

DAITSS DIGITAL PRESERVATION SYSTEM: RE-ARCHITECTED, RE- WRITTEN, AND OPEN SOURCE

Priscilla Caplan

Florida Center for Library Automation



The future's so bright....

This briefing...

- A little bit of history
- Functions
- Key features
- DAITSS architecture
- DAITSS technical environment
- Open Source site
- VM Demo



The future's so bright....

History

- DAITSS in use by Florida Digital Archive since 2005
- Written by FCLA with support from the IMLS
- OAIS-based preservation repository system
- Designed for consortial / multi-user environment
 - 11 publicly funded universities of Florida



The future's so bright....

DAITSS 2 project

Rewritten from scratch in 2009-2011

Keep existing functionality but

- compartmentalize format support
- make it easier to modify code
- make it easier for other sites to implement
- improve performance and scalability
- improve interoperability
- eliminate unnecessary complexity
- improved support for operations



The future's so bright....

Functions supported

- Submit = accept or reject a package 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
- User interface for submission, requests, tracking



The future's so bright....

Key features: (1) production

- Designed for heavy duty production
- Can do individual submissions, but oriented to batches
- Maximum throughput, minimum hand-holding
- Throttle for simultaneous ingest, disseminate, withdrawal
- FDA stats:
 - On 12/1: 309,534 AIPS, **41 million files, 105 TB**
 - Growing by 4 TB/month (one copy)
 - Used by 11 universities
 - 1 FDA manager (librarian), 1 operations (tech)



The future's so bright....

Key features: (2) Standards

- Highly standards conformant
- Serious about OAIS
- METS for SIP, AIP and DIP descriptors
- XML for reports and outputs
- PREMIS and format-specific technical metadata
- RESTful Web services
- Support for TIR RXP submission and export in process



The future's so bright....

Key features: (3) Interoperability

- Standards
- Uses existing tools when possible (JHOVE, JHOVE2, DROID, etc.)
- Web services can stand alone, be used in other environments (e.g. pim.fcla.edu)
- Dark archive can be integrated with access system on front end or back end



The future's so bright....

Key features: (4) Auditability

- No software or operator action without audit trail
- Permanent record of operator and preservation events
- No “under the covers” actions required
- Highly controlled access to storage
- Peek to view storage without dissemination



The future's so bright....

Key features: (5) Format support

- Per-file processing:
 - Identification, validation, description
 - Action Plan defines additional steps:
 - Create a derivative in a normalized format
 - Create a derivative in a forward-migrated format
 - Download missing schema required for validation
- Per-file done during Ingest, Refresh and Dissemination
- Automatic migration, mass migration, migration on request
- PREMIS file and bitstream description
- Can add support for a new format with no coding (sometimes, depending)



The future's so bright....

