



HRA – Visual Resources | Tutorial

# User Manual for VRA Core 4 Transform Tool

(Version 1.0.0)

Version: 26.02.2016 | Author: MA

# Contents

1.	Introd	duction and project history	3	
2.	Data preparation			
	2.1.	Predefined headers	4	
	2.2.	Repeated values	4	
	2.3.	Display values	5	
	2.4.	Controlled data	6	
3.	Introducing the template			
	3.1.	Downloading the template	7	
	3.2.	The template explained	7	
	3.3.	Creating a .csv file	9	
4.	Accessing the transform tool online			
	4.1.	The basic user interface	10	
5.	Uploa	iding a file	11	
6.	Generating VRA Core 4 XML			
	6.1.	Selecting a mapping	13	
	6.2.	Selecting a transformation	14	
	6.3.	Defining the records to be processed	14	
	6.4.	Generating XML		
	6.5.	Allowing pop-ups	17	
7.	The ir	nterface in Advanced Mode		
	7.1.	Applying XSLs	18	
	7.2.	Validating the XML	20	
8.	If validation fails			
	8.1.	Example 1 (element type):	22	
	8.2.	Example 2 (element type):		
	8.3.	Example 3 (date pattern):		
	8.4.	Example 4 (work/image ID):		
	8.5.	Example 5 (output empty):		
9.		rating RDF XML		
	Processing works with multiple image records			
	Contact			
	Appendix: Elements, displays and repetitions			
	Appendix: Columns Full Template			

# 1. Introduction and project history

The VRA Core 4 XML Transform Tool converts descriptive image metadata from flat tables (.csv) to structured VRA Core 4 XML. It makes use of a template with predefined headers. Users may work with the tool in a simple or advanced mode. The transform tool performs XML validation and provides feedback in case of errors.

This manual covers the whole workflow of using the *VRA Core 4 XML Transform Tool* including descriptions of all its features:

- preparation of data
- uploading data
- adjusting usage settings
- data transformation
- data validation
- handling validation feedback
- downloading the XML files

It also covers the (still experimental) transformation to RDF XML and includes a number of appendices for further reference.

Development of the *VRA Core 4 XML Transform Tool* was based on an initiative by Susan Jane Williams and resulted in a successful application by Matthias Arnold for a <u>Project Grant</u> funded by the <u>Visual Resources Association Foundation</u> (VRAF) in fall 2014. Additional support was given by the <u>Heidelberg Research Architecture</u> (HRA) at the <u>Cluster of Excellence "Asia and Europe in a Global Context"</u>, Heidelberg University.

The template csv files and the mappings to XML were developed by Susan Jane Williams, Greg Reser, and Matthias Arnold. Implementation and development of the tool was coordinated by the Heidelberg Research Architecture, programming was conducted by HRA's senior developer Matthias Guth.

The tool was released as version 1.0 in February 2016. Although we put a lot energy in testing there may still be some hidden bugs or problems. If you find anything strange, or wish to provide us feedback on issues, bugs or problems encountered, or if you want to share your ideas on how to further improve the tool please do not hesitate to write us. Any feedback is very much appreciated and may be directed to the Matthias Arnold at <a href="matthias.arnold@uni-hd.de">matthias.arnold@uni-hd.de</a> (for more information see below, section "Contact").

The VRA Core 4 XML Transform Tool is developed as Open Source, its source code can be found at: https://github.com/exc-asia-and-europe/csv2xml

Note: This user manual will be updated with every new feature or functionality. The latest version can always be found on GitHub:

https://github.com/exc-asia-and-europe/csv2xml/tree/master/doc.

# 2. Data preparation

The VRA Core 4 XML Transform Tool uses **predefined headers** to identify data in the template in order to map the values to the correct elements (or attributes) in XML. These headers can be arranged in templates (for example in .xlsx or .ods format tables).

Any descriptive image metadata can be used for transformation, be it an export from your institution's Digital Asset Management System, metadata embedded in images, or a personal image database. You only need to do two things for preparation:

- 1. Use the predefined template headers
- 2. Export data to comma separated value (.csv) file

#### 2.1. Predefined headers

The template offers 231 columns with fixed headers for your data. To be able to transform your .csv data into XML you have to use these headers.

