

# Florida Digital Archive (FDA) Policy Guide

Version 2.4, August 2007

superseded versions:

version 2.3, December 2006

[version 2.2, August 2006](#)

[version 2.1, February 2006](#)

[version 2.0, January 2006](#)

Sections changed in this version:

Succession Plan (Removed mention of subaccounts)

This document covers mission, governance, division of responsibilities, archivable materials, rights, services, preservation strategies and other topics of interest to FDA users.

## **CONTENTS**

Mission.....	3
Background .....	3
Governance .....	3
FDA Affiliates .....	4
Designated Community .....	4
Terminology: SIP, AIP and DIP .....	4
Model of operation and division of responsibilities .....	5
Archivable materials .....	7
Rights in archived materials.....	7
Billing .....	7
Services .....	8
Preservation Strategies.....	10
Storage .....	11
Security .....	11
Continuity of Operations Plan .....	12
Succession plan .....	13
Related documentation.....	13

## **Mission**

The mission of the Florida Digital Archive (FDA) is to provide a cost-effective, long-term preservation repository for digital materials in support of teaching and learning, scholarship, and research in the state of Florida.

In support of this mission, the FDA guarantees that all files deposited by agreement with its affiliates remain available, unaltered, and readable from media. For those materials designated to receive full preservation treatment, the FDA will maintain a usable version using the best format migration tools available.

## **Background**

Planning for the FDA began in 2001 in response to the perceived need of the directors of the libraries of the public universities of Florida to ensure the permanent availability of digital library materials such as electronic dissertations. Development was expedited by the award in 2002 of a three-year grant from the Institute of Museum and Library Services (IMLS) which concluded in September, 2005.

In order to implement the FDA, FCLA staff designed and developed the DAITSS application (Dark Archive in the Sunshine State). When this application is complete it will be released as open source software under a license yet to be determined. The FDA began ingesting materials in production in the fall of 2005. As of August 2006, the dissemination function is still in development.

The technical design, procedures and policies of the FDA are based on [OAIS - Open Archival Information System Reference Model](#) (ISO 14721:2003) and on ongoing work to define and certify trusted digital repositories, including [Trusted Digital Repositories: Attributes and Responsibilities](#) (RLG May 2002) and the [RLG/NARA Audit Checklist for Certifying Digital Repositories](#) (RLG August 2005).

## **Governance**

The FDA is run by the Florida Center for Library Automation, a Type-1 center of the Florida Division of Colleges and Universities (DCU), responsible for providing information technology support to the libraries of the public universities of Florida. It is under the administrative control of the Director of FCLA.

The FCLA Advisory Board acts as the Advisory Board of the FDA, recommending policy decisions to the FCLA Director. The FCLA Advisory Board consists of the directors of the libraries of the eleven public universities, plus a representative from the

DCU, a representative from the Division of Community Colleges, and the State Librarian of Florida.

## **FDA Affiliates**

FDA affiliates are institutions who have signed Agreements with the FDA allowing them to use the services of the FDA. Institutions eligible to be affiliates are public university libraries in the DCU, and PALMM partners. A PALMM partner is an institution that has signed a formal partnership agreement with a library in the Florida state university system to participate in one or more PALMM projects. Non-library units within the public university system may submit materials for archiving only indirectly, by arranging for a Library to include these materials in its Agreement.

## **Designated Community**

An Open Archival Information System (OAIS) is defined as an archive that has accepted the responsibility to preserve information and make it available to a Designated Community. The Designated Community is defined as an identified group of potential Consumers who should be able to understand a particular set of information. The designation of one or more Designated Communities is important for a preservation repository, because the repository must assure that archived digital information is independently understandable to that group.

For the FDA, the Designated Community is the set of professional staff of the FDA affiliates. Staff members interact with the FDA and serve as proxies for the constituencies they serve in the academic and research communities. They must be able to render materials disseminated to them by the FDA and present these materials to users in understandable form. This may require them to write or acquire rendering software, for example METS-based page turners or media players, but it will not require extraordinary efforts, such as digital archeology or the acquisition of obsolete software or hardware.

