AMNH Functional Requirements for an Archives Content Management System

created October 9, 2013
Draft Version 1, February 25, 2014

OVERVIEW

To establish core functional requirements for an Archives Content Management System. The system will be the repository for the master records supporting archival collections held in the American Museum of Natural History Library. Accessions records, finding aids and entity records, are some examples of the kinds of records specific to archives. Data management, editing and publication of approved records are key components of the system. Long-term preservation of the data is also an important consideration for choosing a system.

Note: For the 2012 CLIR grant project, the AMNH Library is creating entity records using Encoded Archival Context-Corporate Bodies, Persons and Families (EAC-CPF). The metadata standard is relatively new in the archival community, and at the time of this document, few systems exist to support it (out-of-the-box). The project team is developing workflows to maintain the descriptive records as individual files with an eye towards integrating them into supporting systems when they become available.

AUDIENCE

For high-level administration and funding:
Library/Archive administrators, Library/Archive managers

For technical implementations, database configurations and publication setup:
IT, programming specialists

For data input, management and publication:
Librarians, archivists, data managers, data entry personnel

TEAM MEMBERS

Barbara Mathé, Museum Archivist and Principal Investigator for the 2013 CLIR Hidden Collections grant
Rebecca Morgan, Project Archivist
Iris Lee, Metadata Analyst
Nick Krabbenhoeft, Database and Metadata Consultant (Educopia Institute, MetaArchive Cooperative)
Gregory Raml, Special Collections Reference Librarian
Jen Cwiok, Digital Lab Manager
Susan Lynch, Interlibrary Loans Librarian and Technical Consultant
Mai Reitmeyer, Research Services Librarian

MAIN SECTIONS

1.0   Data Creation
2.0   Data Management
3.0   Data Dissemination
4.0   Technical Administration

RATING SCALE

3 = Essential; 2 = Very useful; 1 = Wish the system supports it

1.0	DATA CREATION	Function	Use case example	Rating (1-3)	Date rated
1.1	**Batch import**	Database will be able to batch import legacy data from a number of sources, and will retain significant relationships between records as well as external links. Field mapping should be designated by the user or specified in xml to eliminate the need for data cleanup.			
1.1.1	Spreadsheet import	Ingests legacy data from spreadsheets		3	11/12/2013
1.1.2	Spreadsheet import	Can associate spreadsheet headers to specific data fields within the database	Minimal catalog records captured in spreadsheets, mapped to MARC, can be mapped to specified data fields in the database	3	11/12/2013
1.1.3	XML import	Ingests xml files		3	11/12/2013
1.1.4	XML import	Imports ead		3	11/12/2013
1.1.5	XML import	Imports eac-cpf		3	11/12/2013
1.1.6	XML import	Imports marcxml or marc21slim		1	11/12/2013
1.1.7	XML import	Imports Dublin Core		1	11/12/2013
1.1.8	XML import	Imports extended schemas	Dublin Core with extended metadata fields (Image database)	1	11/12/2013
1.1.9	XML import: identifiers	Metadata imported from xml retains record identifiers for name entities	Name entities with assigned unique identifiers will not lose their <eac-cpf:recordId> during batch import processing	3	11/12/2013
1.1.10	XML import: links	Ingested metadata retains relationships that link between individual records	Finding aid links to name entity; Name entity links to other name entities	3	11/12/2013
1.1.11	XML import: links	Metadata imported from xml retains links to external sources (xlink:href)	Links to URIs are retained in <eac-cpf:sources>	3	11/12/2013
1.1.12	XML import: links	Metadata imported from xml retains xlink attributes that specify the behavior when		1	11/12/2013

		user clicks on external links			
1.1.13	Database migration	Complete data migration from external database using a single action	All data from AT can be migrated as a whole: Accessions, Names, Subjects, Resources, Digital Objects	1	11/12/2013
1.1.14	Controlled vocabularies	Imports controlled vocabularies, such as Library of Congress authorities, Getty Thesauri, VIAF, etc.		2	11/12/2013
1.1.15	Controlled vocabularies	Imports locally-controlled vocabulary lists		3	11/12/2013
1.2	**Data input**	Database will support various data set descriptions and be able to link to other records within the system.			
1.2.1	Finding aids	Can create descriptions in the ead standard for finding aid records		3	11/12/2013
1.2.2	Entities	Can create descriptions in the eac-cpf standard for name entity records		2	11/12/2013
1.2.3	Digital object description	Can create descriptions for digitized image-based material, at the collection- and item-level in Dublin Core or MODS		2	11/12/2013
1.2.4	Accession description	Can create an accession record		3	11/12/2013
1.2.5	Multi-level description	Supports description for archival records at the collection-, series-, and item-levels		3	11/12/2013
1.2.6	Linking records: internal	Within the system, records can link to other records; must link one-to-many	Entity records link to other entity records as well as collection records; collection records link to entities, collections, items, accessions, and subjects	3	11/12/2013
1.2.7	Linking records: internal	Records can link within their own table (cpf to cpf, for example)		3	11/12/2013
1.2.8	Linking records: internal	Records can link outside their table (ead to cpf, for example)		3	11/12/2013

