **govinfo** begin with those received by GPO in October 2002. The documents received from the Educational Resources Information Center have dates that range from 1995 to 2004. Prior reports are available from select Federal depository libraries nationwide in microfiche. A larger selection of ERIC reports is available from the [ERIC program web site](#). Files are available in Adobe Portable Document Format (PDF) only.

The text, "[ERIC program web site](#)", links to: <http://eric.ed.gov/>

The **Help** button on the Browse page for this collection will link to: <https://www.govinfo.gov/help/eric>

7.4.2. Browse Hierarchy

There will be one method for collection browsing for RIC2: by year of the date issued. Articles within each year are sorted by their "treesort" values.

| Level | Display Name | Navigator | Purpose / Special Instructions |
|-------|------------------|------------|---|
| 1 | Date Issued Year | pubyearnav | Not applicable. Display year as YYYY from {PM/dateIssued} |

The hierarchy method will be: Ancestor siblings.

7.4.3. Presentation of articles

The articles at the leaf of the browse-by-date tree should be presented the same as specified in section 7.1, with the following exceptions:

1. Add the links from line 4 to the end of line 1.
2. Remove the collection name and date from line 2.
3. Remove line 3 and 4 completely.

For example:

ED 471 693 - High School Counselor's Handbook, 2003-04

[PDF](#) [DETAILS](#) [SHARE](#)

Office of Student Financial Assistance (ED), Washington, DC.

Note: Articles are to be sorted by "treesort" (see section 6.3).

7.4.4. Complete Collection Browsing Example

Browse Education Reports from ERIC

The screenshot shows a web interface for browsing ERIC Education Reports. It features a tree view on the left with years 2004, 2003, and 2002. The 2002 year is expanded, showing a list of three articles. Each article entry includes a title, a subtitle, the issuing office, and three buttons: PDF, DETAILS, and SHARE. A 'Sort' dropdown menu is visible in the top right corner of the article list.

| Year | Article ID | Title | Office | PDF | DETAILS | SHARE |
|------|------------|---|--|-----|---------|-------|
| 2004 | | | | | | |
| 2003 | | | | | | |
| 2002 | ED 463 411 | Effective Advisory Committees. In Brief: Fast Facts for Policy and Practice | Office of Vocational and Adult Education (ED), Washington, DC. | PDF | DETAILS | SHARE |
| 2002 | ED 463 445 | High Schools That Work: Best Practices for CTE. Practice Application Brief No. 19 | Office of Educational Research and Improvement (ED), Washington, DC. | PDF | DETAILS | SHARE |
| 2002 | ED 463 446 | Learning Disabilities and Career Development. Practice Application Brief No. 20 | Office of Educational Research and Improvement (ED), Washington, DC. | PDF | DETAILS | SHARE |

7.4.5. Related Resources

The following is the list of "Related Resources" which should be displayed at the bottom of the page when browsing this collection:

- Additional ERIC Reports
 - Display text: From the Department of Education website
 - Links to: <http://eric.ed.gov/>
- Contact your local Federal Depository Library
 - Display text: To access a physical copy of the ERIC Education Reports.

- Links to: <http://catalog.gpo.gov/fdlpdir/public.jsp>
- Purchase printed education related publications
 - Display text: From GPO's Online Bookstore
 - Links to: <https://bookstore.gpo.gov/catalog/education-libraries>

7.5. Citation Search

There is no citation search for this collection.

8. Content Details

This section covers any collection-specific needs for accessing the documents and metadata on the Content Details page. Here is an example of a Content Details page for this collection.

