Ingest and Dissemination with DAITSS

Presented by Randy Fischer, Programmer,
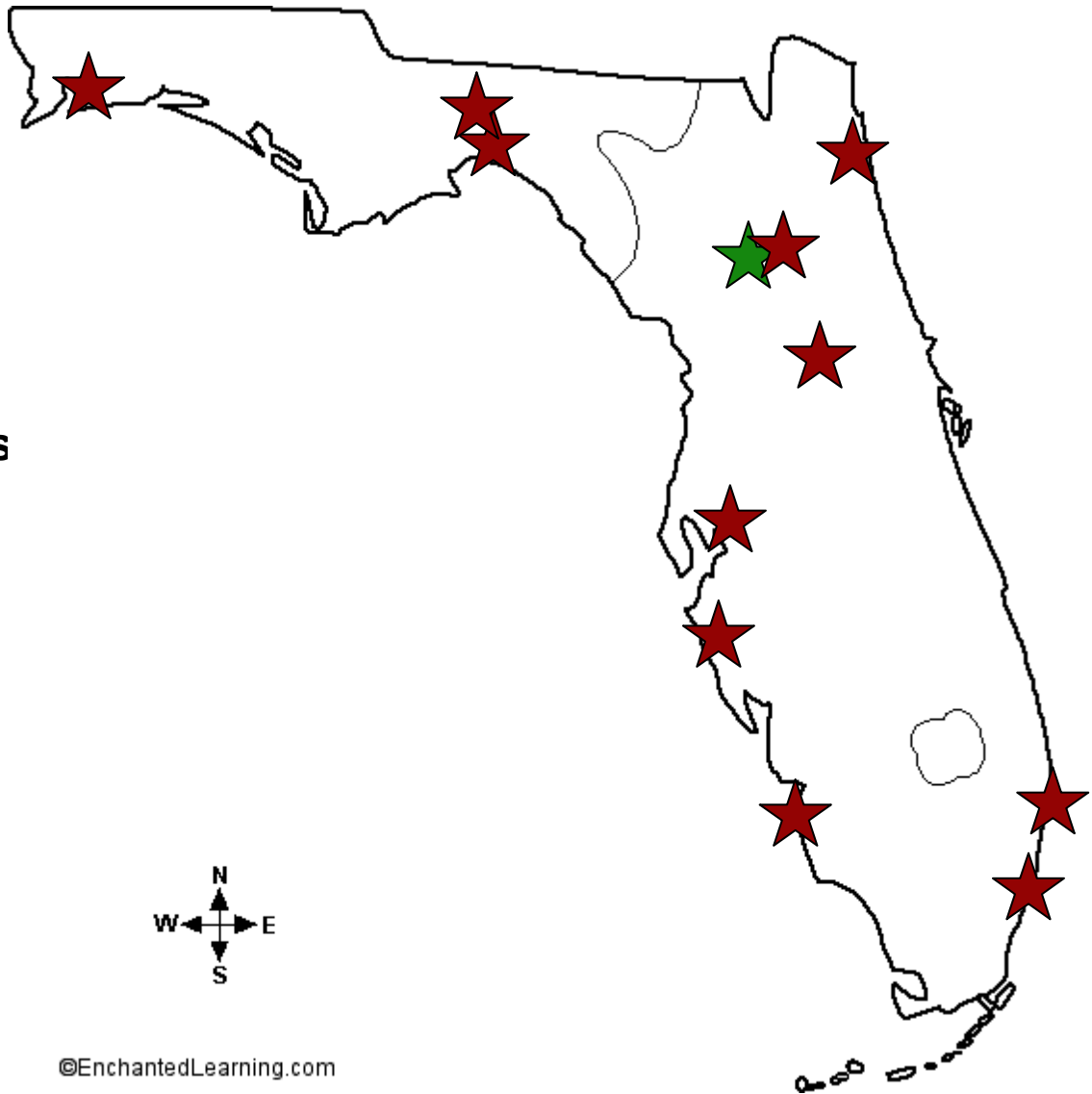Florida Center for Library Automation, University of Florida

DigCCurr2007
Raleigh, Durham

Florida Digital Archive

- What's the FDA?
    - Preservation Repository
    - Operated by the Florida Center for Library Automation
    - Serves the State Universities in Florida