These include references to your local system, like "IMAGE\_Filename" or "IMAGE\_LocalRepoName". In the list a number of elements are repeated to allow the transformation of multiple values, for example it offers up to three agents and up to eight subjects for the WORK record. The template also includes many columns for references to controlled vocabularies, like the name and Ref-ID of a subject in a vocabulary, and data types controlled by the VRA Core 4 schema.

For example, each of the eight subjects of a WORK record can be expressed in four columns:

WORK\_SubjectType1 WORK\_SubjectVocab1 WORK\_SubjectRefid1

The complete list of headers can be found in "Appendix: Columns Full Template" below.

#### **IMPORTANT:**

You do not need to fill each column with data. The sequence of columns (headers) may be changed and you can even delete columns you will not need.

It is, however, essential **not to change the headers** themselves, i.e. not their "text" or "values".

# 2.2. Repeated values

For some of the eighteen elementSet's of VRA Core 4 XML for WORK and IMAGE records elements can be repeated. These values usually are combined with other information, like references to authority files.

**WORK** 

Agent: 3
CulturalContext: 2
Date: 2
Location: 2
Material: 4

Measurements: 8 (in 2 sets)

Relation: 2 (plus 1 work-image)

Style/Period: 4
Subject: 8
Technique: 4

Title: 2 (preferred/alternative)

Worktype: 3

**IMAGE** 

Subject: 3

An expanded list of repeated elements can be found below, in "Appendix: Elements, displays and repetitions."

# 2.3. Display values

The template is designed to include data for each of the eighteen elementSet's of VRA Core 4 XML for WORK and most of the IMAGE records. For a large number of elementsSets structured (or "qualified") information can be transformed.

All elementSets may have a display element, with the exception of the following four IMAGE elementSets: CulturalContext, StateEdition, StylePeriod, and Textref. These are NOT included in the template.

For an overview of the elementSets, including information about which only contains display values and which also includes other elements and attributes please refer to "Appendix: Elements, displays and repetitions" below.

#### 2.3.1. Auto-filling display values

The VRA Core 4 XML Transform Tool allows transforming both, structured data for sub-elements and attributes, as well as unstructured data for display values.

If users do NOT have <display> data in their .csv file, VRA Core 4 XML Transform Tool will automatically fill the content of the <display> element based on available data from the structured information.

For example, if the template contains

WORK AgentDisplay [empty]

WORK\_Agent1Name Michelangelo Buonarroti

WORK\_Agent1Role painter WORK\_Agent1Attribution school of

The following the display element is created:

<display>school of Michelangelo Buonarroti (painter)</display>

However, if users do provide data for the display element (in the example "WORK\_AgentDisplay") these values will be used and NOT overwritten or changed.

## 2.4. Controlled data

VRA Core 4.0 in the restricted version uses controlled type lists and date formats.

#### 2.4.1. Data type values

For the **type values** please refer to the <u>VRA Core 4.0 Restricted Schema Type Values</u> document.

# 2.4.2. Date values

For date values there is no individual documentation file. <u>Vra-strict.xsd</u> states for dateValueType:

"...Defines a date which follows the ISO 8601 date format, and allows right truncation. [...] In brief, the following formats are allowed:

present	
2006	(2006)
2006-12	(December, 2006)
2006-12-31	(31 December 2006)
-44	(44 BCE)
-44-03	(March, 44 BCE)
-44-03-15	(15 March, 44 BCE)
-10000000	(10 Million Years Ago)
-100000000000	(100 Billion Years Ago)"

Source: http://loc.gov/standards/vracore/vra-strict.xsd, last accessed December 12, 2015

#### 2.4.3. Controlled data in the Transform tool

#### **IMPORTANT:**

To be able to convert your data into valid VRA Core 4 XML these type values must be used and date format rules must be followed.

The transform tool will assist you as far as possible.

For example, in <date> the sub-element <latestDate> must not be empty. But if you only entered <earliestDate> the tool will automatically fill <latestDate> with the <earliestDate> value.

If you accidentally mistyped a controlled type value, the validation will point you to this error in the validation results (see chapter "Validation errors" below).

