#### **Example EML Error Messages**

Example EML error messages here are gathered from a **Metacat harvest report** and three diagnostic tools

- oXygen XML editor
- KNB emlParser
- pasta quality checks in evaluation mode

Here are a batch of eml docs which are all identical except for one intentional mistake in each. Two are base cases, completely valid and correct (#1 for eml 2.1.0 and #11 for eml 2.0.1.) In some cases I made the same intentional mistake in the older and newer versions of eml to demonstrate how some mistakes were allowed in the older version.

See the accompanying .zip archive for the eml docs, harvest list, and harvest report used for these examples. The example eml docs are named by their docId, as in the left column of the table. View them in oXygen, not in a browser. When these eml docs are viewed in the oXygen eml editor, the mistakes are marked with a red squiggle as well as a message at the bottom of the editor pane.

The KNB parser <u>Validation service for EML</u> will validate that your EML file conforms to the rules in the EML spec. It also runs a SAX validation parse to check if your document is schema valid according to the EML schemas. The KNB parser tool gives instant feedback.

Sometimes error messages contain eml tags and when rendered in a browser those tags dissapear (because they're not html). I tried to escape them but in some cases you will see end tag "" instead of end tag "</title>".

Harvest log reports may contain errors in four categories:

- (1) xml not well-formed
- (2) xml not matching eml schema (eml not schema-valid)
- (3) parser errors with references (missing id, or missing customUnit, or duplicate id).
- (4) harvest issues (duplicate revision, ie a docId already submitted, or the url to the eml file is not-found)

The KNB emlParser will find any errors in categories 1, 2 or 3. An xml editor alone will not alert you to category 3 errors but once you have the error message from the parser you can then locate the problem id. The category of the error will determine your approach. So it is useful to recognise which category of error has happened.

example EML doc docType	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
-------------------------------	---------------------------------	-------------------------	------------------	----------------------------	------------------------------

	2.1.0	This one validates and parses ok.	no error	EML specific tests: Passed.  The tests which are specific to EML, including validation that IDs are present and properly referenced, have passed.  XML specific tests: Passed.  Document is XML-schema valid. There were no XML errors found in your document.	success	Value of the packageId attribute must conform with LTER best practices. The fake scope knb-test-fls is in all examples here so I wont repeat this message.
example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test-fls.2.1	eml 2.1.0	This one is mal-formed XML, with double-quoted root element. Missing  Category (1) xml not well-formed	The element type "title" must be terminated by the matching end-tag "".	EML specific tests: Failed.  The following errors were found: Error running xpath expression: //dateTimeDomainl //nonNumericDomainl //numericDomainl//accessl //attributeListl//constraintl //coveragel//temporalCoveragel //geographicCoveragel //taxonomicCoveragel //dataSourcel //dataTablel//otherEntityl //citationl//addressl //conferenceLocationl//partyl //originatorl//creatorl//contactl //publisherl//editorl//recipientl //performerl//institutionl	Error running xpath expression:  //dateTimeDomainl //nonNumericDomainl //numericDomainl //accessl//attributeListl //constraintl//coveragel //temporalCoveragel //temporalCoveragel //taxonomicCoveragel //dataSourcel //dataTablel //otherEntityl//citationl //addressl //conferenceLocationl //partyl//originatorl //creatorl//contactl //publisherl//editorl	Unable to parse packageId from document. The document may not be well-formed XML.

				//metadataProviderl //associatedPartyl//personnell //physicall //connectionDefinitionl //distributionl//researchProjectl //projectl//relatedProjectl //softwarel//spatialRasterl //spatialReferencel //spatialVectorl //storedProcedurel//viewl //protocoll//additionalMetadata : The element type "title" must be terminated by the matching end-tag "".  XML specific tests: Failed  The following errors were found: The element type "title" must be terminated by the matching end-tag "".	//recipientl//performerl //institutionl //metadataProviderl //associatedPartyl //personnell//physicall //connectionDefinitionl //distributionl //researchProjectl //projectl //projectl //softwarel //spatialRasterl //spatialReferencel //spatialVectorl //storedProcedurel //viewl//protocoll //additionalMetadata: The element type "title" must be terminated by the matching end-tag "".	
example EML doc	docТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test-fls.3.1	eml 2.1.0	This one is mal-formed XML, with single-quoted root element. Missing  Category (1) xml not well-formed	The element type "title" must be terminated by the matching end-tag "".	EML specific tests: Failed.  The following errors were found: Error running xpath expression: //dateTimeDomainl //nonNumericDomainl //numericDomainl//accessl //attributeListl//constraintl //coveragel//temporalCoveragel	Error running xpath expression: //dateTimeDomainl //nonNumericDomainl //numericDomainl //accessl//attributeListl //constraintl//coveragel //temporalCoveragel //geographicCoveragel //taxonomicCoveragel	Unable to parse packageId from document. The document may not be well-formed XML.