Download

 PDF

 MODS

 PREMIS

 ZIP

Actions

 Browse Education Reports

 CGP Record

 Share

Content Details

ED463563 - Preventing Bullying. ERIC Digest

Summary

| | |
|----------------------------------|---|
| <u>Category</u> | Executive Agency Publications |
| <u>Collection</u> | Education Reports from ERIC |
| <u>SuDoc Class Number</u> | ED 1.615: |
| <u>Date Issued</u> | March 1, 2002 |
| <u>Author</u> | Lumsden, Linda |
| <u>Source Institution</u> | ERIC Clearinghouse on Educational Management, Eugene, OR |
| <u>Sponsoring Agency</u> | Office of Educational Research and Improvement (ED), Washington, DC |
| <u>Publication Type</u> | ERIC Publications - - ERIC Digests in Full Text |
| <u>Subject</u> | Academic Achievement, Antisocial Behavior, Bullying, Elementary Secondary Education, Group Dynamics, Intervention, Peer Acceptance, Peer Influence, Peer Mediation, Prevention, School Responsibility, Violence |
| <u>Identifiers</u> | ERIC Digests |
| <u>Abstract</u> | Students who are the target of bullying episodes commonly suffer serious, long-term academic, physical, and emotional consequences. Unfortunately, school personnel often minimize, underestimate, tolerate, or ignore the extent of bullying and the harm it can cause. This digest examines the problem of bullying and some of its effects, and discusses preventive steps schools and states are taking. Bullying occurs when a person willfully and repeatedly exercises power over another with hostile or malicious intent. It can be physical or psychological in form. In a survey of more than 15,000 sixth through tenth graders, 30 percent reported bullying others, being targets, or both. Everyday approximately 160,000 students stay home from school out of fear of being bullied. Victims often have difficulty concentrating on schoolwork, show low performance, and have relatively high rates of absenteeism. Psychological symptoms of victims include anxiety, depression, and low self-esteem. To counteract bullying, schools can distribute a written anti-bullying policy to everyone in the school community and apply the policy consistently and fairly. Positive socialization and anger-management skills can be taught. An anti-bullying culture can be developed among peers. Some states are beginning to require schools to adopt anti-bullying policies. (Contains 17 references.). |

As can be seen in the example above, the Content Details page is organized as follows:

- Along the left-hand side of the page is a column of links that are partitioned into two sections:

- The top section consists of Download links regarding the specific document
- The bottom section offers some additional Actions
- The remainder of the page is devoted to the Contents Details of the specific document:
 - At the top of this section is a Header
 - Below the Header is a single tabbed pane containing a Summary of metadata fields specific to this document

8.1. Downloads

The following are available from the Content Detail page, which will be provided to the public.

- PDF rendition of the document.
- MODS metadata for the document.
- PREMIS metadata for the package.
- ZIP file of the package contents including metadata files.

All URLs are in the **govinfo** standard format. See the SDD Volume XIX – Common Metadata and Standard References for more details.

8.2. Actions

The following HTML action links will be available when browsing the Content Details page of a Congressional Hearing:

- Browse the Congressional Record
 - Goes to the main CREC browse page.
- CGP Record
 - Links to https://catalog.gpo.gov/F?func=find-b&find_code=sys&request={PM/otherIdentifier}@idStandard='ils-system-id'}
 - If the ILS system ID is not available use: <https://catalog.gpo.gov>
- Share
 - Allows the user to share a link to the content detail page of the publication.

8.3. Header

The header at the top of the content-detail page will read the same as line 1 from section 7.1.

8.4. Summary Fields to Display

The following fields are to be displayed on the Content Detail page. Note that the fields are to be displayed in the same order as they are listed below.

| Display Name | "fdsys.xml" Entities & Special Instructions | Example |
|--------------------|---|--|
| Category | {PM/category} | Executive Agency Publications |
| Collection | "Education Reports from ERIC" | Education Reports from ERIC |
| SuDoc Class Number | {PM/classification[@authority='sudocs']} | ED 1.615: |
| Date Issued | {PM/dateIssued} | December 1, 2002 |
| Author | {PCS/personalAuthor}... <i>Set all entries separated by semi-colons.</i> | Ahearn, Eileen M.; Lange, Cheryl M.; Rhim, Lauren Morando; McLaughlin, Margaret J. |
| Source Institution | {PCS/institution} | National Association of State Directors of Special Education, Alexandria, VA. |
| Sponsoring Agency | {PCS/sponsorAgency} | Special Education Programs (ED/OSERS), Washington, DC. |
| Publication Type | {PCS/type}... <i>Include all entries in a comma-separated list.</i> | Reports – Evaluative. Numerical/Quantitative Data |
| Subject | {PCS/subject}... <i>Include all entries in a comma-separated list.</i> | Accountability, Case Studies, Charter Schools, Compliance (Legal), Delivery Systems |
| Identifiers | {PCS/identifiers}... <i>Include all entries in a comma-separated list.</i> | Family Activities, Read Along, Team Learning |
| Abstract | {PCS/abstract} | The message of this series of books, "Parents and Children Together," is that parents should get together with their children, talk about stories, and learn together. This book, "Learning Together," contains... |

Note: All dates must be formatted as: {month} {day}, {year}. Example: May 2, 2008.

9. "mods.xml" Mapping

Since this collection has no granules, there will only be a single "mods.xml" produced for all situations. This "mods.xml" structure is mapped out in the following section.

9.1. "mods.xml" Structures

This section defines the two different structures or components for the "mods.xml".

