

ARCHIVE SERVICES REPORTS

Version 1.2, August 2007

Supercedes Version 1.0 January 2007

Supercedes *Ingest Reports and Error Reports*, version [1.1, February 2006](#) and [1.0, October 2005](#)

Email reports are produced by the Florida Digital Archive (FDA) Ingest, Dissemination, and Withdrawal services. For each successfully ingested Submission Information Package (SIP) an Ingest Report is produced; for each rejected SIP an Ingest Error Report is produced. Successful Dissemination produces a Dissemination Report, and Withdrawal produces a Withdrawal Report. This document describes the format and content of these reports, and provides some suggestions for handling them.

1. Introduction

Ingest, Dissemination and Withdrawal services produce reports from the FDA to the affiliate whose content is affected. Reports are emailed by the FDA system to the address supplied in the FCLA-Library Agreement, Appendix A.

All reports are XML files which can be viewed directly as XML or displayed using an XSL stylesheet. By default, the reports reference the stylesheet "daitss_report_xhtml.xml" which is provided on the Florida Digital Archive website. To use "daitss_report_xhtml.xml" simply download the file into the same directory in which your reports reside. To use a stylesheet of your own choosing, place your own XSL file with the same name in the report directory instead.

Individual report types are described below: Ingest (section 2), Ingest Error (section 3), Dissemination (section 4) and Withdrawal (section 5). Section 6 contains recommendations for local processing of reports.

2. Ingest Reports

An Ingest Report is written when a SIP is successfully ingested into the digital archive, with or without warnings.

2.1. XML Document

The Ingest Report is an XML document that uses the schema *daitssReport.xsd*. The document is contained within a <REPORT> element. The first XML element within <REPORT> is <INGEST>.

```
<REPORT xmlns="http://www.fcla.edu/dls/md/daitss/"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.fcla.edu/dls/md/daitss/
        http://www.fcla.edu/dls/md/daitss/daitssReport.xsd">
  <INGEST IEID="E20060116_AAAAAF" INGEST_TIME="2006-01-17T02:40:11Z"
  PACKAGE="CFE0000038">
    etc.
```

A sample XML Ingest Report is available in the [Digital Archive Information](#) section of the FDA website.

2.2. Formatted Ingest Report

If the default stylesheet is used, the Ingest Report has three main sections: a header, "Agreement Info," and "Files." The header and Agreement Info are formatted as shown below.

Ingest

| Package name | Int. Entity ID | Ingest time |
|--------------|-----------------|----------------------|
| WF00000147 | E20050902_AAAAK | 2005-09-02T17:36:30Z |

Agreement Info

Account: UWF

Project: FHP

"Package name" is the name of the Submission Information Package (SIP) as submitted by the customer. A SIP is a set of one or more content data files and a descriptor, all in the same folder. See [Submitting Materials to the FCLA Digital Archive](#) for more information.

"Int. Entity ID" is the name of the "intellectual entity" (logical object) assigned by the FDA. The archive assumes each SIP contains only one logical object, such as a book, map, recording or dissertation.

"Ingest time" is the time the ingest completed in date/time format in Greenwich Mean Time:

yyyy-mm-ddThh:mm:ssZ

"Account" is the customer account code of the library owning the package.

"Project" is the project code associated with the SIP, used by the system to determine the preservation treatment of the contents of the package. It is taken from the PROJECT attribute of the AGREEMENT_INFO element in the METS package descriptor. (If the METS descriptor was created by program from an MXF file, and if multiple project codes were present in the MXF file, one project code is chosen to be used in the PROJECT attribute.)

Files

The "Files" section of the report contains detailed information about each file that is included in the archived Archival Information Package (AIP) for the package. It can have up to five subsections: Archival Attributes, Message Digests, Events, Broken Links, and Warnings.

The Archival attributes section identifies files and summarizes attributes which are detailed later in the report.

Archival Attributes

| Id | Name | Size | Pr e s e r v a t i o n | Orig i n | Gl o b a l | M e s s a g e D i g e s t s | Ev e n t s | Br o k e n L i n k s | War n i n g s |
|----------------------|---------------------------|----------|------------------------------------------------------|-------------------|------------------------|--------------------------------------------------------------------|------------------------|-------------------------------------------------|------------------------------|
| F20051007- AAAAAB | SN00000005/1893/1.tif | 10408136 | F U L L | DEP OSIT OR | false | 2 | 0 | 0 | 0 |
| F20051007- AAAAAC | SN00000005/1893/2.tif | 10185746 | F U L L | DEP OSIT OR | false | 2 | 0 | 0 | 0 |
| F20051007- AAAAAI | SN00000005/1893/8.tif | 10214154 | F U L L | DEP OSIT OR | false | 2 | 0 | 0 | 0 |
| F20051007- AAAAAJ | SN00000005/WF00000010.xml | 16159 | B I T | DEP OSIT OR | false | 2 | 0 | 1 | 1 |
| F20051007- AAAAAM | WF00000010.xml | 21116 | F U L L | DEP OSIT OR | false | 2 | 0 | 16 | 0 |
| F20051007- AAAAAN | WF00000010.xml.mxf | 16128 | F U L | DEP OSIT OR | false | 2 | 0 | 0 | 0 |