//taxonomicCoveragel/datasetl /eml/datasetl//dataSourcel //dataTablel//otherEntityl //citationl//addressl //conferencel.ocationl//partyl //originatorl//creatorl//contactl //publisherl//editorl//recipientl //performerl//institutionl //metadataProviderl //associatedPartyl//personnell //physicall //connectionDefinitionl //distributionl//researchProjectl //projectl//relatedProjectl //spatialReferencel //spatialReferencel //spatialVectorl //spatialVectorl //storedProcedurel//viewl //protocoll//additionalMetadata : The element type "title" must be terminated by the matching end-tag "".  XML specific tests: Failed  XML specific tests: Failed  //dataSourcel //dataTablel //dataSurce //partyl//partyl//partyl //partyl//originator //recatorl//oratorl //recipientl//recipient //metadataProviderl //ssociatedPartyl //personnell//pipsocial //connectionDefinitionl //festivational //connectionDefinitionl //dataTable //conference //publisher//citationl //metadataProviderl //ssociatedPartyl //personnell//pipsocial //connectionDefinitionl //distributionl //researchProjectl //softwarel //spatialVec
--

	eml 2.1.0	This one is invalid EML, with double-quoted root element. Missing entityName  Category (2) eml not schema-valid	E [Xerces] cvc-complex- type.2.4.a: Invalid content was found starting with element 'entityDescription'. One of '{alternateIdentifier, entityName, references}' is expected.	EML specific tests: Passed.  The tests which are specific to EML, including validation that IDs are present and properly referenced, have passed.  XML specific tests: Failed  The following errors were found: cvc-complex-type.2.4.a: Invalid content starting with element 'entityDescription'.  The content must match '((((((((((((((((((((((((((((((((((((	cvc-complex- type.2.4.a: Invalid content was found starting with element 'entityDescription'. One of '{alternateIdentifier, entityName, references}' is expected.	String index out of range: 0
example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion

knb-test-fls.5.1	eml 2.1.0	This one is invalid EML, with single-quoted root element. Missing entityName  Category (2) eml not schema-valid	E [Xerces] cvc-complex- type.2.4.a: Invalid content was found starting with element 'entityDescription'. One of '{alternateIdentifier, entityName, references}' is expected.	EML specific tests: Passed.  The tests which are specific to EML, including validation that IDs are present and properly referenced, have passed.  XML specific tests: Failed  The following errors were found: cvc-complex-type.2.4.a: Invalid content starting with element 'entityDescription'.  The content must match '((((((((((((((((((((((((((((((((((((	cvc-complex- type.2.4.a: Invalid content was found starting with element 'entityDescription'. One of '{alternateIdentifier, entityName, references}' is expected.	String index out of range: 0
_	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion

knb-test- fls.6.1 2.	ml .1.0	This one fails the parser: duplicate ID.  Category (3) parser errors with references	Error not detected because parser errors are not validation errors.	//metadataProviderl //associatedPartyl//personnell //physicall //connectionDefinitionl //distributionl//researchProjectl //projectl//relatedProjectl //softwarel//spatialRasterl //spatialReferencel //spatialVectorl //storedProcedurel//viewl //protocoll//additionalMetadata : Error in xml document. This EML document is not valid	Error running xpath expression: //dateTimeDomainl //nonNumericDomainl //accessl//attributeListl //constraintl//coveragel //temporalCoveragel //temporalCoveragel //dataSourcel //dataSourcel //dataTablel //otherEntityl//citationl //addressl //conferenceLocationl //partyl//originatorl //creatorl//contactl //publisherl//editorl //recipientl//performerl //institutionl //metadataProviderl //associatedPartyl //personnell//physicall //connectionDefinitionl //distributionl //researchProjectl //spatialRasterl //spatialReferencel //spatialReferencel //spatialVectorl //storedProcedurel //viewl//protocoll //additionalMetadata:	Resolve issues with IDs and references.
-------------------------	------------	--	---	---	---	---