1.2.9	Linking records: external	Records can link to other records in external databases within the institution	Enitity or collection records link to specimen collections in KE EMu Science databases	2	11/12/2013
1.2.10	Controlled vocabularies	Lookup lists for controlled vocabularies, such as Library of Congress authorities, Getty Thesauri, VIAF, etc.		2	11/12/2013
1.2.11	Controlled vocabularies	Lookup lists for locally-controlled vocabulary lists		3	11/12/2013
1.2.12	Control description	History of the record descriptions can be documented	<eac-cpf:mainentanceHistory>	3	11/12/2013
1.2.13	Multimedia	Stores multimedia objects as supporting data, such as, images (jpeg, tiff), documents (pdf), email correspondence, and moving images	Deed of gift documentation saved as a pdf to support an accession record; Born-digital records, such as email, can be linked to a collection of personal papers	2	11/12/2013
1.2.14	Multimedia	Playback for moving images, image, and graphic files		2	11/12/2013
1.2.15	Multimedia	Multimedia can be added to system as special collection digital objects with descriptive metadata	*Are video files considered digital objects? May be the same functionality as 1.2.3*	2	11/12/2013
1.2.16	Multimedia	Multimedia can be added to system as administrative records with no descriptive metadata		2	11/12/2013
1.2.17	Multimedia	If 1.2.16 is available, collection multimedia and administrative multimedia are stored separately in the system to distinguish their value as records		3	11/12/2013
1.2.18	Special characters	Supports UTF-8		3	11/12/2013
1.3	**Identification**	Unique identifiers for different types of records must be supported within the system. The library has various ways to enumerate collections; the system should support manual input and automate a unique string based on a specified format.			
1.3.1	Record	Unique identifiers can be		3	11/12/2013

		identifiers	assigned to record entries			
1.3.2	Record identifiers		Unique identifiers can be expressed in multiple naming structures	Collections: MSS .K35 Entities: amnhc_2000012 Accessions: FY2010-001 (example)	3	11/12/2013
1.3.3	Record identifiers		Different components of a record entry can be assigned a unique identifier		3	11/12/2013
1.3.4	Record identifiers		Allows the format of the identifier to be specified	Assign prefix followed by a sequential number, or manually input cutter number	3	11/12/2013
1.3.5	Record identifiers		Generates unique identifiers automatically		3	11/12/2013
1.3.6	Record identifiers		Allows input of a unique identifier, but validate that it is unique before it is assigned		3	11/12/2013

2.0	DATA MANAGEMENT	Function	Use case example	Rating (1-3)	Date rated
2.1	**Management tools**	Management tools offer functionality to: 1) perform actions on descriptive data (collection- or item-level metadata and the structure which supports them), and 2) track movement or actions performed on the physical objects or items in the collection. Other management capabilities include running reports (data analysis) and performing queries. The system should manage actions for repeat use and batch processing. A graphic user interface (GUI) for queries should be available for basic searches performed by archivists, managers, and data entry specialists. Programmed queries can be run as a script by technical staff.			
2.1.1	Merge records	Merge two or more records into a single record		3	11/20/2013
2.1.2	Split records	Split a single record into multiple records		3	11/20/2013
2.1.3	Delete records	Ability to permanently delete records		3	11/20/2013
2.1.4	Deaccession	Deaccession records and store them in the system without permanently deleting the data		3	11/20/2013
2.1.5	Deaccession (object)	Stores data about deaccessioned objects in the collection		3	11/20/2013
2.1.6	Preservation of collections (object)	Record preservation and conservation needs of the physical collections	<phystech> *Physical Characteristics and Technical Requirements* in EAD. Do we need it to be more descriptive or discrete in details?	3	11/20/2013
2.1.7	Location (object)	Can record physical location of the collection or item	<physloc> – *Physical Location* in EAD. Oversize prints in	3	11/20/2013