## **Terminology: SIP, AIP and DIP**

The terms SIP, AIP and DIP appear frequently below.

A SIP, or submission information package, is defined in OAIS as an information package delivered to the repository for archiving. In the OAIS specification, the SIP may contain one digital file or many files. In the FDA implementation, a SIP must follow certain rules documented in the [FDA SIP Specification](#).

An AIP, or archival information package, is the contents of the SIP as ultimately represented in archival storage, after additions, deletions, and transformations are performed by the repository.

A DIP, or dissemination information package, is a set of content disseminated from the repository. In the OAIS specification, the DIP may contain content from one or more AIPs. In the FDA implementation, a DIP disseminates all or part of a single AIP.

## **Model of operation and division of responsibilities**

The FDA model of operation is based on the notion that responsibility for digital preservation is shared between the FDA and its affiliates. The FDA attempts to ensure that archived materials can be disseminated back to the depositor in renderable form; that is, that the materials can be displayed, played or used as appropriate for the format on software currently available at the time of dissemination. However, the FDA is not itself a system for discovery, public access or rendering. As a result, affiliates should note these limitations of the FDA:

- Information packages are disseminated from the FDA on request in batch form and may not be immediately usable. Disseminated materials may require the provision of appropriate rendering software and/or some transformation, such as the creation of derivative service copies.
- The FDA is responsible only for the materials it archives. If an affiliate chooses to archive both the master and service formats of a particular item, the FDA will attempt to ensure the future renderability of both files via forward migration and other techniques. If the affiliate archives only the master format, the FDA is not responsible for making service copies.

The following are the responsibility of an FDA affiliate institution:

- Negotiate an Agreement for the use of the FDA. A [template](#) for the *FCLA – Library Agreement* is available on the [Florida Digital Archive website](#). Agreements must be countersigned by representatives of the affiliate and of FCLA.
- Maintain current Appendices to the Agreement, specifying which materials will be archived, which file formats get bit level or full preservation, and which staff are authorized to perform certain actions.
- Select the content to submit for archiving, and make sure that adequate descriptive metadata is maintained locally.

- Ensure the affiliate has the right to deposit the content, and to give the FDA all permissions it needs to store, back up, and carry out preservation strategies on the material. Affiliate institutions assume liability for any breach of intellectual property rights occasioned by the deposit, copying and/or preservation treatment of deposited objects.
- Submit the content to the FDA in the format required by the [FDA SIP Specification](#). At this time submission via FTP is required, although other forms of submission may be supported in the future. The SIP specification requires a descriptor conforming to the [DAITSS METS SIP Profile](#).
- Maintain records of what is archived with the FDA. The affiliates records should include at a minimum the entityID of the SIP and a link to any locally maintained metadata. Recording the filenames and checksums of any submitted content files and the local identifier and/or title included in the METS descriptor is also recommended. Administrative information may also be useful, such as the date the package was sent to the FDA, the date confirmation of ingest was received via an Ingest report, and additional checksums contained in the Ingest report.
- Use the information in Ingest Reports and Error Reports to verify the disposition and condition of materials submitted to the FDA. See [Ingest Reports and Error Reports](#) on the FDA website for more information about these reports.
- If for any reason the archiving of any materials is no longer required, make a request to the FDA to withdraw that content.
- Request dissemination of content when needed back for any reason, for example, in order to make new derivatives from a master.

The following are the responsibility of the FDA:

- Implement bit level or full preservation for materials submitted by affiliates as specified in the Agreement with the affiliate.
- Restrict authorization to deposit materials, withdraw materials, disseminate materials, and request reports to the individuals specified in the Agreement with the affiliate.
- Provide detailed Ingest or Error information for every SIP received.
- Preserve original files exactly as submitted, with demonstrated integrity, viability and authenticity.
- When full preservation is requested for a file in a supported format, use appropriate preservation strategies to maintain a renderable version of the file at all times.

- Provide dissemination information packages (DIPs) on request. The DIP may contain original files as submitted and/or modified renderable versions.
- Provide appropriate reports to affiliates for management purposes.
- Achieve and maintain certification as a Trusted Digital Repository.