"mods.xml" for the entire issue:

```
<?xml version="1.0" encoding="UTF-8"?>
<mods version="3.3" ID="{PH/@id}"
  xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-3.xsd"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://www.loc.gov/mods/v3">

  {Component 1: Publication Metadata independent of package or granule}
  {Component 2: Package Metadata for the package as a whole}

</mods>
```

9.2. "mods.xml" Components

This section specifies the mapping of "mods.xml" collection-specific metadata entities from "fdsys.xml" entities.

Notes:

1. All elements below are children of the top-level <mods> tag which is specified above.
2. Items indicated with an asterisk (*) below are the same for all collections.

| MODS Schema Entry | MODS Entity Attributes | "fdsys.xml" Entity(ies) & <i>Special Instructions</i> |
|-----------------------|--|---|
| titleInfo/title | | {PM/title} |
| location/url | displayLabel="PDF rendition" access="raw object" | https://www.govinfo.gov/content/pkg/{PCS/accessId}/pdf/{PCS/accessId}.pdf |
| location/url | displayLabel="Content Detail" access="object in context" | https://www.govinfo.gov/app/details/{PCS/accessId} |
| identifier | type="preferred citation" | {PCS/ericNumber} <i>Format ERIC number as "ED {ddd} {ddd}". Where "ddd" segments are from numeric suffix. For example, ED464761 is formatted as "ED 464 761".</i> |
| extension/searchTitle | | {PCS/ericNumber}; {PM/title}; {eric-number-formatted} <i>{eric-number-formatted} is the ERIC number formatted as "ED {ddd} {ddd}". Where "ddd" segments are from numeric suffix. For example, ED464761 is formatted as "ED 464 761".</i> |

| MODS Schema Entry | MODS Entity Attributes | "fdsys.xml" Entity(ies) & Special Instructions |
|-----------------------------------|------------------------|---|
| extension/ ericNumberFormatted | | {PCS/ericNumber}; {eric-number-formatted} <i>{eric-number-formatted} is the ERIC number formatted as "ED {ddd} {ddd}". Where "ddd" segments are from numeric suffix. For example, ED464761 is formatted as "ED 464 761".</i> |
| extension | | {PCS/*} |

Note: There are no standard references to be set for this collection.

10. Preservation Planning

Designated Community

The Primary Designated Community for the ERIC collection is the U.S. Department of Education. This collection will be preserved at the linguistic, contextual, and practical level of understanding, knowledge, and experience essential to educational practice and policy as well as interpreting relevant materials sources such as think tanks, institutional repositories, research organizations, state and district offices, and technical assistance providers.

The Secondary Designated Community for the ERIC collection includes depository and non-depository libraries serving education professionals and services, as well as educators, parents, administrators, policymakers, researchers, public and private entities, childhood services organizations, and technical assistants providing support and services to the Department of Education the general public at large. This community also includes all organizations and offices within the Federal Executive, Legislative, and Judicial Government. This level of understanding can render and access content by successfully discovering content by branch, category, citation, ERIC number, government author, identifiers, series, source institution, sponsoring agency, subject, SuDoc class number, and title.

The Designated Community at-large for the ERIC collection includes all other state and local governments, data consumers, legal professionals, internal GPO units, news media, non-profit organizations, private industries, transparency organizations, libraries, academics, or interested citizens with access and the capability to render electronic government information products.

Content Submission

Content has been migrated to **govinfo** from previous system; no day-forward content submission.

Information Packages

Information packages will conform to the packaging implementation data and logical models of the **govinfo** Architecture System Design Document. All information packages can be located at the file path location and at the logical level in relation to other packages within the system by the METS package unique identifier of the package.

Content Information

Content packages contain MODS, PREMIS, METS metadata files and submitted PDF renditions.

Preservation Description Information

Provenance: Persists preservation actions, including source of ingest and processing.

Reference: All content packages are assigned a globally unique identifier throughout the repository which is independent of the underlying technical infrastructure and the physical location of the digital object.

Fixity: SHA-256 Checksum Algorithm calculated and persisted in the PREMIS file.

Context: Linking agent identifiers persistent the contextual relationship between the SIP, ACP, and AIP packages created from the submitted content.

Representation Information

The DROID file format identification processing component will identify that content is submitted in a PDF rendition file and that the file format type designation information is sustained in the MODS and PREMIS metadata. The author, physical description, relevant cataloging information, and legal citations are persisted in the MODS metadata file. This information is not parsed; this information is provided by the content submitter.

File formats (by MIME type) accepted for this collection include:

- application/pdf

See File Format Transformation Strategy document for risk-based analysis of file format obsolescence within the **govinfo** repository.

Dissemination Presentation

Content will be disseminated in a presentation which persists the look and feel of printed publications (PDF) as received from the content submitter