# 3. Introducing the template

## 3.1. Downloading the template

The full template for VRA Core 4 XML Transform Tool is available for download at

https://github.com/exc-asia-and-europe/csv2xml/tree/master/doc

in these formats:

- Comma-separated values text file (.csv)
- Excel workbook (.xslx)
- OpenDocument spreadsheet (.ods)

# 3.2. The template explained

The template consists of three worksheets.

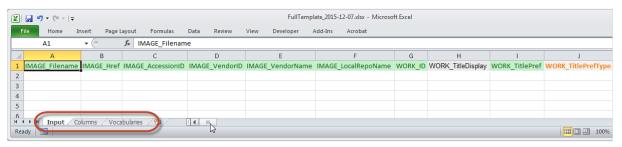


Figure 1: Full template - Input worksheet (in Excel)

The "Input" worksheet contains all the columns with their respective header in horizontal layout (colour coded), so that it can be used for editing data.

The "Columns" worksheet contains two columns. In the first, all headers are arranged vertically and colour coded, while in the second some hints to the content of the data are given.

The colour codes are included to provide visual feedback in case different types of data are expected for a "field".

*Display values,* for which the tool will auto-fill content if no data was provided for display are shown in black on grey background.

WORK\_TitleDisplay

Type values are shown in orange on light grey. Allowed values are provided in the second column.

IMAGE\_RightsType copyrighted, publicDomain, undetermined, other

References to external vocabularies are shown in dark blue on orange background.

WORK\_Agent1NameVocab

*Special content* relevant for validation is shown in black on red background, together with the validation rules. This is used for dates.

	-		-	-			_		
WORK_EarliestDate1	VALIDATION <	xsd:patter	n value	="present (-)	)*[0-9]{1	,12}(-[	0-9]{2}(-[0-9	9]{2})*)	*"/>
WORK_LatestDate1	VALIDATION <	xsd:patter	n value	="present (-)	)*[0-9]{1	,12}(-[	0-9]{2}(-[0-9	9]{2})*)	*"/>

The "normal" columns are shown in dark green on lighter green.

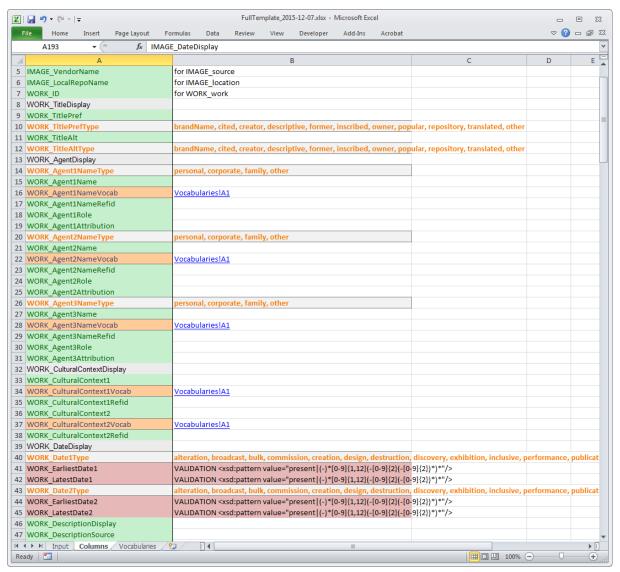


Figure 2: Full template - Columns workbook (in Excel)

The "Vocabularies" worksheet contains a reference list of external authorities or controlled vocabularies. It is by no means comprehensive. The values of the "vocab" attribute are not controlled by the VRA Core 4 schema, but it is strongly recommended to make sure the content is consistent within your data set. This is especially of interest if you intend to transform your data to RDF and use it as Linked (Open) Data.