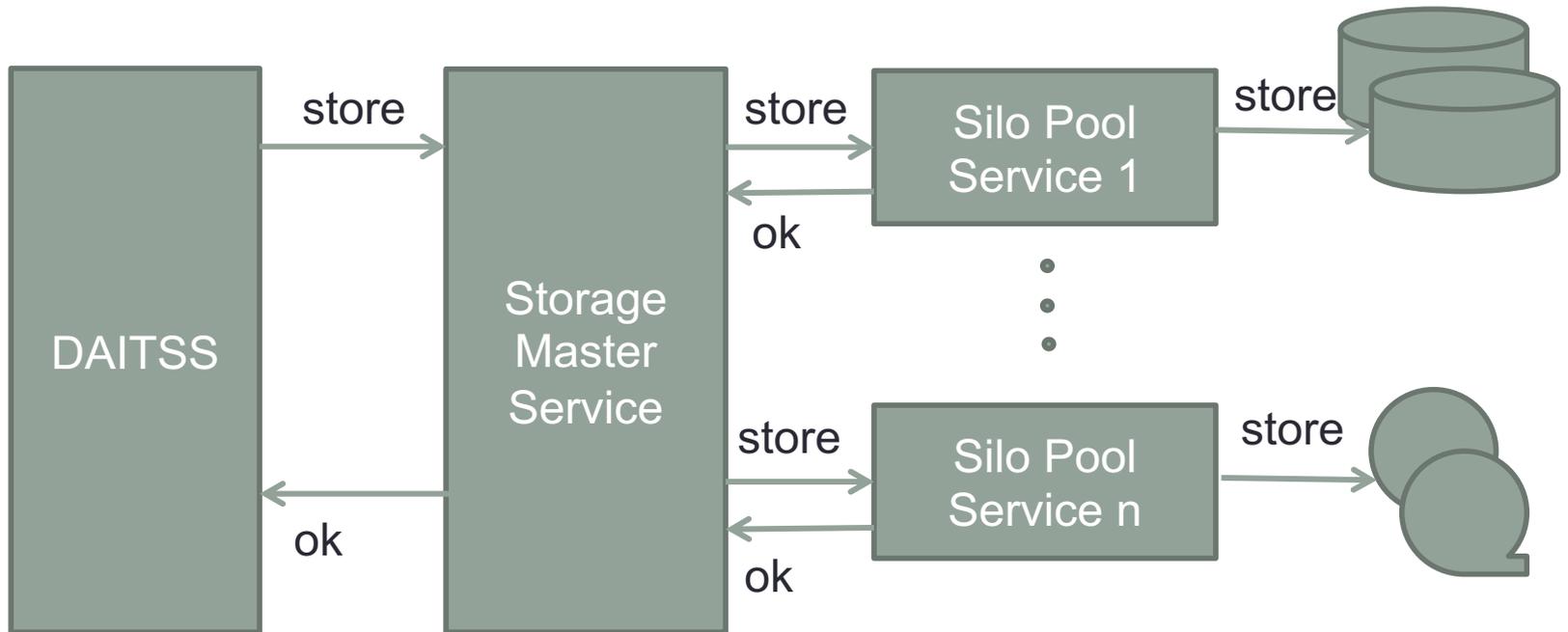
Key Features: (6) Storage management

- All access to Archival Storage (AIPS) through Storage Master Service
- Storage silos (n GB) assigned to silo pools (managed sets of similar silos)
- Storage master ensures that n AIP copies are written to n discrete silo pools
- Silo pool service will fill silos most efficiently
- Silo pool services run continuous fixity checks



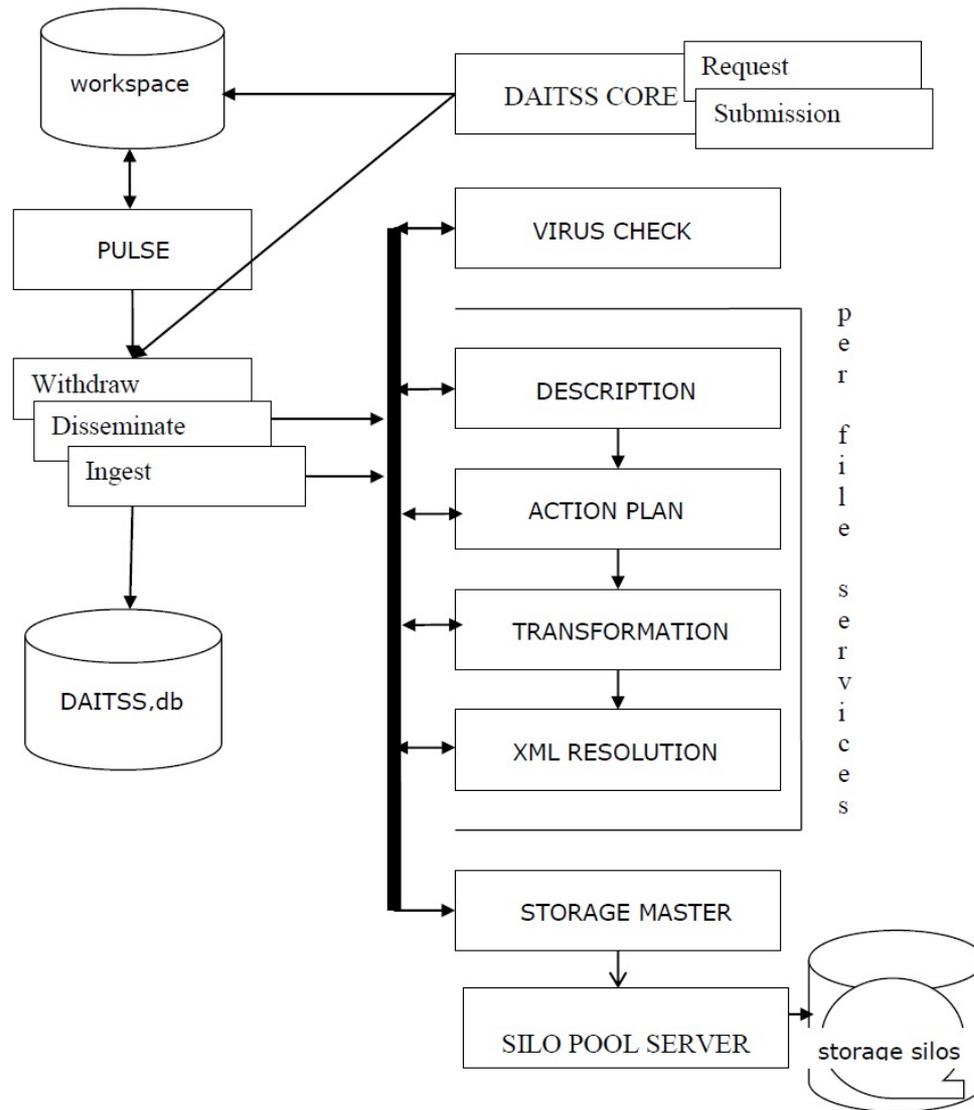
The future's so bright....

