DAITSS WEB SERVICES

Priscilla Caplan

Florida Virtual Campus

The future's so bright….

History

- DAITSS in use by Florida Digital Archive since 2005
- Written by the Florida Center for Library Automation (FCLA) with support from the IMLS
- Completely rearchitected and rewritten in 2009-2011
- OAIS-based preservation repository system
- Designed for consortial / multi-user environment
  - 11 public universities of Florida
- As of 6/1/2012:
  - 308,554 AIPs, 40.7 million files, 111 TB (one copy)
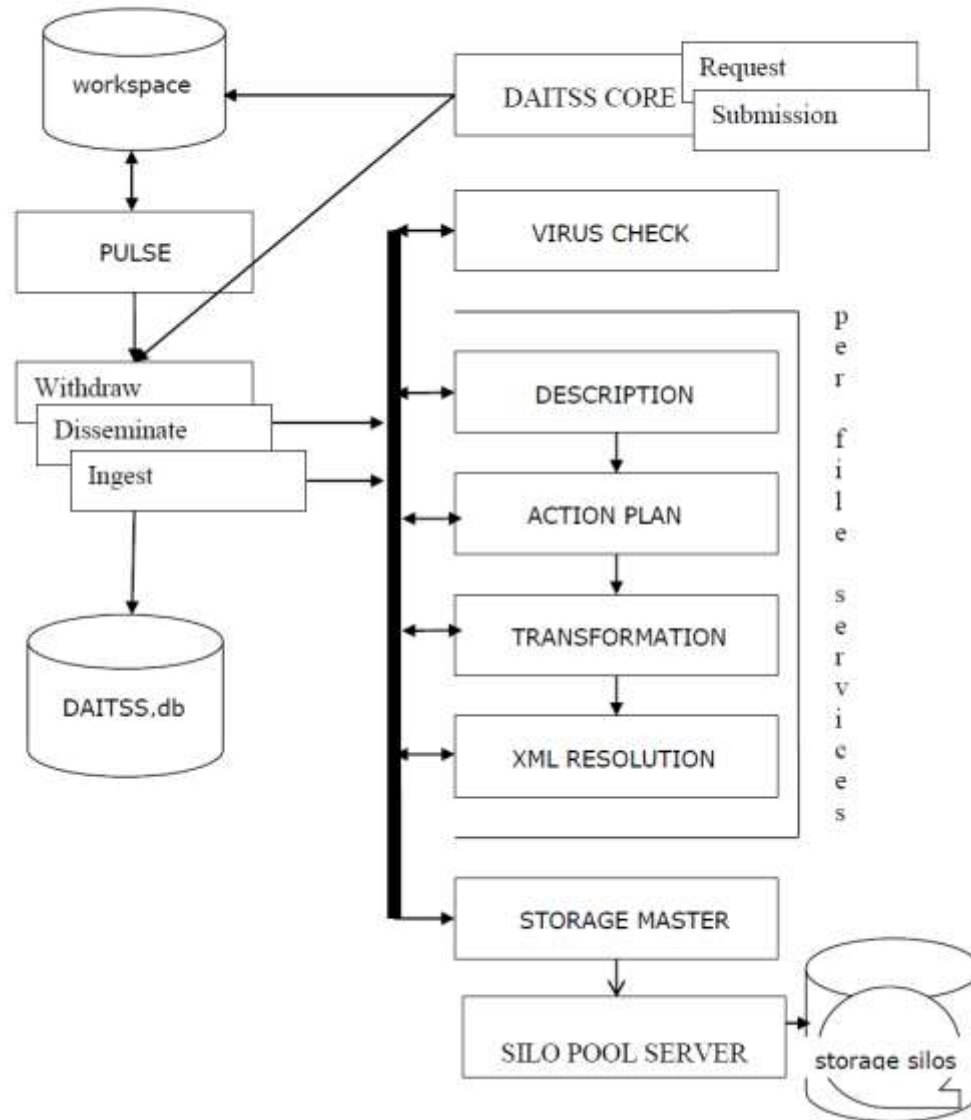
The future's so bright….