| | | | | | | | | | |
|------------|-----------------------------|-------|---|---|------|-------|-------------------|---|--------------------|
| F20051007_ | | | L | | | | | | |
| AAAAAO | WF00000010_LOC.xml.mxf | 16152 | F | U | ARC | false | 2 | 0 | 0 |
| | | | L | L | HIVE | | | | |
| | | | L | | | | | | |
| F20051007_ | | | F | U | ARC | false | 2 | 0 | 16 |
| AAAAAP | WF00000010_LOC.xml | 22717 | L | L | HIVE | | | | 0 |
| | | | L | | | | | | |
| F20051006_ | links_20051006161421/www.w | | F | U | INTE | | | | |
| AAAAAM | 3.org/2001/03/XMLSchema.dtd | 16018 | L | L | RNE | true | 2 | 0 | 0 |
| | | | L | L | T | | | | 1 |
| | | | L | | | | | | |
| F20051007_ | | | F | U | DEP | | | | |
| AAAAAL | MXF.dtd | 14268 | L | L | OSIT | true | 2 | 0 | 0 |
| | | | L | L | OR | | | | 0 |
| | | | L | | | | | | |

"Id" is the identifier assigned to the file by the system.

"Name" is the name of the file in the SIP, relative to the package directory.

"Size" is the size of the file in bytes.

"Preservation" is the preservation level assigned to the file, which will be either "Full" or "Bit." Files with preservation level "None," if any were included in the SIP, are not listed in the report, because these are not included in the Archival Information Package (the set of files and metadata actually stored by the repository).

"Origin" indicates whether the file was included in the original SIP ("Depositor") or if it was created by the FDA ("Archive") or downloaded from the Internet ("Internet").

"Global" has a value of "true" if the file is a global file and "false" otherwise. Global files are commonly-used files that are owned and managed by the repository. They are logically, but not physically, included in the Archival Information Package.

"Message Digests" is a count of the number of message digests (checksums) calculated for the file. This functions as a link to a display of the message digests later in the report.

"Events" is a count of the number of events reported for the file. If non-zero, this functions as a link to a display of the events later in the report. Not all event types are reported.

"Broken links" is a count of the number of links in the file which point to files that are referenced in the Archival Information Package but which would not be located either in the SIP or on the Internet. If non-zero, this functions as a link to a display of the missing files later in the report.

"Warnings" is a count of the number of warnings associated with the file during ingest. This functions as a link to a display of the warning messages.

Message Digests

The Message Digests section has an entry for each file, showing the value ("Message digest") and type ("Algorithm") of each message digest created.

[F20050902_AAABCL](#)

| Message Digest | Algorithm |
|-----------------------------------------|-----------|
| 4b906e3fd6a7e57b46034db01abddf6a | MD5 |
| bdc8d92bd26d91eaad0aa304255ac3d01d84323 | SHA-1 |

Events

The Events section has an entry for each event recorded for the file. Many events happen to files on Ingest, but only normalization and migration event types are recorded in the Ingest Report. Other event types, such as the verification of message digests, virus checks, and validation, can be assumed to occur on Ingest to all files and so are not reported.

[F20050902_AAABCL](#)

- Normalized version created on Ingest

Broken Links

The Broken Links section, if present, has an entry for each file containing one or more links to a file not included in the AIP. The entry lists the names of the missing files as they are referenced within the SIP. This may be a relative or absolute path, or a URL.

[F20050902_AAABDI](#)

- 1.jpg
- 2.jpg
- 3.jpg
- 4.jpg
- 5.jpg

- 6.jpg

Warnings

Warnings are given when files contain anomalies that do not prevent their being ingested into the archive. These anomalies are recorded permanently in the metadata for the file. Some warnings will cause the preservation level of a file to be downgraded from full to bit-level preservation. Only warnings of interest to the customer are given in the Ingest report.

[F20051007_AAAAAJ](#)

- **A_XML_BAD_FORMAT:** Unknown format exception (parse error)

3. Ingest Error Reports

An Ingest Error Report is created when a problem with one or more files within the SIP is so severe that the package is not ingested. The SIP is copied to a "rejects" directory and is available to the customer on request.