    - Dark Archive: no online presentation
    - Designed solely as a preservation repository
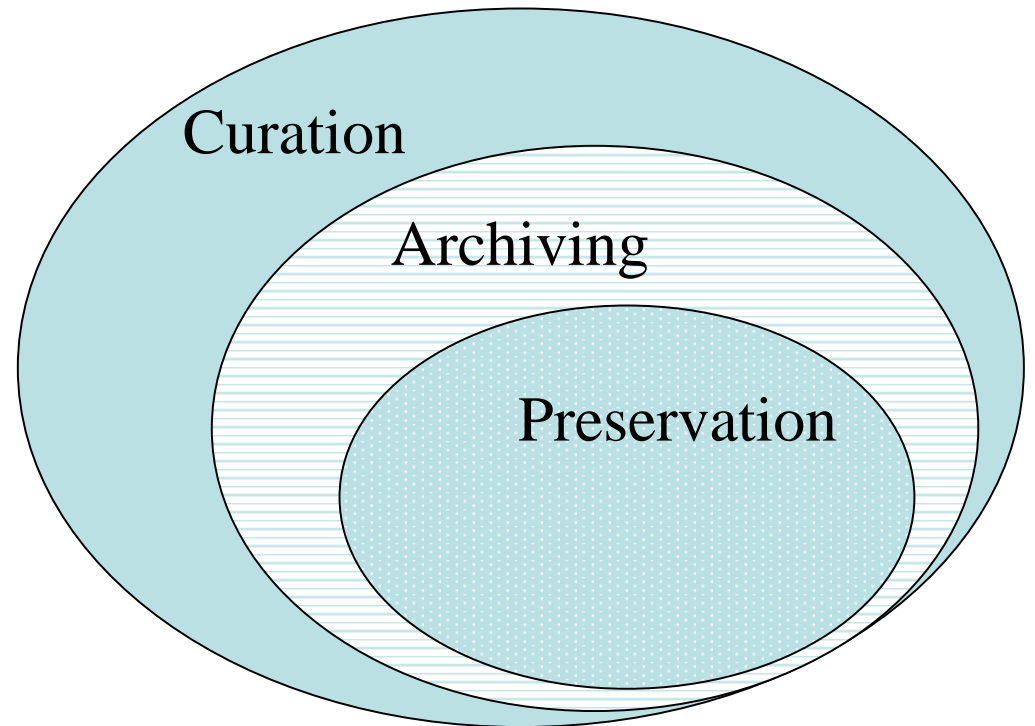
Florida Digital Archive

DAITSS

- What's DAITSS?

  - The Dark Archive In The Sunshine State

  - The software developed for the FDA

  - Implements the OAIS functional reference model

  - Implements the preservation strategies of Format Migration and Normalization
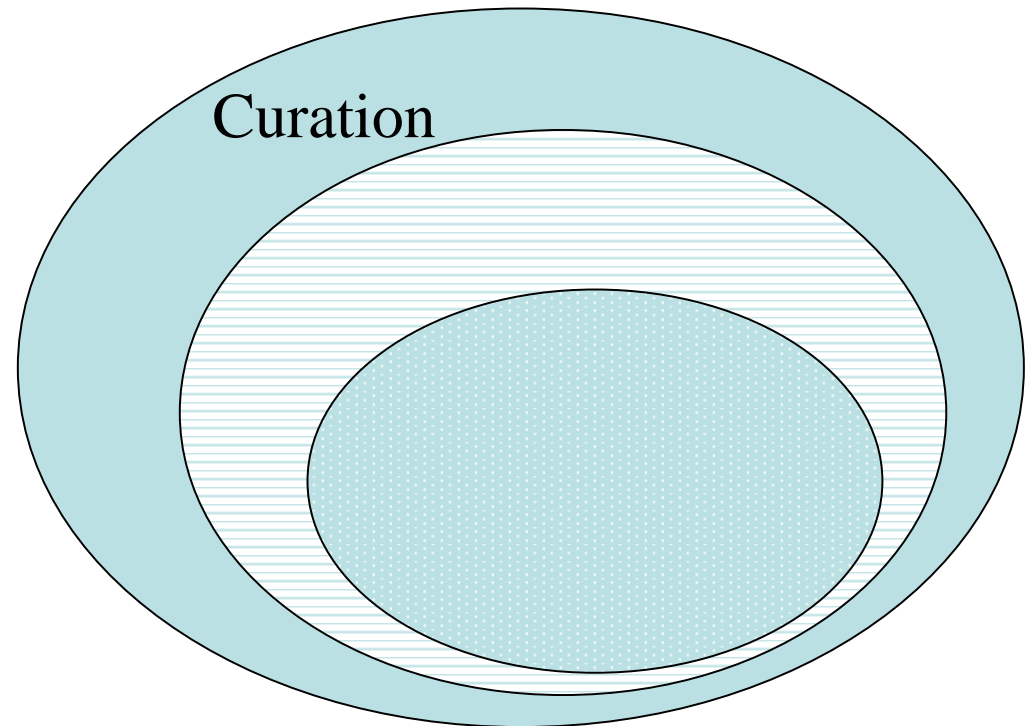
Roles & Responsibities

