CTDA Technology

Fedora Commons

CTDA is built on the open source Fedora Commons architecture and uses a number of open source tools, applications, and services. Fedora (Flexible Extensible Digital Object Repository Architecture) was originally developed at Cornell University and the University of Virginia in the late 1990s. Fedora application development is currently managed by Duraspace, a non-profit organization dedicated to preserving scientific and cultural heritage digital assets. The Fedora repository architecture is currently in production in hundreds of academic, governmental, and cultural organizations around the world.

Islandora

CTDA’s current management and presentation platform is Islandora. Islandora is an open-source software framework designed to help institutions and organizations and their audiences collaboratively manage and discover digital assets. Islandora was originally developed by the University of Prince Edward Island’s Robertson Library, but is now implemented and contributed to by an ever-growing international community. Islandora development is now managed by the Islandora Foundation, a non-profit membership organization.

Handle.Net

CTDA implements the Handle System to provide persistent and unique identifiers for digital assets. The Handle System is a component of CNRI’s Digital Object Architecture and provides a means of managing digital information in a network environment. The Handle System includes an open set of protocols, a namespace, and a reference implementation of the protocols that enable a distributed computer system to store identifiers, known as handles, of arbitrary resources and resolve those handles into the information necessary to locate, access, contact, authenticate, or otherwise make use of the digital assets. This information can be changed as needed to reflect the current state of the identified digital asset without changing its identifier, thus allowing the name of the item to persist over changes of location and other related state information. All metadata is covered under the Creative Commons Zero License.

File Format Support

CTDA is committed to providing long-term access to all deposited content by applying best practices for digital preservation while also acknowledging the complexities involved in preserving digital information. CTDA will work within evolving “best practices” to be responsible stewards and will work to preserve the metadata and supported digital and media objects. CTDA will research the “state of the art” preservation practices and offer preservation ideas and recommendations. Choices regarding preservation will be based on CTDA’s community demand. Assessments will be continuously undertaken regarding ongoing technical feasibilities and digital repository “best practices.”
Any format can be deposited into CTDA. However, CTDA through Islandora provides services such as the creation of derivatives and technical metadata extraction for the following formats: doc, docx, ods, txt, html, sxc, csv, tsv, odf, ppt, pptx, xls, xlsx, warc, PDF, tiff/tif, jpg/jpeg, png, bmp, wav, mp3, ogg, mp4, mov, qt, m4v, and avi.

Technical Outline

Institutional affiliation and other indicators of viability of the project

- Name of environment: Connecticut Digital Archive
- Name of major systems: Fedora and Islandora
- Version of major systems: Fedora 3.8.1 and Islandora 7.x.1-7 (Drupal 7)
- URL of project homepage: http://ctdigitalarchive.org
- Age of project: Launched November 13, 2013
- Notes on long-term viability of project: CTDA is a service of the University of Connecticut Libraries in collaboration with the Connecticut State Library. It is supported with a base funding from these institutions.
- Degree of open-source license: CTDA is built using open source technologies. All metadata has been dedicated to the public pursuant to Creative Commons’ CC0 public domain dedication (see CTDA Metadata Guidelines).
- Other documentation: Information about CTDA, including governance, mission, vision, resources for participants, or services are available at: http://ctdigitalarchive.org.

Technical requirements, maintenance, and scalability

- Local install or ASP?: There is one implementation of Fedora and multiple Islandora presentation sites.
- Operating system: Red Hat Enterprise Linux Server release 6.6 (Santiago)
- Hardware: Dell Intel-based servers running VMware
- Application server: Fedora, Islandora, Handlebar, MySQL
- Web server requirements: Apache Tomcat, Apache HTTP
- Primary programming language: PHP
- Auxiliary programming language(s): Drush, Perl
- Application framework: System and software development in CTDA is driven by the need to fulfill its mission and vision. This has resulted in a modular architecture where discrete systems fulfilling different OAIS functions (e.g. object ingest, storage, indexing) communicate and interoperate as an integrated whole. Disaggregation of the functional components of the repository allows agile response to problems that arise (e.g., issues with ingest, storage, or access systems are localized and may be addressed separately) and sharing of development responsibilities across partner institutions. Although many repository systems and services sit on central servers, the modular architecture and orientation toward open standards and open systems make it possible for partner institutions to develop services and key pieces of repository functionality
- Database server requirements: MySQL
- Other software requirements: Drupal, Solr, Command Line tools
- Required skills: Significant knowledge of UNIX, PHP, Apache, MySQL, Java
- Internal backup and restore functions: Backup and restore functionality is provided at a system level and consists of a) file backup b) database backup and c) virtual machine. Backup services are currently provided by Unix Systems Administrator.
- Disaster recovery plan: The University of Connecticut Libraries maintains a disaster recovery plan.
- Scalability
  - Application: Applications are lightweight and served by multiple servers; additional web servers can be added to increase application performance.
  - Data: CTDA uses Dell equal logic storage.
- Batch ingest: Batch ingest is handled by Islandora and relies on a zip submission package.
- Batch export: CTDA does not support batch export at this time.
- Support for OAI harvesting: OAI harvesting is handled by Islandora and does not include support for deletions at this time.

Security

- Access control: Access to digital items is determined by rights statements and is handled through Islandora’s XACML Policies.
- User management: The University of Connecticut Libraries manages CTDA users as described in the CTDA Managing Users Guidelines.

Storage

- Add content: Content is added through Islandora.
- Access content: Content is accessed and presented through Islandora.
- Remove content: Content may be removed at the request of the creator/contributor, participating institution’s discretion, CTDA’s discretion, or by legal order. Content files are removed and a tombstone record is made available.
- Manage metadata: Descriptive metadata is managed through Islandora. This metadata includes information on rights. Structural metadata is managed by Fedora’s RELS-EXT and RELS-INT relationship data streams. Technical metadata is managed through Islandora as a data stream. Preservation metadata is created on the fly through Islandora.

Aggregation

- Multiple roles: The ability to change user roles and permissions is held by a limited group of developers and administrators.
Create aggregation: CTDA aggregates first digital objects based on namespaces, which identify the institution that holds or acts as steward for the particular digital objects for management and harvesting purposes. Second, CTDA aggregates digital assets for the Connecticut History Illustrated using the metadata tag, CHI.

Remove aggregation: All participating institutions are required to be assigned and implement their unique namespace. However, the aggregator metadata tag CHI can be removed from any digital object.

Change aggregation membership: Digital objects can be shared or migrated with other collections. Aggregation can only be changed for CHI.

Find aggregation members: All digital objects are part of a namespace. Namespaces are easily identified and found in the repository. It is possible to facet search results in CTDA to limit results from a particular institution, which corresponds with a namespace.

Administrative functions

- Multiple roles: The ability to change user roles and permissions is held by a limited group of developers and administrators.
- Configuring roles: Roles are configured using Drupal.
- Versioning: Data streams are versioned in Fedora.
- Archiving: CTDA provides management, secure storage, preservation solutions, and – whenever possible- open access to digital assets of enduring value, as determined by CTDA participants.

Access and formats functions

- Accessibility of system: The CTDA system and interfaces are currently not designed to provide access for users with disabilities.
- Internationalization support: Unicode encoding (utf-8 basic multilingual plane) is supported in repository applications.
- Citation linking: Each digital assets receives a persistent and unique identifier using the Handle System.
- RSS feed: There is both an RSS feed and CSV download.
- Search and retrieval: CTDA offers full text, faceted, simple and advanced searching.
- Indexing: Indexing is provided through Solr and Gsearch.
- Help Support: A contact email address, ctda@uconn.edu, is provided on the website.