## **Archivable materials**

An FDA affiliate may archive any content in any format, as long as the affiliate has appropriate rights (see *Intellectual Property Rights* below). Content to be archived is specified in Appendix A of the Agreement between FCLA and the affiliate. The affiliate is considered the "owner" of the content. Content must be identified by project code and file format; that is, every SIP must have a project code contained in it, which is associated with a list of file formats and the level of preservation treatment desired for that format by the owner.

For more information about specifying content to be archived, see [Setup Procedures for New FDA Affiliates](#) on the FDA website.

Appendix A can be modified by the affiliate at any time by providing FCLA with a dated replacement in writing. The revised Appendix must be signed by the primary contact for the affiliate, as listed in the Agreement. Modifications will be implemented by FCLA within 10 working days.

## **Rights in archived materials**

The affiliate retains all ownership and management rights in materials submitted by the affiliate.

The affiliate is responsible for compliance with all applicable copyright laws and other laws applicable to deposited materials, and must have the authority to grant FCLA non-exclusive rights to copy, display and create derivative versions of deposited files.

If FCLA receives notification in writing that any digital file is held in the FDA in violation of applicable law, the FDA will disseminate a DIP to the owning affiliate and withdraw the entire AIP containing the file. If this is challenged and resolved in favor of the affiliate, FCLA will re-ingest the material at no charge.

## **Billing**

FCLA does not charge for the use of the FDA. It is possible that at some time in the future, billing will be instituted with the advice and consent of the FDA Advisory Board.

If billing is instituted, written notice of no less than 180 days will be provided to affiliates. No charges will be incurred by affiliates for FDA services provided up to the time that billing is implemented. If an affiliate wishes to withdraw any of its materials from the FDA before billing is implemented it may do so at no charge by following standard procedures for withdrawal requests.

If billing is implemented, the billing algorithm and rates will be posted on the FDA website along with terms and conditions of payment established by FCLA. Affiliates who allow their content to remain in the archive commit to meeting their financial obligations to the FDA.

## **Services**

Standard services of the FDA are listed below. Special services, such as the automatic creation of METS descriptors for a set of content files, may be available upon request. Contact FDA staff for information.

### **1. Ingest of SIPs.**

SIPs submitted according to FDA instructions will be ingested into the archive. For every SIP submitted, the affiliate will receive an Ingest Report for successful ingest or an Error Report if the SIP is rejected. Ingested SIPs are transformed into Archival Information Packages (AIPs) for storage.

As part of the ingest process, content files are checked for viruses and for conformance to format specifications (validity). Finding a virus in any file will cause the entire SIP to be rejected with an Error Report. Finding a validity error will not normally cause the SIP to be rejected, but it will be recorded as an anomaly in the metadata stored for the file and reported to the affiliate in the Ingest Report.

Note that if changes are required to an information package already submitted to and archived by the FDA, the package cannot be changed by resubmitting the SIP. The FDA has no ability to update archived content. To correct a package without adding a duplicate, the existing AIP must first be withdrawn before submitting a SIP containing the corrected content. (See item 5. Withdrawal, below).

### **2. Secure storage and management of AIPs.**

This includes maintaining onsite and offsite copies, periodic fixity checking and periodic verification that the AIP is complete. It also includes periodic copying to new storage media (media refreshment).



Secure storage and management is referred to as “bit-level preservation” because the integrity of the original AIP is preserved for later dissemination.

### 3. Full preservation for supported formats.

Full preservation treatment will be performed for files in supported formats if full preservation is specified for the file by the affiliate in its Agreement. Supported formats are computer file formats with Action Plans listed on the FDA website.

Full preservation includes bit-level preservation of the originally submitted files, as well as services intended to ensure that the information content of the files will remain usable into the indefinite future. These services vary according to the file type but may include the creation of normalized forms of the file and/or the reformatting of obsolete formats to reasonably comparable successor formats. See Preservation Strategies below for more information.

The FDA does not guarantee that normalized or migrated versions of any file will be identical in appearance or behavior to the original file. Note also that if an AIP is comprised of content files in both supported and unsupported formats, there is no guarantee that the intellectual entity will remain usable as intended.