Functions supported

- Submit = accept or reject a package (SIP) for Ingest
- Ingest = transform a SIP into a stored AIP
- Refresh = re-run per-file processing to update an AIP
- Disseminate = export a refreshed copy of an AIP
- Peek = view an AIP as it is in storage
- Withdraw = remove an AIP from storage (but retain provenance in the database)
- Operator interface for tracking, interventions, storage management
- User interface for submission, requests, tracking

The future's so bright….

DAITSS Web Services

The future's so bright….

Description service

- Takes a file as input
- Uses DROID, JHOVE, PRONOM, other to
  - Identify file type
  - Validate according to file type specifications
  - Record inhibitors and anomalies
  - Supply technical metadata appropriate to file type
- Returns an XML file of description information in PREMIS format

The future's so bright….

Action Plan Service

- Takes a PREMIS document describing a file as input
- Extracts file format info from the PREMIS document
- Uses the file format info to look up the appropriate Action Plan for the format
- If a transformation is required:
  - Extracts a transformation identifier from the Action Plan
  - Returns status 303 with location field set to the URL of the Transformation Service
  - Returns the transformation identifier which will be used by the Transformation Service to create a derivative version of the file

The future's so bright….

Transformation Service

- Takes as input a request including a transformation identifier
- Extracts the transformation identifier from the request
- Looks up the transformation instructions in a configuration file
- Performs the format transformation
- Caches the derivative file
- Creates a PREMIS description of the Transformation event and agent
- Returns the PREMIS description and a link to the derivative file

The future's so bright….

XML Resolution Service

- Downloads schema, DTDs and stylesheets referenced by XML files
- Must be called at least 3 times:
  - 1st call says "get ready to process a bunch of xml files"
  - 2 – nth calls pass XML files to resolve
  - Last call picks up a tar file of all downloaded files

The future's so bright….

Using DAITSS Services

- All available in GitHub

  - https://github.com/daitss/

- All services can be invoked as RESTful services

- Can be invoked by CURL

- Can be used standalone or can be embedded in other services or applications

- Description Service and Action Plan service available on public servers

- Description Service included in public PREMIS-in-METS service (PIM)

The future's so bright….

Description Service
identify, validate and extract

Description Service performs format identification, validation and characterization on a given file. The result is expressed in PREMIS schema containing a PREMIS object for the identification and characterization result, a PREMIS event describing the validation status with applicable anomaly, and a PREMIS agent identifying the software agent that generates PREMIS output, in this case, the description service.

Submit | Information

Submit : file   uri

Submit a file to the description service. The result of format description may be saved locally using the browser "Save Page As" function.

File: [Choose File] No file chosen   [Submit]

Alternatively, users can use HTTP clients such as curl to submit local files directly to the description service via HTTP protocol. For example, to use curl to upload a local file named 00001.pdf to the description service and save the result to premis.xml,

curl -F "document=@00001.pdf" -F "extension=pdf" http://description.fcla.edu/description > premis.xml

The file upload is limit to 50MB.

DAITSS Description Service

The future's so bright….

