MDE Data Analyis - Methodology

			Test Question	Test Description	Test Method	Primary Focus	Note	Done?
METRICS	1		**Conformance: Does the data conform to CDWA Lite and CCO?**					
	Structure	1.1	Are the data elements CDWA Lite Spec. marks as “required’ present? In additon, are the data elements marked "highly recommended" present?	Account for "required" elements.	Machine	entire record		Y
		1.2	Are the data elements CCO marks as “required” present?	Account for "required" elements.	Machine	entire record	Patricia Harpring CCO analysis	Y
	Content	1.3	Does the data content comply with the CDWA Lite definition of its element/attribute?	Machine process assembles sample. Human inspection of sample records	Machine/Human	all present	Since CDWA and CCO are identicial twins, and Patricia will answer this question for CCO	Y
		1.4	For a data element marked “format controlled” in CCO, check whethe the data values correspond to the applicable format.	Evaluate whether algorithmic testing viable. If so, test algorithmically. If no, human inspection of sample records.	Machine/Human	only CDWA required / highly recommended and non-display	None of the "required" fields are format controlled	Y
		1.5	Does the data content comply with the guidelines CDWA Lite specifies as "rules"?	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	only CDWA required / highly recommended	See 1.3	Y
		1.6	Does the data content comply with the guidelines CCO specifies as "rules"?	Review CCO book "rules for…" sections. Evaluate if there are field for which algorithmic testing viable.	(Machine)/Human	only CDWA required / highly recommended or CCO required	Patricia Harpring CCO analysis	Y
	Value	1.7	For a data element marked “controlled list” in CCO, evaluate consistency of data values.	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	only CDWA required / highly recommended (expanded to all present)		Y
		1.8	For a data element marked "controlled" in CDWA Lite / CCO, establish whether use of a controlled vocabulary has been declared.	Check for each "controlled" data element whether TERMSOURCE attribute has been declared. Check for consistency across all records for each data element.	Machine	only CDWA required / highly recommended		Y
		1.9	For a data element marked "controlled" in CDWA Lite / CCO, check the data value against applicable controlled vocabularies.	Establish a ratio of controlled vocabulary hits in data elements wit or without TERMSOURCE. Establish whether hits occurr on preferred or alternate terms.	Machine	entire record		Y
	Format	1.1	Does the data validate against the CDWA Lite XML schema	Schema validatior	Machine	entire record / XML		Y
		1.11	Does the data conform to the declared character set encoding (UTF 8)?	Character set validation	Human	entire record / XML		Y
	2		**Connections: What relationships between records does the data support?**					
	Structure	2.1	What is the lowest common denominator of CDWA Lite data elements consistently present for institutions / the aggregate?	Check for data elements which have a value for every single record, bound by institution or for the whole aggregation.	Machine	entire record		Y
	Content	2.2	Evaluate the consistency of the data – are the same terms used to denote the same concept (i.e. each artist is represented by 1 name and 1 name only throughout the data)?	? Check whether hits on controlled vocabulary occur on differer synonyms (prefered and alternate) versions of the same term. (see 1.9) ? Ratio of unique terms / documents ? Human inspection of sample records	?	only CDWA required / highly recommended and non-display		

			Test Question	Test Description	Test Method	Primary Focus	Note	Done?
	Value	2.3	Are the terms used to describe a collection object in record set A usable to find similar objects in another record set?	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		
	Value	2.4	Are common terms from one institutional set of records usec commonly in others?	Cross-match common terms.	Machine	entire record		Y
	Value	2.5	Are low frequency terms in one institutional set of records found i any other institutional record sets? (Low frequency terms are often high quality terms.)	Cross-match low frequency terms.	Machine	entire record		
EVALUATION		3	**Suitability: How well do the records support search, retrieval, aggregation?**					
	Structure	3.1	Is the set of fields required by CDWA Lite adequate for searching?					
	Structure	3.2	Is the set of fields consistently present across an institution or th aggregation adequate for searching?					
	Structure	3.3	Evaluate the impact of the lack of subject data for discovery.	? (Check hits against search terms used by users of other museum resources, i.e. CAMIO.)	?			
	Structure	3.4	Evaluate the impact of an unknown creator on the usefulness of th data	?	?			
		4	**Enhancement: How can suitability for search, retrieval, aggregation be improved?**					
	Content	4.1	Which data elements can be normalized to a more common form for cross-database searching?		Machine	only CDWA required / highly recommended		
	Content	4.2	Can works without creation dates inherit a ca. date range?	For works without a creation date, apply birth/death dates for th creator to generate a ca. date	Machine			
	Value	4.3	Does intermediation with controlled vocabularies demonstrably enhance retrieval?	Resolve search terms to synonyms/related terms according t controlled vocabularies, and evaluate how this impacts the relationship among records.	Machine	only CDWA required / highly recommended		
	Value	4.4	Can names be enriched in order to match name authorities?	For names that do not include birth/death dates, but where th associated work has a creation date, use the creation date to help disambiguate the name in ULAN or similar authorities	Machine			
			For definitions of data structure, data content, data value and data format, see:					
			Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives and Museums by Mary W. Elings and Günter Waibel					
			First Monday, volume 12, number 3 (March 2007),					
			URL: http://firstmonday.org/issues/issue12_3/elings/index.html					