Ordinarily the Ingest Error Report is emailed to the email address associated with the affiliate's account as given in the ACCOUNT attribute of the <AGREEMENT_INFO> element in the SIP. However, if the error being reported is that the ACCOUNT value is missing or unrecognized, the Error Report will be sent to the archive administrator, who will attempt to determine the submitting affiliate and forward the report.

3.1. XML Document

The Ingest Error Report is an XML document that uses the schema *daitssReport.xsd*. The document is contained within a <REPORT> element. The first XML element within <REPORT> is <ERROR>.

```
<REPORT xmlns="http://www.fcla.edu/dls/md/daitss/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.fcla.edu/dls/md/daitss/
http://www.fcla.edu/dls/md/daitss/daitssReport.xsd">
  <ERROR REJECT_TIME="2006-01-04T02:50:45Z">
    etc.
```

A sample XML Error Report is available in the [Digital Archive Information](#) section of the FDA website.

3.2. Formatted Ingest Error Report

Ingest Error reports contain an error message, an identification of the SIP that was rejected, and the date and time of rejection. The contents of the error message will vary depending on the error. An example of an Ingest Error report is shown below:

Error

Descriptors must be valid
could not validate

/daitss/prod/data/ingest/work/WF00012154/E20051229_AAAAMR/WF00012154.xml

File:

/daitss/prod/data/ingest/work/WF00012154/E20051229_AAAAMR/WF00012154.xml

WF00012154 rejected 2005-10-06T20:52:09Z

4. Dissemination Reports

A Dissemination Report is written when a dissemination request is processed. If the dissemination is successful, the AIP as stored at the time of processing is re-ingested and the resultant, possibly updated AIP is exported as a DIP. Since all disseminations involve re-ingest, the bulk of the Dissemination Report contains the same information as an Ingest Report.

4.1. XML Document

The Dissemination Report is an XML document that uses the schema *daitssReport.xsd*. The document is contained within a <REPORT> element. The first XML element within <REPORT> is <DISSEMINATION>.

```
<REPORT xmlns="http://www.fcla.edu/dls/md/daitss/"
```

```

xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.fcla.edu/dls/md/daitss/
http://www.fcla.edu/dls/md/daitss/daitssReport.xsd">
  <DISSEMINATION IEID="E20070103_AAAAAA"
    INGEST_TIME="2007-01-03T16:05:22Z" PACKAGE="FDA0000200">
    etc.

```

A sample XML Dissemination Report is available in the [Digital Archive Information](#) section of the FDA website.

4.2. Formatted Dissemination Report

If the default stylesheet is used, the Dissemination Report has three main sections: a header, "Agreement Info," and "Files." The header and Agreement Info are formatted as shown below.

Dissemination

dissemination event info

| Package name | Int. Entity ID | Ingest time |
|--------------|------------------|----------------------|
| FDA0000200 | E20070103_AAAAAA | 2007-01-03T16:05:22Z |

Agreement Info

| | |
|-----------------|-----|
| Account: | FDA |
| Project: | FDA |

"Package name" is the name of the Submission Information Package (SIP) as originally submitted by the customer.

"Int. Entity ID" is the name of the "intellectual entity" (logical object) assigned by the FDA. The archive assumes each SIP contains only one logical object, such as a book, map, recording or dissertation.

"Ingest time" is the time the re-ingest completed in date/time format in Greenwich Mean Time:

yyyy-mm-ddThh:mm:ssZ

"Account" is the customer account code of the library owning the package.

"Project" is the project code associated with the AIP.

Files

The "Files" section of the report contains detailed information about each file that is included in the Dissemination Information Package (DIP). It can have up to five subsections: Archival Attributes, Message Digests, Events, Broken Links, and Warnings. The contents of the first four subsections are the same as described above for the Ingest Report. The contents of the Warnings subsection always includes a note indicating who requested the dissemination.

5. Withdrawal Reports

Withdrawal reports are created when a request to withdraw material from the FDA is processed. One Withdrawal Report is created for each package listed in the withdrawal request. The Withdrawal Report is emailed to the email address associated with the affiliate's account as given in the ACCOUNT attribute of the <AGREEMENT_INFO> element in the SIP.

5.1. XML Document

The Error Report is an XML document that uses the schema *daitssReport.xsd*. The document is contained within a <REPORT> element. The first XML element within <REPORT> is <WITHDRAWAL>.

```
<REPORT xmlns="http://www.fcla.edu/dls/md/daitss/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.fcla.edu/dls/md/daitss/
  http://www.fcla.edu/dls/md/daitss/dev/daitssReport.xsd">
  <WITHDRAWAL IEID="E20061211_AAAAAU"
  etc.
```

A sample XML Error Report is available in the [Digital Archive Information](#) section of the FDA website.

