

Requirements for discovery and delivery of document level objects

Proposed list of metadata fields to drive the discovery and delivery of document level objects

The following is a list of descriptive, technical and administrative metadata that may accompany a document in Hypatia. Documents can be either a single file or a grouping of files that exist together in either a folder, a zipped archive or a disk image. Not all documents will necessarily have metadata for all of these metadata fields.

Field Name	ISAD(G) Element	Description	Searchable?	Facet?	Display?	Sortable?	Allow edit?	Metadata Source
		Fields required for Standards-Compliant Archival Description						
Repository	ISDIAH 5.1.2	Name of archival unit responsible for the collection.	Yes	Yes	Yes	Yes		collection object
collection call number	3.1.1		Maybe	Yes	Yes	Yes		collection object
collection title	3.1.2		Yes	No	Yes	Yes		collection object
accession number			No	No	Maybe?	No	Yes	?
Document identifier	3.1.1		Yes	No	Yes	Maybe	No	autogenerated
archival context		Location of the document in an intellectual arrangement (series, subseries, etc.)	No	Yes (with collection title)	Yes	No	Yes (archivist)	FTK / Parent object(s) / EAD
Level of description	3.1.4	Identifies the level of arrangement of the unit of description	No	Maybe	Yes	No	Yes (archivist)	
Conditions governing access (facet)	3.4.1	To provide information on the legal status or other regulations that restrict or affect access to the unit of description. Peter: I used the following controlled vocabulary for AR - Access restrictions: AR:Owner; AR:Archivist; AR:Invited person; AR: Public; AR:Reading room	Yes (need controlled vocabulary)	Yes	Yes	Maybe	Yes (archivist)	FTK?
Conditions governing access (note)	3.4.1	To provide information on the legal status or other regulations that restrict or affect access to the unit of description. Peter: I used the following controlled vocabulary for AR - Access restrictions: AR:Owner; AR:Archivist; AR:Invited person; AR: Public; AR:Reading room	Maybe	No	Yes	No	Yes (archivist)	EAD?
Conditions governing use /reproduction	3.4.2		Yes (need controlled vocabulary)	Yes	Yes	Maybe	Yes (archivist)	FTK / EAD
Conditions governing use /reproduction (note)	3.4.2		Maybe	No	Yes	No	Yes (archivist)	EAD
Scope and contents	3.3.1		Yes	No	Yes	No	Yes	
Creator	3.2.1		Yes	Yes	Yes	No	Yes (archivist)	FTK / parent object ?
subject heading, name, etc. (manually assigned)			Yes	Yes	Yes	No	Yes (archivist)	FTK /
subjects, name, place (software generated)			Yes	Yes	Yes	No	Yes (archivist)	Entity extraction software /service (e.g. OpenCalais)
Citation			No	No	Yes	No	Yes (archivist)	
document title	3.1.2	Title supplied by archivist describing the document	Yes	No	Yes	Yes	Yes (archivist)	EAD?
document date	3.1.3	Is this the creation date or last modified date. Do we need both?	Yes	Yes (need both)	Yes	Yes		FTK / Ingest
document size	3.1.5	Indicates the file or document's size on a filesystem	No	No	Yes	No		FTK / Ingest
		Additional fields required for assets						
source media	3.4.4	Description of the physical carrier for a record (floppy disk, hard disk, etc.) Peter: I used the following controlled vocabulary for CM - Computer media: CM:5.25 floppy; CM:3.5 floppy; CM:Punch card; CM: CD/DVD; CM: Hard Drive; CM: Zip Disk: CM: Tape; CM: Cloud Storage;	No	Yes (need controlled vocabulary)	Yes	No		FTK /
operating system and version (if known)	3.4.4	Peter: I think this field is not necessary. Also, I don't know any tools I can get this info. Files can be created by different os and stored in 1 computer.						
document type		Controlled value list. Is this a text document, image, audio, video, forensic image etc. Where is this list coming from? Peter: I used the following controlled vocabulary for FT - Format Type: FT:Document; FT:Spreadsheet; FT:Computer Program; FT: Image; FT: Video; FT: Audio; FT: Email	No	Yes (need controlled vocabulary)	Yes	No	Yes (archivist)	FTK /
file or document name		Document or file name assigned to an object by an operating system	Yes	No	Yes	No	Maybe	FTK / Ingest
document location		Location of the document on a filesystem. This is different from the archival location of a document in a series / subseries.	No	No	Yes	No		FTK /

mime type (original)	3.4.4	The mime type indicates the type of document and may indicate the application that was used to create the document	Maybe	No	Maybe	No		Ingest
mime type (presentation version)			No	Maybe	Maybe	No		Ingest
application software and version (if known)	3.4.4		No	No	Yes	No	Yes (archivist)	FTK /
thumbnail image		image that represents the document type (eg. PDF, text, image etc.) Peter: If the file is an image, it should be the relative thumbnail.	No	No	Yes	No		FTK for image thumbnail /
"Download" this		button that allows the archivist or end user to download the document (if permitted) Peter: We may also consider adding digital signature of the institution to the files.	No	No	Yes	No	No	
checksum			No	No	Yes	No	No	FTK / Ingest
Take-down request / policy			No	No	Yes	No	Yes (public for request)	Web UI
Original file			No	No	Yes	No	No	
Display version of the original file			No	No	Yes	No	No	
Presentation format history		Automated? piece to say that original file X was converted by Person Y using software Z on this date						
		User-generated content						
annotations ("stories")			No	No	Yes	No	Yes (creator / invited public / public)	Web UI
archivist created tag		tags that archivists/curators add - become facets (How are these different from access points)		yes				Web UI
creator tag		tags by creator - become facets by creator (How are these different from access points)		yes				Web UI
(pre-)approved user tag		tags that are added by approved users outside of the repository/library - should show up in facet as similar to an approved editor in Wikipedia (?)		yes				Web UI
user created tag		tags created by non-approved users; might go through vetting process by repository /library or be listed as unverified/unvetted editor (like Wikipedia?)		?				Web UI