### 4. Dissemination.

Archived content will be disseminated by FTP upon the request of an authorized agent of the affiliate. (Note that dissemination does not remove content -- the AIP remains in the FDA unless it is explicitly withdrawn.) The dissemination package (DIP) will be placed in the affiliate's FDA FTP directory, and the affiliate will receive a Dissemination Report by email.

The DIP always includes the original version of all files deposited in the original SIP. If a preferred version of any file has been created, the DIP will also include the "last, best" version of the file. The DIP Descriptor is a METS file with a structural map section (<structMap>) describing the original representation and, if applicable, a second structural map section describing the current representation.

Dissemination is carried out in three steps. First, a copy of the AIP is exported to a special re-ingest directory, where it is treated as a SIP. Second, the AIP/SIP is re-ingested, at which time file formats are re-identified using the most current identification routines, and files are normalized and/or migrated using the most current file processing routines. Third, the new AIP is written as a DIP to the appropriate output directory. This process ensures that the disseminated package is always as complete and up-to-date as possible.

### 5. Withdrawal.

An AIP will be withdrawn from the FDA upon the request of an authorized agent of the affiliate. Files belonging to the withdrawn AIP are deleted entirely from storage, but the FDA retains a permanent record that the intellectual entity was ingested and withdrawn. The affiliate will receive a Withdrawal Report by email.

A common reason for withdrawal is to correct information in a previously submitted package. The existing AIP must be withdrawn, and a SIP containing the correct information submitted for Ingest.

AIPs may also be withdrawn on the initiative of the FDA, if it receives information that content has been archived in violation of copyright.

## 6. Reporting.

The FDA will provide periodic statistical reports to affiliates about their own use of the FDA. Ad hoc reporting will also be available upon request.

## Preservation Strategies

Preservation strategies supported by the FDA are based on format transformation, that is, changing files formats to delay or accommodate format obsolescence. The FDA performs three kinds of transformations:

- **Localization.** If a file contains one or more links to other files (for example, an XML file containing a link to a schema definition), a localized version will be created if possible. The linked-to file will be located and downloaded into the AIP if it is not already contained in the SIP. A localized version of the original, linked-from file will be created in which the link is replaced by a reference to the name of the linked-to file in the AIP. The purpose of localization is to ensure that AIPs are complete and that references between files in the AIP can always be resolved.
- **Normalization.** If a file is in a format considered to be less than optimal for digital preservation a version of the file may be created in a more preservation-worthy format. In general, preferred formats are non-proprietary, well documented, and well understood by FDA staff. Normalized versions may not be equivalent to originals in appearance or functionality. For example, a PDF file will be normalized into a set of page-image TIFFs. In this case the appearance of the content is retained but functionality such as actionable hyperlinks is lost.
- **Migration.** If a file is in a format considered at risk of obsolescence, a version will be created in a format considered to be a reasonable successor to the original format. All effort will be made to retain the appearance and behaviors of the original version, although this can not always be guaranteed. The successor

format may be a higher version of the original format (for example, PDF 1.4 might be migrated to PDF 1.6) or it may be another format altogether.

The preservation strategies that will be implemented for any file format are documented in the Action Plan for the file format, available on the FDA website. Action plans are reviewed periodically and revised when appropriate.

All preservation strategies are applied at the time a SIP is ingested, as part of ingest processing. They are only applied to files receiving full preservation treatment, as specified by the affiliate in the *FCLA – Library Agreement* Appendix A. Localized, normalized and migrated versions of files contained in the SIP become part of the AIP.

If a file is designated as receiving full preservation treatment in Appendix A, but there is not yet an Action Plan for the file format, bit-level preservation will be carried out for the file until the time when full preservation becomes available. At that time, the AIP containing the file will be disseminated and re-ingested, causing the full preservation treatment to be applied.

## **Storage**

All AIPs are stored on tapes in an IBM robotic tape library managed by Tivoli Storage Manager.