	А	В	С	D	Е	F		
1	AAT	http://vocab						
2	TGN	http://vocab	http://vocab.getty.edu/tgn/					
3	ULAN	http://vocab	http://vocab.getty.edu/ulan/					
4	CONA	http://vocab						
5	LCNAF	http://id.loc.gov/authorities/names/						
6	LCSH	http://id.loc.gov/authorities/subjects/						
7	LCGFT	http://id.loc	http://id.loc.gov/authorities/genreForms/					
8	TGM	M http://id.loc.gov/vocabulary/graphicMaterials/						
9	VIAF	http://viaf.o	rg/viaf/					
10	ICONCLASS	http://iconcl	lass.org/					
11								
12								
14 →	I ← ► ► I Input / Columns Vocabularies							

Figure 3: Full template - Vocabularies workbook (in Excel)

# 3.3. Creating a .csv file

Export records from your local system directly into .csv files if you can include the template headers.

Alternatively, use your preferred spreadsheet software (e.g. Microsoft Excel, OpenOffice Calc, or LibreOffice Calc) as intermediary to adjust the headers and export from there into .csv files. In the doc folder on GitHub an Excel (.xslx) version of the template is available.

#### **IMPORTANT:**

**Only .csv files can be transformed** so make sure you convert spreadsheets (e.g. from Excel or Calc) to .csv text files before transformation.

Use UTF-8 as encoding to conserve possible special characters in your data.

For a guide on how to convert Microsoft Excel files into comma separated value (.csv) files see the file "Converting .xslx file to .csv using OpenOffice Calc" available in the documentation folder on GitHub.

[To be covered:

Other data sources: EMWG export-import tool]

# 4. Accessing the transform tool online

At the moment, the *VRA Core 4 XML Transform Tool* can be accessed online for testing at <a href="http://kjc-ws2.kjc.uni-heidelberg.de:8081/exist/apps/csv2xml/index.xq">http://kjc-ws2.kjc.uni-heidelberg.de:8081/exist/apps/csv2xml/index.xq</a>

Please note this is a server dedicated to testing software and developments.

#### 4.1. The basic user interface

The VRA Core 4 XML Transform Tool interface offers a Simple and an Advanced Mode.

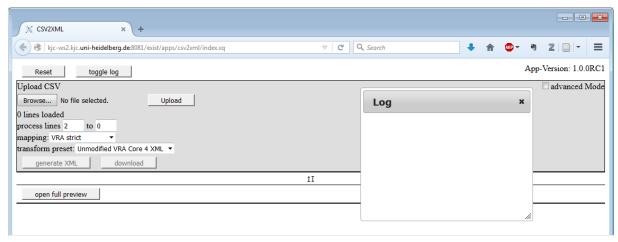


Figure 4: User interface - Simple mode.

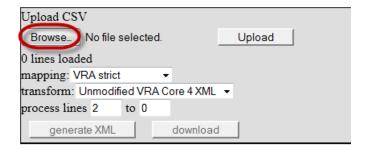
In simple mode the XML generation and validation are processed in one step finishing with the optional download of the XML files.

The advanced interface provides additional functionalities for the advanced user including the possibility to apply additional transformations.

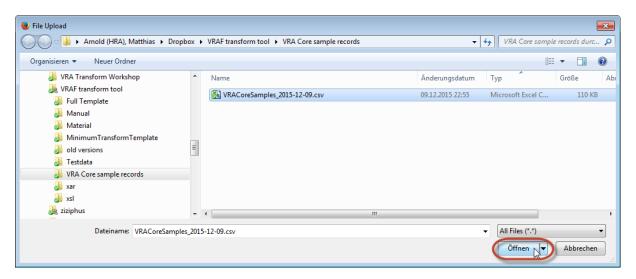
Both interfaces offer a "Log" window which is visible by default. While being used, the tool displays messages here, for example which record is currently processed. It can be switched off/on using the "toggle lock" button.

# 5. Uploading a file

To transform data you need to upload it first. Click the "Browse" button:

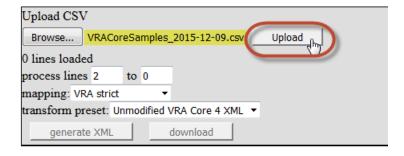


Select the .csv file on your computer and click "Open".