example		Intentional error	oXygen	XML specific tests: Passed.  Document is XML-schema valid. There were no XML errors found in your document.  KNB	Error in xml document. This EML document is not valid because the id FLS occurs more than once. IDs must be unique.  Metacat 2.0.0	qualityChecker
EML doc	docТуре	& Category	red squiggles	emlparser	errorMessage	suggestion
knb-test-fls.7.1	eml	This one fails the parser: missing ID.  Category (3) parser errors with references	Error not detected because parser errors are not validation errors.	EML specific tests: Failed.  The following errors were found: Error processing keyrefs: //references: Error in xml document. This EML instance is invalid because referenced id Ficticious_LTER_Site does not exist in the given keys.  XML specific tests: Passed.  Document is XML-schema valid. There were no XML errors found in your document.	invalid because	Resolve issues with IDs and references.
example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test- fls.8.1	eml 2.1.0	This one is missing a customUnit.  Category (3) parser errors with references	Error not detected because parser errors are not validation errors.	EML specific tests: Failed.  The following errors were found: Error processing	Error processing keyrefs: //unit /customUnit : Error in xml document. This EML instance is	Resolve issues with IDs and references.

				keyrefs: //unit/customUnit: Error in xml document. This EML instance is invalid because referenced id thingBob does not exist in the given keys.  XML specific tests: Passed.  Document is XML-schema valid. There were no XML errors found in your document.	invalid because referenced id thingBob does not exist in the given keys.	
example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test-fls.9.1	eml	An empty field: creator/organizationName Category (2) eml not schema-valid	E [Xerces] cvc-type.3.1.3: The value " of element 'organizationName' is not valid.	EML specific tests: Passed.  The tests which are specific to EML, including validation that IDs are present and properly referenced, have passed.  XML specific tests: Failed  The following errors were found: cvc-type.3.1.3: The value " of element 'organizationName' is not valid.	cvc-pattern-valid: Value " is not facet-valid with respect to pattern '[\s]* [\S][\s\S]*' for type 'NonEmptyStringType'.	Make this doc valid.
example EML doc	docТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion

	2.0.1	An empty field: creator/organizationName	(see note) Not detected in eml 2.0.1.	EML specific tests: Passed.  The tests which are specific to EML, including validation that IDs are present and properly referenced, have passed.  XML specific tests: Passed.  Document is XML-schema valid. There were no XML errors found in your document.		(Just the pkgId and eml version messages) This is an example of why eml 2.1.0 is needed: to find empty fields
example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test- fls.11.1		This one validates and parses ok.	(see note)no error	EML specific tests: Passed.  The tests which are specific to EML, including validation that IDs are present and properly referenced, have passed.  XML specific tests: Passed.  Document is XML-schema valid. There were no XML errors found in your document.	success	Use an approved namespace. (eml 2.0.1 is not approved for pasta)
example EML doc	docТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion

	2.0.1	Miss-formated dates	(see note) Not detected in eml 2.0.1.	EML specific tests: Passed.  The tests which are specific to EML, including validation that IDs are present and properly referenced, have passed.  XML specific tests: Passed.  Document is XML-schema valid. There were no XML errors found in your document.	success (Not detected in eml 2.0.1)	(Just the pkgId and eml version messages) Date format not checked in eml 2.0.1
example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test-fls.13.1	eml 2.1.0	Miss-formated dates	The date error is not detected. Shouldn't that be detected?	EML specific tests: Passed.  The tests which are specific to EML, including validation that IDs are present and properly referenced, have passed.  XML specific tests: Passed.  Document is XML-schema valid. There were no XML errors found in your document.	success	(Just the pkgId and eml version messages) Date format not checked in eml 2.0.1
example EML doc	docТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion

