

MDE Data Analysis - Methodology		Test Question	Test Description	Test Method	Primary Focus	Comment	
METRICS	1	<b>Conformance: Does the data conform to CDWA Lite and CCO?</b>					
		Structure	Content	Account for "required" elements. Machine process assemblies sample. Human inspection of sample records	Machine Machine Machine/Human	entire record entire record all present	
				Evaluate whether algorithmic testing viable. If so, test algorithmically. If no, human inspection of sample records. Review CDWA "rules" (see sheet CDWA-CCO). Evaluate for which fields algorithmic testing viable. If so, test algorithmically. If no, human inspection of sample records.	Machine/Human Machine/Human	only CDWA required / highly recommended and non-display only CDWA required / highly recommended	none of the "required" fields are format controlled
				Review CCO book "rules for..." sections. Evaluate if there are fields for which algorithmic testing viable. Reconstruct the list of data values for each field. Evaluate the consistency and length of that list. Compare to data values suggested in CDWA Lite (see sheet CDWA-CCO) and CCO.	(Machine)/Human Machine/Human	only CDWA required / highly recommended or CCO required only CDWA required / highly recommended only CDWA required / highly recommended and non-display	contract with Patricia Harpring for CCO evaluation of sample?
		Value		Check for each "controlled" data element whether TERMSOURCE attribute has been declared. Check for consistency across all records for each data element. Establish a ratio of controlled vocabulary hits in data elements with or without TERMSOURCE. Establish whether hits occur on preferred or alternate terms. Schema validation	Machine Machine Human	only CDWA required / highly recommended entire record / XML entire record / XML	
				Character set validation	Human	entire record / XML	
				<b>2 Connections: What relationships between records does the data support?</b>			
		Structure	Content	Check for data elements which have a value for every single record, bound by institution or for the whole aggregation. ? Check whether hits on controlled vocabulary occur on different synonyms (preferred and alternate) versions of the same term. (see 1.9) ? Ratio of unique terms / documents ? Human inspection of sample records Cross-match all terms. Establish for each data value: if this data value were used as a search term, how many hits would it produce across institutional record sets and the aggregate set?	Machine ?	entire record only CDWA required / highly recommended and non-display	
				Cross-match common terms.....	Machine	entire record	
		Value		Cross-match low frequency terms.	Machine	entire record	

EVALUATION		Test Question	Test Description	Test Method	Primary Focus	Comment	
3	Structure	<b>Suitability: How well do the records support search, retrieval, aggregation?</b>					
		3.1	Is the set of fields required by CDWA Lite adequate for searching?			Need to define "adequate search" first.	
		3.2	Is the set of fields consistently present across an institution or the aggregation adequate for searching?				
		3.3	Evaluate the impact of the lack of subject data for discovery.	?			
		3.4	Evaluate the impact of an unknown creator on the usefulness of the data?	?			
	Content	<b>Enhancement: How can suitability for search, retrieval, aggregation be improved?</b>					
		4.1	Which data elements can be normalized to a more common form for cross-database searching?		Machine	only CDWA required / highly recommended	
		4.2	Can works without creation dates inherit a ca. date range?	For works without a creation date, apply birth/death dates for the creator to generate a ca. date	Machine		
		4.3	Does intermeditation with controlled vocabularies demonstrably enhance retrieval?	Resolve search terms to synonyms/related terms according to controlled vocabularies, and evaluate how this impacts the relationship among records.	Machine	only CDWA required / highly recommended	
		4.4	Can names be enriched in order to match name authorities?	For names that do not include birth/death dates, but where the associated work has a creation date, use the creation date to help disambiguate the name in ULAN or similar authorities	Machine		
		<b>For definitions of data structure, data content, data value and data format, see:</b>					
		Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives and Museums by Mary W. Elings and Günter Waibel <i>First Monday</i> , volume 12, number 3 (March 2007).					
		URL: <a href="http://firstmonday.org/issues/issue12_3/elings/index.html">http://firstmonday.org/issues/issue12_3/elings/index.html</a>					