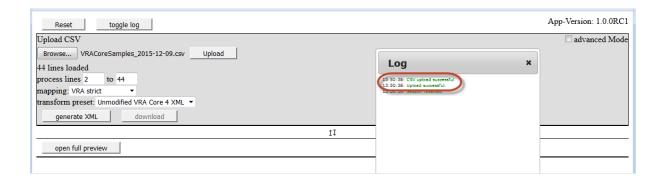
The name of the .csv file will be displayed in the interface. In this example we use "VRACoreSamples\_ 2015-12-09.csv" which contains data from the cataloguing examples on the VRA Core support pages<sup>1</sup>.

Now click on "Upload".



Once the upload is finished you will see an message "CSV upload successful" in the log window.

<sup>&</sup>lt;sup>1</sup> http://core.vraweb.org/vracore\_examples.html



# 6. Generating VRA Core 4 XML

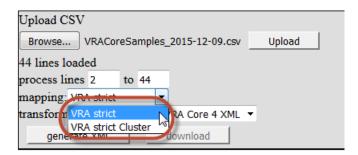
After uploading a .csv file to the *VRA Core 4 XML Transform Tool* you can just hit the "generate XML" button. This will process all records into VRA Core 4 XML using the default settings (transform all records to VRA strict without any modifications).

However, even in Simple Mode you can manipulate how the XML is generated. You have three options:

- Select another mapping
- Select another transform
- Define the range of records to be processed

# 6.1. Selecting a mapping

The VRA Core 4 XML Transform Tool supports generating XML based on different mappings. For VRA Core, two variants are available: "VRA strict" and "VRA strict Cluster".



#### 6.1.1. VRA strict

VRA strict is the **default mapping**. It is based on the VRA Core 4 restricted version (<a href="http://www.loc.gov/standards/vracore/vra-strict.xsd">http://www.loc.gov/standards/vracore/vra-strict.xsd</a>). It uses controlled values for type attributes and specifies date formats.

For more information please refer to http://www.loc.gov/standards/vracore/schemas.html.

## 6.1.2. VRA strict Cluster

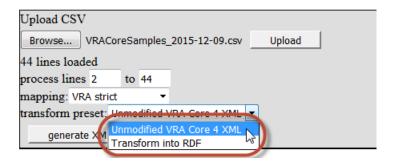
The VRA strict Cluster mapping is an expansion of the restricted VRA Core 4 version developed at the Cluster of Excellence "Asia and Europe in a Global Context", University of Heidelberg. It contains additional elements and attributes, for example, attributes for multilingual data, role attributes for agents, and an element for geo-coordinates (<a href="http://cluster-schemas.uni-hd.de/vra-strictCluster.xsd">http://cluster-schemas.uni-hd.de/vra-strictCluster.xsd</a>).

For more information please refer to the document "VRA Core – Extensions".

# 6.2. Selecting a transformation

The tool supports different transformations.

The current version of *VRA Core 4 XML Transform Tool* offers two built-in transformations. Additional .xsl files may to be added in the source code. In a future version, this may be changed (e.g. upload transformation stylesheets in the user interface).



Currently, two transformations are implemented:

- Unmodified VRA Core 4 XML
- Transform into RDF (explanations see below)

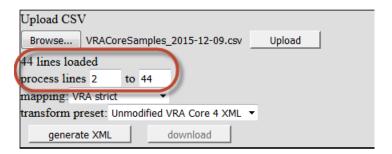
To transform the .csv data into VRA Core 4 XML choose "Unmodified VRA Core 4 XML".

# 6.3. Defining the records to be processed

By default, all records listed in the uploaded .csv file will be processed.

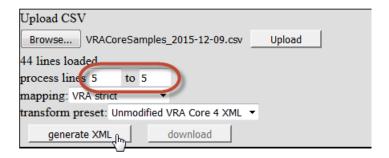
## 6.3.1. Calculating records for processing

Within the .csv file a record equals a line. In the template, the first record (first line) contains the column headers. Therefore, the *VRA Core 4 XML Transform Tool* will by default process data beginning with line #2 (i.e. record #1).



# 6.3.2. **Defining a range of records**

The VRA Core 4 XML Transform Tool also allows you to define the range of records to be processed. This is very helpful if you want to check a smaller range of records or even individual ones, look at their xml and test if they validate.