knb-test-fls.14.1	eml 2.1.0	less-than sign in the methods.	of elements must consist of well-formed	EML specific tests: Failed.  The following errors were found: Error running xpath expression: //dateTimeDomainl //nonNumericDomainl //nonNumericDomainl //nonNumericDomainl //nonNumericDomainl //nonNumericDomainl //nonNumericDomainl //nonNumericDomainl //nonNumericDomainl //nonNumericDomainl //coessl //datributeListl//constraintl //coveragel//temporalCoveragel //geographicCoveragel //dataSourcel //geographicCoveragel //dataSourcel //dataTablel//otherEntityl //citationl//addressl //conferenceLocationl//partyl //originatorl//creatorl//contactl //publisherl//editorl//recipientl //performerl//institutionl //metadataProviderl //associatedPartyl//personnell //physicall //connectionDefinitionl //distributionl//researchProjectl //spotialReferencel //spatialReferencel //spatialReferencel //spatialReferencel //spatialVectorl //storedProcedurel//viewl //protocoll//additionalMetadata : The content of elements must consist of well-formed character data or markup.  XML specific tests: Failed	Error running xpath expression: //dateTimeDomainl //nonNumericDomainl //accessl//attributeListl //constraintl//coveragel //temporalCoveragel //temporalCoveragel //taxonomicCoveragel //dataSourcel //dataTablel //otherEntityl//citationl //addressl //conferenceLocationl //partyl//originatorl //creatorl//contactl //publisherl//editorl //recipientl//performerl //institutionl //metadataProviderl //associatedPartyl //personnell//physicall //connectionDefinitionl //distributionl //researchProjectl //projectl //softwarel //spatialRasterl //spatialReferencel //spatialVectorl //storedProcedurel //viewl//protocoll //additionalMetadata:	Unable to parse packageId from document. The document may not be well-formed XML.
-------------------	--------------	--------------------------------	---	---	---	---

				The following errors were found: The content of elements must consist of well-formed character data or markup.	The content of elements must consist of well-formed character data or markup.	
example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
<u>knb-test-fls.15.1</u>	eml 2.1.0	This one has an html tag in the abstract so it is invalid.  Category (2) eml not schema-valid	E [Xerces] cvc-complex- type.2.4.a: Invalid content was found starting with element 'br'. One of '{itemizedlist, orderedlist, emphasis, subscript, superscript, literalLayout, ulink}' is expected. Oxygen points to line 70.	EML specific tests: Passed.  The tests which are specific to EML, including validation that IDs are present and properly referenced, have passed.  XML specific tests: Failed  The following errors were found: cvc-complex-type.2.4.a: Invalid content starting with element 'br'. The content must match '((((((("":itemizedlist)  ("":orderedlist))  ("":emphasis))  ("":subscript))  ("":superscript))  ("":literalLayout))  ("":ulink) {0-1}) {0-UNBOUNDED}'.	cvc-complex- type.2.4.a: Invalid content was found starting with element 'br'. One of '{itemizedlist, orderedlist, emphasis, subscript, superscript, literalLayout, ulink}' is expected.	Make this doc valid.
example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test-fls.16.1	eml	This one has a weird Microsoft Word encoding in the abstract.	F [Xerces] The content of elements must consist of well-formed character data or	EML specific tests: Failed.  The following errors were	Error running xpath expression: //dateTimeDomainl //nonNumericDomainl	Make this doc valid.

			found: Error running xpath expression: //dateTimeDomainl	//numericDomainl //accessl//attributeListl
			//nonNumericDomainl	//constraintl//coveragel
				//temporalCoveragel
				//geographicCoveragel
			//coveragel//temporalCoveragel	
			88	/datasetl/eml/datasetl
				//dataSourcel
				//dataTablel
			<b>-</b>	//otherEntityl//citationl
			// Citation // address:	//addressl
			1 3	//conferenceLocationl
			$\mathcal{S}$	//partyl//originatorl
			· F · · ·	//creatorl//contactl
			1	//publisherl//editorl
				//recipientl//performerl
			······································	//institutionl
	Category (2) eml not	markup. Oxygen points	, physican	//metadataProviderl
		to line 70.	//connectionDefinitionl	//associatedPartyl
			3	//personnell//physicall
			. FJ	//connectionDefinition
			r	//distributionl
			1	//researchProjectl
			1	//projectl
				//relatedProjectl
			, Protocolii, accertionali, rotace	//softwarel
				//spatialRasterl
				//spatialReferencel
				//spatialVectorl
				//storedProcedurel
				//viewl//protocoll
				//additionalMetadata :
				The content of
			found: The content of elements	elements must consist
			must consist of well-formed	of well-formed
			character data or markup.	character data or