Storing an AIP



The future's so bright....

DAITSS Web Services



The future's so bright....

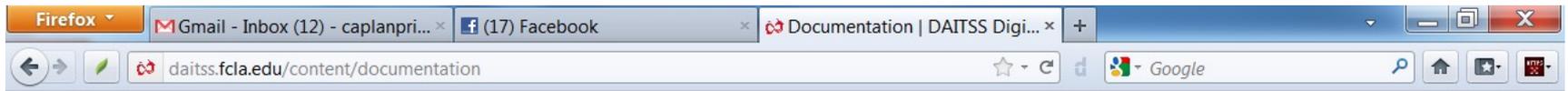
Technical

- Red Hat or CentOS linux
- Coded in Ruby
- Java / RJB for external tools (DROID, JHOVE, etc.)
- Databases PostGreSQL
- Web services run under Thin proxied by Apache
- Squid caching server

- Complete VM version available on DAITSS website (daitss.fcla.edu)



The future's so bright....



DAITSS Digital Preservation Repository Software

[Home](#) [About DAITSS](#) [Documentation](#) [Download](#) [News Archive](#) [Contact](#)

- [Home](#)
- [About DAITSS](#)
- [Documentation](#)
- [Download](#)
- [News Archive](#)
- [Contact](#)

[Home](#)

Documentation

[DAITSS Installation Guide](#) (PDF)

[DAITSS Demo Virtual Machine Quick Start Guide](#) (PDF) *(last updated: December 7, 2011)*

[DAITSS Operations Manual](#) (PDF)

- [Chapter 1: DAITSS Overview](#) *(last updated: October 25, 2011)*
- [Chapter 2: Getting Started](#) *(last updated: October 28, 2011)*
- [Chapter 3: DAITSS Software Components](#) *(last updated: October 25, 2011)*
- [Chapter 4: DAITSS Preservation Services](#) *(last updated: October 25, 2011)*
- [Chapter 5: The DAITSS Archiving Process](#) *(last updated: November 28, 2011)*
- [Chapter 6: DAITSS Workflow Interface](#) *(last updated: October 26, 2011)*
- [Chapter 7: DAITSS Archival Storage](#) *(last updated: October 25, 2011)*
- [Chapter 8: Fixity](#) *(last updated: October 25, 2011)*
- [Appendix A: Sample AIP Descriptor](#) *(last updated: October 25, 2011)*

[DAITSS Data Entity Dictionary](#) (PDF) *(DRAFT)*

[DAITSS METS SIP Profile](#) (PDF)

© 2011 The Florida Center for Library Automation, 5830 NW 39th Ave, Gainesville, FL 32606; 352.392.9020



The future's so bright....