			Memorabilia Room.		
2.1.8	Reports	Different types of reporting options available		3	11/20/2013
2.1.9	Reports	Performs analysis to descriptive data	Error reports, records inconsistencies, record completion rate.	3	11/20/2013
2.1.10	Reports	Tracks usage		3	11/20/2013
2.1.11	Reports	Gathers statistics		3	11/20/2013
2.1.12	Queries/Sort	Performs sophisticated queries within a single table of data	Run a query just in finding aid collections, or just in the container lists within finding aid collections	3	11/20/2013
2.1.13	Queries/Sort	Performs sophisticated queries in multiple tables of data	Run a query in both entities and accession records	3	11/20/2013
2.1.14	Queries/Sort	Allow users to save and reuse queries			11/20/2013
2.1.15	Global changes	Makes changes to multiple records at a time		3	11/20/2013
2.1.16	Public access restrictions	Allow fields to be suppressed from display	For instance, suppress specific descriptive elements from public finding aid	3	11/20/2013
2.2	**Workflow**	In addition to management tools, it is important for the archivists and data managers to define a workflow process specific to individual projects, create staging zones for various points in record creation and description maintenance, assign status terms to records to inform the approval process, and maintain a record history for data as it is created, updated or otherwise changed. Allowing a "preview" mode for finding aids and entities is also an important functionality that facilitates the approval process.			
2.2.1	Defining workflow	Ability to create a custom workflow		3	11/20/2013
2.2.2	Defining workflow	Allows defining of multiple workflows		3	11/20/2013
2.2.3	Defining workflow	Can select specific workflows for specific projects		2	11/20/2013

2.2.4	Status	Ability to assign a status to a record	Such as "draft", "audit", "publish"	3	11/20/2013
2.2.5	Status	Status names or values can be customized	Examples of statuses are "new", "in progress", "draft"	2	11/20/2013
2.2.6	Record history	System is able to restore a previous version of a record		2	11/20/2013
2.2.7	Record history	System is able to restore all previous versions of records		1	11/20/2013
2.2.8	Record history	History for individual records is captured in the record's metadata	You can see all the changes made to an individual finding aid or individual entity	3	11/20/2013
2.2.9	Record history	Records the person (userID or equivalent) responsible for the changes made to individual records		3	11/20/2013
2.2.10	Preview	Ability to preview a draft publication	Available in html view or formats data in a print preview.	3	11/20/2013
2.3	**Permissions**	Different levels of users should be allowed structured access to the system and its records.			
2.3.1	Permissions	Restrict access to system functions according to a user's role. Database will be able to assign certain functionality and access to defined sets of user groups.		3	11/20/2013
2.3.2	Restrict data	Specified records restricted from specified user groups		3	11/20/2013
2.3.3	Restrict data	Specified data fields restricted from specified user groups	Interns restricted from editing repository information	3	11/20/2013
2.3.4	Users	**Admin** Allowed to change interface, publishing preferences/profiles, add plugins/extensions, manage data exchanges, manage users and specify permissions, no restrictions		3	11/20/2013
2.3.5	Users	**Project Manager** Allowed to manage users, add/delete/edit/merge/split records, publish reports and public-facing documents, manage local vocabulary lists, restricted from Admin		3	11/20/2013

		functions			
2.3.6	Users	**Data Editor** Allowed to add and edit records in "draft" mode, can link records, resctricted from Admin and Project Manager functions		3	11/20/2013
2.3.7	Users	**Training/Probationary** account Allowed to view records in the "cataloging" interface, restricted from making any changes		3	11/20/2013

3.0	DATA DISSEMINATION	Function	Use case example	Rating (1-3)	Date rated
3.1	**Public access**	Pending approval from administrators and managers, records will be published to the internet for public access. The system should support output for web display or print. Any linking attributes should be retained in the output. The project team will discuss the best way to integrate APIs for linked open data as applications become available.			
3.1.1	Finding aids	Publishes finding aids for online access (html or xml)		3	12/10/2013
3.1.2	Finding aids	Publishes finding aids for print (pdf)		3	12/10/2013
3.1.3	Entities	Publishes entity pages for online access (html)		3	12/10/2013
3.1.4	Entities	Publishes entity pages for print (pdf)		3	12/10/2013
3.1.5	Exporting to external databases	Exports records to search engines and existing data portals	Exports catalog records to the library OPAC; Finding aids exported to a future AMNH portal	3	12/10/2013
3.1.6	Linking records	Published documents retain links to records and URIs	Finding aid with links to entities are retained in the web or print output	3	12/10/2013
3.1.7	Web integration	Web publishing platform available with the system	Similar to the exhibition 'mode' in Omeka	3	12/10/2013
3.1.8	Social integration	Records may be enriched by the general public within the public interface	Public can offer edits or additions to records (bio/hist notes, or related resources) that can be vetted and approved by an administrator.	1-2	12/10/2013
3.1.9	API integration	Allows integration of APIs	*Look for automatic integration or setting up a manual sync.*	TBD	12/10/2013