					markup.	
example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test-fls.17.1	eml 2.1.0	title is just a space Category (2) eml not schema-valid	valid: Value ' ' is not	referenced, have passed.  XML specific tests: Failed	cvc-pattern-valid: Value ' ' is not facet-valid with respect to pattern '[\s]* [\S][\s\S]*' for type 'NonEmptyStringType'.	Make this doc valid.
example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test- fls.18.1	eml 2.0.1	lititle is illet a shace	(see note) Not detected in eml 2.0.1.	ireferenced have nassed		(Just the pkgId and eml version messages) Title with just a space not checked in eml 2.0.1

example EML doc	dосТуре	Intentional error & Category	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test-fls.19.1	eml 2.1.0	This one has an abstract/section with no para.  Category (2) eml not schema-valid	E [Xerces] cvc-complex- type.2.4.b: The content of element 'section' is not complete. One of '{title, para, section}' is expected. Oxygen points to line 70.	referenced, have passed.  XML specific tests: Failed	cvc-complex-type.2.3: Element 'section' cannot have character [children], because the type's content type is element-only.	Make this doc valid.
example EML doc	dосТуре	Intentional error	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion

# **Example Metacat Harvest Error Messages**

These errors are related to the harvest process, not to eml itself.

example EML doc	docТуре	Intentional error	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion
knb-test- fls.20.1	eml 2.1.0	harvest list specifies docType eml 2.0.1 but eml is 2.1.0	no error	na	na	na

		Category (4) harvest issues				
	eml 2.0.1	harvest list specifies docType eml 2.1.0 but eml is 2.0.1 Category (4) harvest issues		na	na	na
knb-test- fls.22.1	eml 2.1.0	Non-existent eml file: expect File Not Found error.  Category (4) harvest issues	na	Cant run a non-existent file thru the knb parser, so no message.	Error getting EML	URL: http://mcr-dev.lternet.edu /~gastilbuhl/metacat/FLS/eml/knb-test- fls.22.1.xml; IOException: http://mcr- dev.lternet.edu/~gastilbuhl/metacat /FLS/eml/knb-test-fls.22.1.xml
	eml 2.1.0	Duplicate harvest: trying to re-insert same revsion. Category (4) harvest issues	no error	na	Attempting to update knb-test-fls.1 to revision 1. Metacat has document revision 1. Metacat already has this EML document	no error in evaluation mode

### Special notes on eml 2.0.1

The eml 2.0.1 schema itself has a problem. There is a <u>patched version</u> of that schema which fixes the problem by wrapping the contents of additionalMetadata in a metadata tag. If the .xsd in the root element points to the offical version of eml 2.0.1, not the patched version, then oXygen cannot validate the eml doc. Instead, it will give the below message:

E [Xerces] cos-nonambig: describes and WC[##any] (or elements from their substitution group) violate "Unique Particle Attribution". During validation against this schema, ambiguity would be created for those two particles.

## Examples determine Error Messages, not vice versa

There can be more than one cause for the same error message. A given error message, can be caused more than one way. For example, in #4, that same message could be the result of the entityDescription appearing above the entityName rather than the entityName actually being missing. This is another reason why looking at validation with an xml editor is going to provide more specific information. It may be the same message but the red squiggle will happen on a different line.

Another way of saying this is that a given error in the eml will always produce the same message but a given message may be caused by more than one error in the eml.

#### **Details**

These tests were run on June 8-10, 2012 using Metacat 2.0.0 interactively on a dataone development server and via harvest on an LNO development server. Pleae try to avoid submiting test docs or eml with known errors to a production metacat.