The future's so bright….

```xml
<premis xmlns="info:lc/xmlns/premis-v2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" version="2.0">
  <!-- premis file object -->
  <object xsi:type="file">
    <objectIdentifier>
      <objectIdentifierType>URI</objectIdentifierType>
      <objectIdentifierValue>/tmp/object20120720-12140-zxkrpo-0.jpg</objectIdentifierValue>
    </objectIdentifier>
    <objectCharacteristics>
      <compositionLevel>0</compositionLevel>
      <fixity>
        <messageDigestAlgorithm>MD5</messageDigestAlgorithm>
        <messageDigest>2630c6bd816041e4e10668d74e7bb19b</messageDigest>
        <messageDigestOriginator>Archive</messageDigestOriginator>
      </fixity>
      <fixity>
        <messageDigestAlgorithm>SHA-1</messageDigestAlgorithm>
        <messageDigest>e53539cb3022a1b4ec9e3ef08caf0cdad971229e</messageDigest>
        <messageDigestOriginator>Archive</messageDigestOriginator>
      </fixity>
      <size>103127</size>
      <format>
        <formatDesignation>
          <formatName>JPEG File Interchange Format</formatName>
          <formatVersion>1.01</formatVersion>
        </formatDesignation>
        <formatRegistry>
          <formatRegistryName>http://www.nationalarchives.gov.uk/pronom</formatRegistryName>
          <formatRegistryKey>fmt/43</formatRegistryKey>
        </formatRegistry>
      </format>
      <format>
        <formatDesignation>
          <formatName>JFIF</formatName>
        </formatDesignation>
        <formatNote>Alternate Format</formatNote>
      </format>
      <objectCharacteristicsExtension>
        <mix:mix xmlns:mix="http://www.loc.gov/mix/v20" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/mix/v20
        http://www.loc.gov/standards/mix/mix20/mix20.xsd">
          <mix:BasicDigitalObjectInformation>
            <mix:ObjectIdentifier>
              <mix:objectIdentifierType>JHOVE</mix:objectIdentifierType>
            </mix:ObjectIdentifier>
            <mix:byteOrder>big endian</mix:byteOrder>
            <mix:Compression>
              <mix:compressionScheme>JPEG</mix:compressionScheme>
            </mix:Compression>
          </mix:BasicDigitalObjectInformation>
          <mix:BasicImageInformation>
```


The future's so bright….

pimtools
v1.0.1

| about | validation | conversion | describe | resources |

PIM

PREMIS:METS
conversion & validation tools

FCLA

PREMIS in METS Toolbox

The PREMIS in METS Toolbox is a set of open-source tools developed to support the implementation of PREMIS in the METS container format. The tools were created by Florida Center for Library Automation for the Library of Congress in 2009.

Validate PREMIS in METS document. Given a PREMIS or PREMIS in METS document via an input method, return a list of errors or a confirmation message if the document conforms. The document is validated against any applicable schema and if applicable the PREMIS in METS best practice.

Convert between PREMIS & PREMIS in METS. Given a PREMIS document generate a METS document containing the PREMIS elements in multiple or single METS metadata sections. Given a METS document containing PREMIS elements in metadata sections generate a stand-alone PREMIS document

Describe a file with PREMIS using the DAITSS 2 Description service. The object identifier type and value are required. The file may also link to an intellectual entity.

flaz@ufl.edu   Valid XHTML | CSS

The future's so bright….

pimtools

v1.0.1

PREMIS:METS

conversion & validation tools

FCLA

Describe: uri | file upload

Generate PREMIS for an uploaded file.

File: [Choose File] Dog DISH cr...7-18-12.jpg

Object Identifier (required):

Type: [local]

Value: [1]

Linking Intellectual Entity Identifier (optional):

Type: [ ]

Value: [ ]