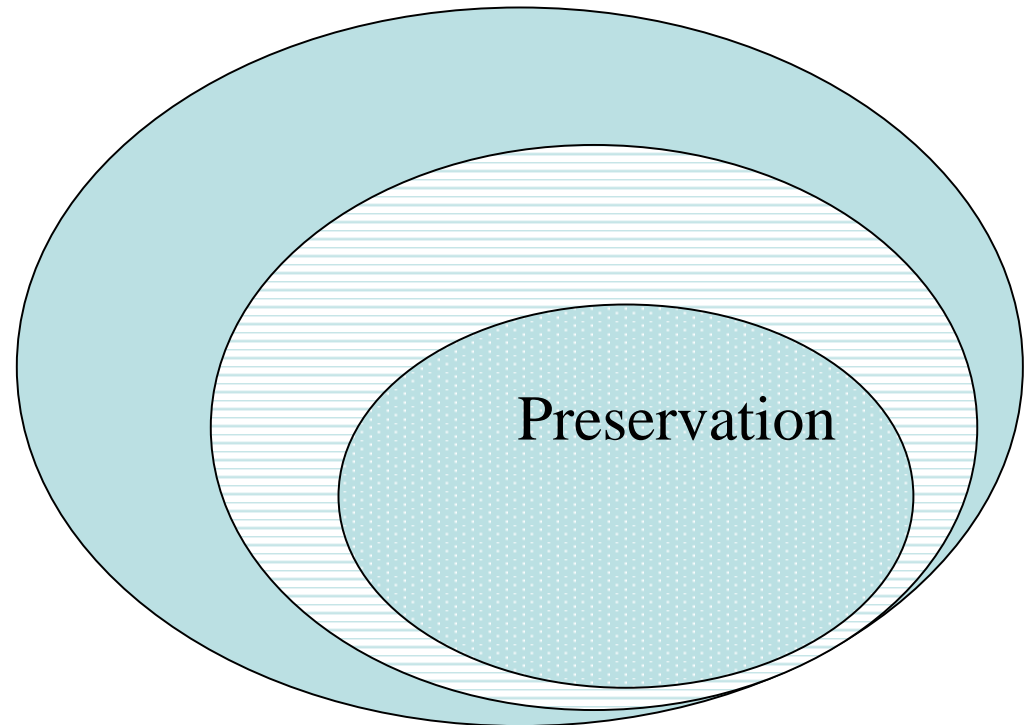- Curation
- Archiving
- Preservation

Responsibility of Library Affiliates

The activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use. For dynamic datasets this may mean continuous enrichment or updating to keep it fit for purpose. Higher levels of curation will also involve maintaining links with annotation and with other published materials.
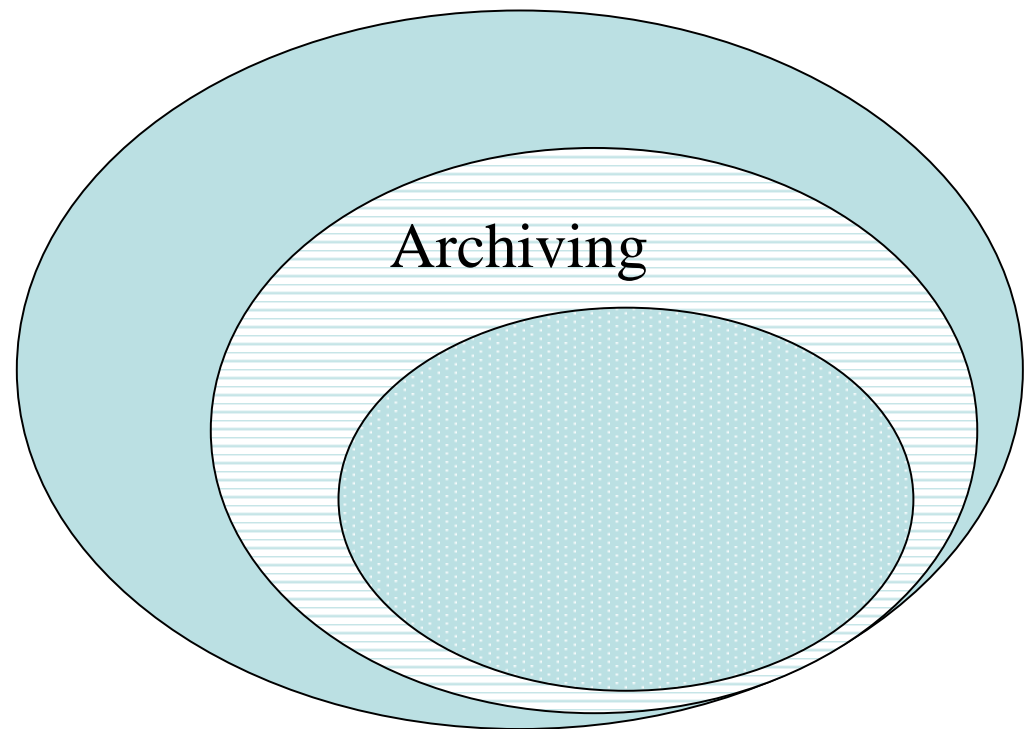
Curation

Responsibility of the FDA

- An activity within archiving in which specific items of data are maintained over time so that they can still be accessed and understood through changes in technology

- Preservation strategies e.g. migration, emulation, normalization

Joint Responsibility

- A curation activity which ensures that data is properly selected, stored, can be accessed and that its logical and physical integrity is maintained over time, including security and authenticity.