For every file in the AIP, three master copies are written. Two copies are stored at the UF Computing & Network Services facility in Gainesville (CNS) and one copy is stored at the Northwest Regional Data Center in Tallahassee (NWRDC).

The three master copies are treated as a single file by DAITSS, the repository software application underlying the FDA. This means that when any action is performed on a file, it must be successfully performed on all three copies to be considered complete. For example, a fixity check involves calculating a message digest over the bits of a file and comparing this to a previously stored message digest. For a fixity check to be complete, message digests must be calculated for each of the three master copies of the file and verified to match the stored message digest.

## **Security**

Data security is ensured by a combination of physical security and cybersecurity.

The UF Computer Network Services (CNS) in Gainesville and the North West Regional Data Center (NWRDC) in Tallahassee are responsible for the physical security of computer tapes containing archived data, computer servers running the archive application, and other computer hardware necessary for the operation of the FDA.

CNS is a secure data center at the University of Florida. The central machine room and all of the core network fiber huts are secured with an electronic lock system based on the Lenel OnGuard software and Lenel controllers. All doors that open on to public space are configured to fail to a secure state. Alarm conditions are investigated by operations staff or reported to the University Police Department at operations' discretion.

Proximity cards and fobs are used for access. Access rights are granted according to work requirements and staff interaction needs. Most technical staff have 24/7 access to their work areas (including the machine room and fiber huts in most cases) and work hours access to other areas where there are staff members. For access during non-work hours, PIN codes are required in addition to the card or fob.

Air handlers and leak detectors are continuously monitored by operations staff. Backup power is provided by a single UPS for all computer equipment and a diesel generator which can power the UPS and the air conditioning systems for 24 hours without refueling.

NWRDC was designed and engineered to be a state-of-the-art data center that could guarantee customers' security, accessibility and connectivity. NWRDC employs an advanced security system and is monitored 24 hours a day, 7 days a week. Exteriors, doorways and hallways are under continuous recorded video surveillance. Multiple security points are in place throughout the entire building. Air temperature and humidity levels are continuously monitored and controlled. Redundant cooling systems are implemented. To protect equipment and data, NWRDC is equipped with redundant power and two generators that can run at full capacity for weeks with emergency refueling as needed, a smoke detection and fire suppression system, and lightning protection.

Cybersecurity refers to computer and network security. FCLA has a designated security administrator who enforces security procedures and maintains security tools that exceed the best practice recommendations of the University of Florida. The FDA is a "dark archive" with no online public access to stored data. Network access to archived data and to the FDA server computers is limited to a small number of FCLA staff working from known IP addresses. Unauthorized attempts to gain access to data and server computers are logged and reviewed daily for unusual security threats. Security patches to the operating systems are checked for daily and installed within 24 hours of release.

## **Continuity of Operations Plan**

A Continuity of Operations Plan, or COOP, is a tested plan to ensure that an essential service can continue to function across a wide range of potential emergencies, including localized acts of nature, accidents, technical failures, and physical or technological attacks.

The primary computer server for the FDA resides at CNS in Gainesville, FL (see Security above). A backup server capable of running the DAITSS software is in place at the NWRDC in Tallahassee. According to the COOP for the FDA, a hardware failure at the primary location will result in a switch over to the backup server within hours of the failure. The complete COOP is in development and will be documented in this section when available.

## **Succession plan**

In the event that the Florida Digital Archive ceases production operations for any reason, FDA affiliates will have two options. In option 1, content archived for the affiliate will be returned to the affiliate in standard FDA Dissemination Information Packages (DIPs). In option 2, the content will be sent to another preservation repository of the affiliate's choice in a DIP exchange format, which may not be the same as the standard FDA DIP. (A standard DIP exchange format has not yet been established within the international preservation community, but the FDA is working with other repositories to ensure interoperability of content.)

An affiliate will be able to specify that option 1 or option 2 applies to all AIPs for that account. Alternatively, if the affiliate can identify exclusive subsets of content by project coding, the affiliate can request different options for different subsets.

## **Related documentation**

Related documentation appears in the [Digital Archive Information](#) and [Software and Documentation](#) sections of the [Florida Digital Archive website](#).