4.0	TECHNICAL MANAGEMENT	Function	Use case example	Rating (1-3)	Date rated
4.1	**System administration**	Systematic tasks and reporting should be built into the system, accessible through the administrative interface or programmable. Automatic backup and recovery procedures should be available and be able to be set up by the Technical Administrator. The system should maintain data integrity by performing checksum functions.			
4.1.1	Cron jobs	Run tasks systematically, on a defined schedule	Exporting updated records to online catalogs and portals on a regular schedule	3	1/7/2014
4.1.2	Backup and recovery	Provide automatic backup and recovery procedures		3	1/7/2014
4.1.3	Backup and recovery	Allows administrator to specify the back-up schedule and location		3	1/7/2014
4.1.4	Recovery	Ability to recover selected files		3	1/7/2014
4.1.5	Reports	Ability to run automated daily integrity reports	*How do you run a report? Through the system interface or the backend?*	3	1/7/2014
4.1.6	Sustainability	Ability to run checksum function on data		3	1/7/2014
4.2	**Metadata compatibility**	Established schemas, specifically EAD and EAC-CPF, should be supported within the system. Along with descriptive standards, technical metadata and preservation metadata should also be available, or be made available to the system. Catalog formats, such as MARC, is desirable though not necessary since the system should export records to the online library catalog system. Ability to load or define custom schemas may be desirable in the future but not necessary at this time.			

4.2.1	EAD	Supports ead for collection description		3	1/7/2014
4.2.2	EAC-CPF	Supports eac-cpf or can support eac-cpf with customization, for entity description		3	1/7/2014
4.2.3	MARC	Exports marcxml or marc records (.mrc)		2	1/7/2014
4.2.4	Other	Supports other established or local schemas in conjunction with ead and eac-cpf	Dublin Core for image descriptions, MODS for object descriptions	3	1/7/2014
	Other	Ability to add other established standards for digital preservation or technical metadata (not just descriptive)	METS, PREMIS	3	1/7/2014
4.2.5	Custom	Ability to add custom schemas	Nice to have in case a new standard comes out or it does not support 4.2.3 or 4.2.4 above	2	1/7/2014
4.3	**Data export**	All data in the system can be exported into an open format, such as xml or plain text as comma separated values (csv). The project team will need to discuss how and if the public data should be harvestable.			
4.3.1	XML export	Exports all descriptive data in xml	Exports collection, entity, and item/object descriptions	3	1/7/2014
4.3.2	XML export	Uses existing xml schemas to export data, or user can customize xml export settings to facilitate xml transformations	Collection, entity, item descriptions in ead, eac-cpf, Dublin Core respectively, or can be easily transformed using xslt	3	1/7/2014
4.3.3	CSV export	Exports all data in csv format		3	1/7/2014
4.3.4	CSV export	Exports selected data to csv format		3	1/7/2014
4.3.5	Harvesting	Allows OAI-harvesting of public finding aids and entities		TBD	1/7/2014

4.4	System support	The longevity of a strong system can be measured, in some part, by the support it gets from its developers and user community. Updates that improve functionality and security should be available and announced to its user group conspicuously. User manuals and other tools should be accessible. A technician on staff to answer individual concerns would be helpful, but the project team would need to discuss the cost-benefit if this kind of support requires a financial commitment. Social/community venues for documenting use cases can also provide valuable information. Not listed here as a required function, though perhaps desirable in the current technical landscape, integration of custom enhancements/plugins/APIs created by the user community: to be discussed. Other concerns to discuss in the future involve hosting and storage.			
4.4.1	Program updates	Updates for the system are available online. Notification of updates communicated through the program itself or via email.	*Frequency for maintenance of support? Frequency of updates? Would there be a security risk?*	3	1/7/2014
4.4.2	Documentation	Users manual available in hard copy or online		3	1/7/2014
4.4.3	Technical Support	Technician or programmer available for problems or questions via email, phone or in-person consultation		2	1/7/2014
4.4.4	Community Support	Community of users can contribute questions and issues within an online users group or listserv		3	1/7/2014
4.4.5	System Wiki	Online documentation available in wiki format	(The benefit to this being that it is a collaborative tool open to multiple users and rich user history)	2	1/7/2014