- Joint Responsibilities of Library Affiliates and the FDA

- FDA manages storage

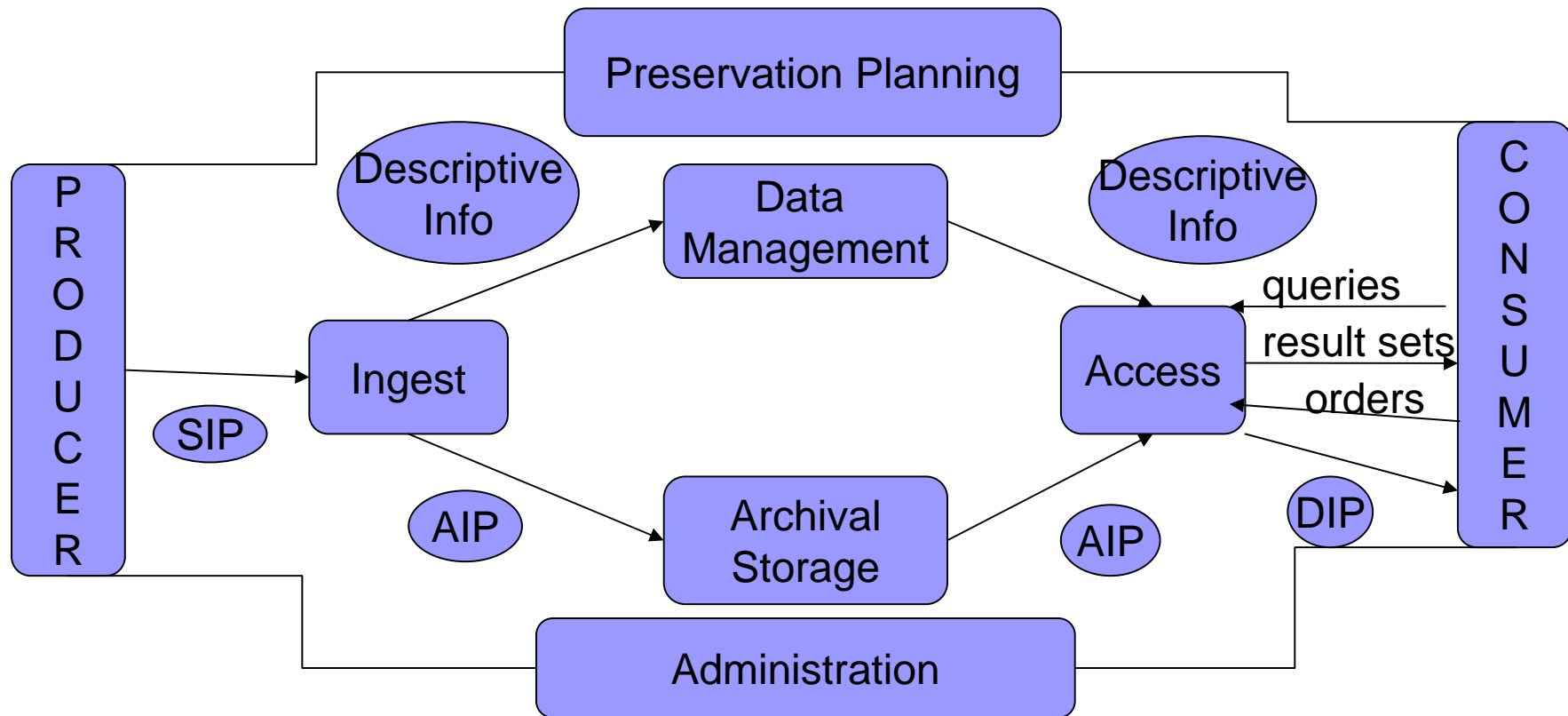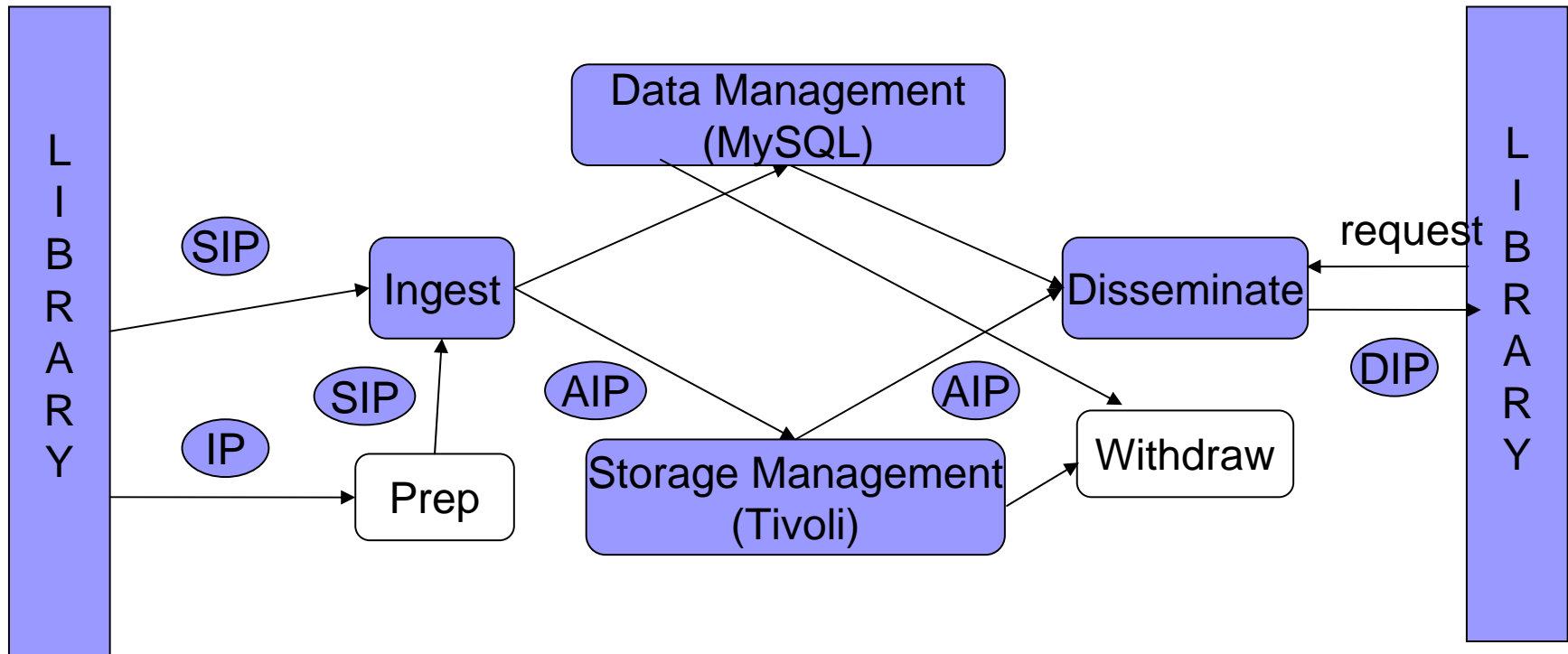- Affiliates select