5.2. Formatted Withdrawal Report

If the default stylesheet is used, the Withdrawal Report has two main sections: a header ("Withdrawal") and a file list ("Files"). These are formatted as shown below.

Withdrawal

| | |
|------------------------|-----------------------------------------------------------------------------------------------|
| Package name | FDA0000200 |
| Int. Entity ID | E20061211_AAAAAU |
| Withdrawal time | 2006-12-11T18:45:57Z |
| Note | Withdrawal of E20061211_AAAAAU in account FDA requested by daitss operator <daitss@localhost> |

"Package name" is the name of the Submission Information Package (SIP) as submitted by the customer.

"Int. Entity ID" is the name of the "intellectual entity" (logical object) assigned by the FDA.

"Withdrawal time" is the time the withdrawal completed in date/time format in Greenwich Mean Time:
yyyy-mm-ddThh:mm:ssZ

"Note" always identifies the contact requesting the withdrawal. This can be used by the affiliate as a double-check that the withdrawal request was legitimate.

Files

F20061211_AAAAEQ

| | |
|--------------|------------------------------------------|
| Path | DSC04975_small.jpg |
| Size | 239339 |
| MD5 | 42ea013588a1ddbdc0dafb85dab12845 |
| SHA-1 | 73d972a507cb845c978ac4bcb3788b356b0e4002 |

In the Files section, each file included in the withdrawn package is listed. For each file, the Data File ID assigned by the FDA is given, followed by the filename and filepath (relative to the SIP/AIP directory), the size of the file in bytes, and any message digests assigned by the FDA.

6. Report handling

The email account to which reports are directed should be dedicated to emails from the FDA. This has two advantages: first, the volume of reports would flood a staff members' regular email account. Second, isolating FDA reports in a single mailbox makes it easier to run scripts to automatically process the reports. One script might separate the different kinds of reports and route them to different subdirectories for processing.

6.1. Ingest Report handling

It is recommended that the affiliate keep a copy of each SIP submitted to the FDA until an Ingest Report is received for that SIP.

FDA affiliates are responsible for keeping track of the materials archived in the FDA. It is recommended that the affiliate's local records store at a minimum the package name of the SIP, the date sent to the FDA, and enough bibliographic information to identify the contents of the SIP.

When an Ingest Report is received, the report can be processed automatically by script and/or program. The package name should be parsed from the report and looked up in the local database. The date that the SIP was ingested or some other flag that ingest was successful can be added to the local record.

It is also recommended that the affiliate store the FDA identifier, file name, and at least one checksum for each file ingested. That way, if the original content is later returned in a Dissemination Information Package (DIP), the affiliate can verify that the data files received are identical to the files that were originally sent.

Finally, the copy of the SIP stored at the local site can be deleted, if desired, when the Ingest Report is received.

6.2. Ingest Error Report handling

Ingest Error Reports received should probably be examined manually by staff to determine what to do about each one. With some experience the affiliate should be able to decide on processing procedures for each type of error. For example, an invalid

checksum may mean that the package was corrupted in transmission, and the routine recovery procedure might be to resubmit the original SIP. On the other hand, if the package was rejected for an invalid SIP descriptor, the error in the METS file must be identified and corrected before the package can be resubmitted.

FDA staff are happy to help in the resolution of ingest errors.

6.3. Dissemination Report handling

Because dissemination includes a re-ingest step, new files may have been created during the dissemination process. Ideally, a script should compare the ingest information from the dissemination report against the information for the package stored in the local database, and the local database updated with new file entries as appropriate.

The Dissemination Information Package (DIP) always includes the unaltered contents of the original SIP. The affiliate may want to compare the checksums of the files in the DIP against those of the original SIP (if these are stored locally, as recommended). This provides assurance that the original files have indeed been preserved with bit-wise integrity.

6.4. Withdrawal Report handling

FDA affiliates are responsible for keeping track of the materials archived in the FDA. It is recommended that the affiliate's local records store at a minimum the package name of the SIP, the date sent to the FDA, and enough bibliographic information to identify the contents of the SIP.

When a Withdrawal Report is received, the report can be processed automatically by script and/or program. The package name should be parsed from the report and looked up in the local database. At the institution's option, the entry for that package could be deleted from the local database. However, a preferred course of action would be to retain the entry, but note the date and time of withdrawal.

The File entries in the withdrawal report can be matched against file entries in the local database in order to check and/or flag each one as deleted.