[Describe]

```xml
http://www.loc.gov/standards/premis/premis.xsd">
 <!-- premis file object -->
▼<object xsi:type="file">
  ▼<objectIdentifier>
    <objectIdentifierType>local</objectIdentifierType>
    <objectIdentifierValue>1</objectIdentifierValue>
  </objectIdentifier>
  ▼<objectCharacteristics>
    <compositionLevel>0</compositionLevel>
    ▼<fixity>
      <messageDigestAlgorithm>MD5</messageDigestAlgorithm>
      <messageDigest>2630c6bd816041e4e10668d74e7bb19b</messageDigest>
      <messageDigestOriginator>Archive</messageDigestOriginator>
    </fixity>
    ▼<fixity>
      <messageDigestAlgorithm>SHA-1</messageDigestAlgorithm>
      <messageDigest>e53539cb3022a1b4ec9e3ef08caf0cdad971229e</messageDigest>
      <messageDigestOriginator>Archive</messageDigestOriginator>
    </fixity>
    <size>103127</size>
    ▼<format>
      ▼<formatDesignation>
        <formatName>JPEG File Interchange Format</formatName>
        <formatVersion>1.01</formatVersion>
      </formatDesignation>
      ▼<formatRegistry>
        <formatRegistryName>http://www.nationalarchives.gov.uk/pronom</formatRegistryName>
        <formatRegistryKey>fmt/43</formatRegistryKey>
      </formatRegistry>
    </format>
    ▼<format>
      ▼<formatDesignation>
        <formatName>JFIF</formatName>
      </formatDesignation>
      <formatNote>Alternate Format</formatNote>
    </format>
    ▼<objectCharacteristicsExtension>
      ▼<mix:mix xmlns:mix="http://www.loc.gov/mix/v20" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/mix/v20
        http://www.loc.gov/standards/mix/mix20/mix20.xsd">
        ▼<mix:BasicDigitalObjectInformation>
          ▼<mix:ObjectIdentifier>
            <mix:objectIdentifierType>JHOVE</mix:objectIdentifierType>
          </mix:ObjectIdentifier>
          <mix:byteOrder>big endian</mix:byteOrder>
          ▼<mix:Compression>
            <mix:compressionScheme>JPEG</mix:compressionScheme>
          </mix:Compression>
        </mix:BasicDigitalObjectInformation>
```

The future's so bright….

pimtools
v1.0.1

PREMIS:METS
conversion & validation tools

FCLA

Convert: uri | file upload | direct input

Accepts either:

1. an XML document conforming to the PREMIS schema and creates a METS document with embedded PREMIS elements.
2. an XML document conforming to the METS schema and creates a PREMIS document from any embedded PREMIS.

Enter XML Document:

```
        <agentIdentifierType>URI</agentIdentifierType>

<agentIdentifierValue>info:fda/daitss/description/v2.1.0</agentIdentifierVal
ue>
    </agentIdentifier>
    <agentName>Format Description Service</agentName>
    <agentType>Web Service</agentType>
    <agentNote>built with JHOVE 1.6 (built on December 12, 2011), DROID 3.0
with DROID signature file version 55.</agentNote>
  </agent>
</premis>
```

☐ Wrap PREMIS in a single digiprovMD section

[Convert]

```xml
<mets xmlns="http://www.loc.gov/METS/" xmlns:premis="info:lc/xmlns/premis-v2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xmlns:xlink="http://www.w3.org/1999/xlink" xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/version18/mets.xsd
 info:lc/xmlns/premis-v2 http://www.loc.gov/standards/premis/premis.xsd http://www.loc.gov/mix/v10 http://www.loc.gov/standards/mix/mix10/mix10.xsd">
 <amdSec>
  <techMD ID="object-1" ADMID="event-1">
   <mdWrap MDTYPE="PREMIS">
    <xmlData>
     <object xmlns="info:lc/xmlns/premis-v2" xsi:type="file">
      <objectIdentifier>
        <objectIdentifierType>local</objectIdentifierType>
        <objectIdentifierValue>1</objectIdentifierValue>
      </objectIdentifier>
      <objectCharacteristics>
        <compositionLevel>0</compositionLevel>
       <fixity>
          <messageDigestAlgorithm>MD5</messageDigestAlgorithm>
          <messageDigest>2630c6bd816041e4e10668d74e7bb19b</messageDigest>
          <messageDigestOriginator>Archive</messageDigestOriginator>
       </fixity>
       <fixity>
          <messageDigestAlgorithm>SHA-1</messageDigestAlgorithm>
          <messageDigest>e53539cb3022a1b4ec9e3ef08caf0cdad971229e</messageDigest>
          <messageDigestOriginator>Archive</messageDigestOriginator>
       </fixity>
        <size>103127</size>
       <format>
         <formatDesignation>
           <formatName>JPEG File Interchange Format</formatName>
           <formatVersion>1.01</formatVersion>
         </formatDesignation>
         <formatRegistry>
           <formatRegistryName>http://www.nationalarchives.gov.uk/pronom</formatRegistryName>
           <formatRegistryKey>fmt/43</formatRegistryKey>
         </formatRegistry>
       </format>
       <format>
         <formatDesignation>
           <formatName>JFIF</formatName>
         </formatDesignation>
         <formatNote>Alternate Format</formatNote>
       </format>
       <objectCharacteristicsExtension>
        <mix:mix xmlns:mix="http://www.loc.gov/mix/v20" xsi:schemaLocation="http://www.loc.gov/mix/v20
         http://www.loc.gov/standards/mix/mix20/mix20.xsd">
         <mix:BasicDigitalObjectInformation>
          <mix:ObjectIdentifier>
            <mix:objectIdentifierType>JHOVE</mix:objectIdentifierType>
          </mix:ObjectIdentifier>
           <mix:byteOrder>big endian</mix:byteOrder>
          <mix:Compression>
            <mix:compressionScheme>JPEG</mix:compressionScheme>
```