Archiving

OAIS

- OAIS is a best practice reference model for long term archiving and preservation

- ISO standard

- Originally developed by NASA

- Everybody uses it (except NASA)

OAIS Functional Model

DAITSS Architecture

Ingest Service

- ## The SIP

  - Must contain one or more data files, and one SIP Descriptor

```
UF009643/

  UF009643.xml

  thesis.pdf
```

Ingest: Validate the SIP

- Validate the Package Directory

- Validate the XML Descriptor

- Administrative Metadata

  – Agreement Information

  – Preservation Policies (bit, full, none)

- Technical Metadata

  – Submitted message digest

  – File size

Ingest: Processing the Package

- Check for viruses

- Identify format, validate & record anomalies

- Extract technical metadata

- Identify & record external references

- Create normalized & migrated versions

Ingest: AIP Processing

- Assemble the files of the AIP

- Create a localized AIP descriptor (XML file)

- Record events & relationships

- Write three copies to storage

- Update the FDA MySQL database

- Send Affiliate Library a report

Dissemination

- Affiliate Requests a Package

- Package restored from tape

- Restored package is enqueued for re-ingest

- Placed into per-affiliate FTP directory

- A report is sent to the affiliate contact

Supported formats

- Bit-level preservation – anything goes

- Full presentation – supported formats

    - TIFF, JP2000

    - WAVE

    - PDF

    - Plain ASCII, SGML, XML

- None – nothing goes

Format Specialist

A Picture of Carol Chou

should go here

Home

Downloads

Documentation

About

Contact

Welcome to DAITSS

News

04.12.2007 -- DAITSS 1.2.1 now available

The DAITSS development team is pleased to announce that DAITSS v.1.2.1 is available for download. This release includes a source and binary distribution, and marks the first public release of the DAITSS digital preservation repository software.

Overview

The DAITSS (Dark Archive in the Sunshine State) digital preservation repository system preserves digital content for the very long term.

DAITSS is a "dark archive"; there is no real time access of content. DAITSS supports use by multiple customers, whether these are different organizations or different units within an organization. DAITSS allows great flexibility in terms of what is archived and how archived content is preserved. It implements forward format migration and normalization as preservation strategies.

DAITSS application was developed by the ⇒ Florida Center for Library Automation.

DAITSS is released as open source software under the ⇒ GNU GPL license.

We welcome you to the DAITSS community!

To learn more:

- Read the DAITSS Overview
- Download and install the software
- Browse the user and technical documentation
- Join the DAITSS mailing list
- Contribute to the DAITSS project

Technical requirements

DAITSS requires Java 5 or later, and is meant to run on a Unix or Linux operating system. The installation script requires the bash shell and makes use of the standard Unix utilities tar, gz, sed and awk. This DAITSS distribution is installed in a minimal configuration, and requires the following software:

- MySQL Version 4.x or later, with InnoDB data table support.
- The Sun Java Virtual Machine (JVM), version 5 or later. The JVM will be present if you have installed Sun's JRE, J2SE or J2EE systems.
- A working Mail Transfer Agent (MTA) such as Sendmail or Postfix