The future's so bright….

pimtools

about | validation | conversion | describe | resources

PREMIS:METS
conversion & validation tools

FCLA

Validate :   uri | file upload | direct input

Takes a METS document containing PREMIS elements and validates it according to the PREMIS in METS Best Practice Guidelines.

Enter XML Document:

```
    <fileGrp>
        <file ID="file-1" ADMID="object-1" OWNERID="1" SIZE="103127"
CHECKSUM="2630c6bd816041e4e10668d74e7bb19b" CHECKSUMTYPE="MD5"/>
    </fileGrp>
  </fileSec>
  <structMap LABEL="SINGLE_FILE">
    <div>
      <fptr FILEID="file-1"/>
    </div>
  </structMap>
</mets>
```

[Validate]

The future's so bright....

DAITSS Digital Preservation Repository Software

Primary links

- Home
- Documentation
- Download
- News Archive
- Contact

07/19/2012 - 13:22

Udated Chapter 5: The DAITSS Archiving Process

An updated version of the DAITSS documentation Chapter 5: The DAITSS Archiving Process is now available from the Documentation section of the DAITSS website. The SIP specifications section is now consistent with the behavior of DAITSS v.2.17.19.

more

The future's so bright, you gotta wear shades!

Welcome to the DAITSS website!

DAITSS is a digital preservation software application developed by the Florida Center for Library Automation (FCLA) with some support from the IMLS. DAITSS is used by the Florida Digital Archive (FDA), a long-term preservation repository service provided by the Florida Virtual Campus for the use of the libraries of the eleven publicly-funded universities in Florida.

DAITSS provides automated support for the functions of Submission, Ingest, Archival Storage, Access, Withdrawal, and Repository Management. It is architected as a set of RESTful Web Services and micro-services but enforces strict controls to ensure the integrity and authenticity of archived content. It implements active preservation strategies based on format-specific processing including, where necessary, normalization and forward migration. It is particularly well suited for materials in text, document, image, audio and video formats.

DAITSS was written for a multi-user environment and supports consortial as well as institutional preservation repositories.

DAITSS is available for use through a GPLv3 license. This website provides access to a fully configured VM version of DAITSS that can be downloaded to run under any VM manager, along with sample SIPs (submission packages) and documentation.

For more general information about DAITSS, see:

- DAITSS, an OAIS-based Preservation Repository
- DAITSS Grows Up: Migrating to a Second-Generation Preservation System
- The Florida Digital Archive and DAITSS
- Chapter 1 of the DAITSS OSS User Manual

For DAITSS documentation, see the Documentation section of this website, or Contact Us.

Download a DAITSS demo virtual machine

The future's so bright….

For more information

- http://Daitss.fcla.edu (Documentation and Software download)

- http://github.com/daitss (Download source code, view README)

- pcaplan@ufl.edu

The future's so bright….