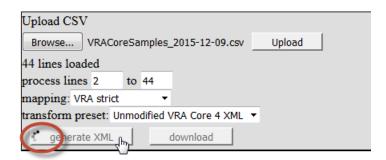


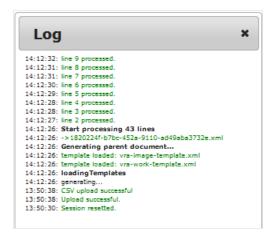
In this example, only line 5 (i.e. record #4) will be processed.

# 6.4. Generating XML

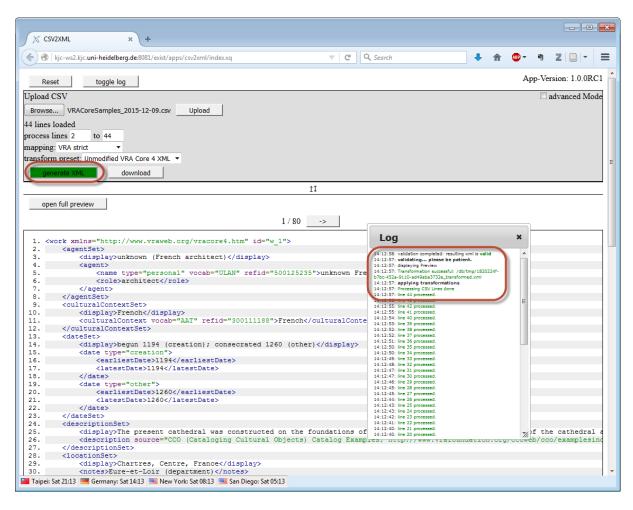
If all settings are made, click the "generate XML" button.

The tool will start generating the XML. You can monitor the progress in the "Log" window.



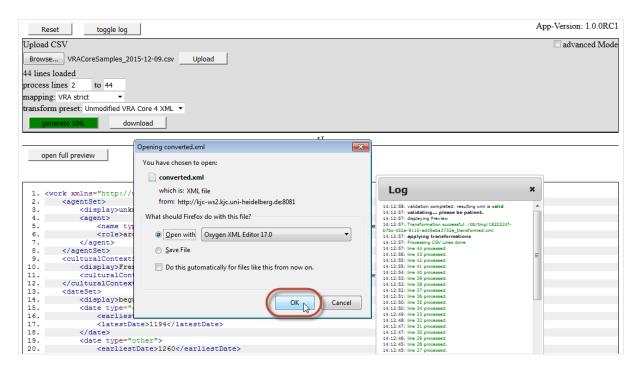


After generating XML a preview of the code is displayed while the tool automatically runs a validation. When all is finished the "generate XML" button changes its colour.



You can browse the preview by clicking the arrow button. The preview shows generated WORK and IMAGE code in XML on separate pages.

You can also directly use the "download" link which will show the download dialog "Opening converted.xml". Note that the default file name is "converted.xml".



You can save the file on a disk or open it in your preferred XML editor.

# 6.5. Allowing pop-ups

You can also click on the "open full preview" button to view the whole XML in your browser. This will open a new tab. Depending on your local browser settings, opening the new tab may trigger a popup blocking message. In Firefox it may look like this:



Choose "Allow pop-ups" from the "Options" menu.



# 7. The interface in Advanced Mode

In the Advanced mode it is possible to

- apply XSL transforms
- choose a schema against which you wish to validate your XML file
- upload your own schema file

It also separates XML generation from validation.



Figure 5: User interface - Advanced mode.

# 7.1. Applying XSLs

During the default generation of "Unmodified VRA Core 4 XML" a transformation script "cleanup-vra.xsl" will be automatically applied. It removes empty attributes and elements and keeps the resulting XML file slim and clean. It is possible to deselect it here. You should note, however, that without running "cleanup-vra.xsl" the resulting XML code may not validate.

If you choose "Transform into RDF" as transform pre-set the "VRA2RDF.xsl" will be loaded and applied by default. For more information about RDF transformation see chapter "Generating RDF" below.

It is possible to add other .xsl files in the source code. In a future version of *VRA Core 4 XML Transform Tool* this may be changed.



Figure 6: Advanced Mode - XML generation without additional transformation

Figure 7: Preview of XML without clean-up

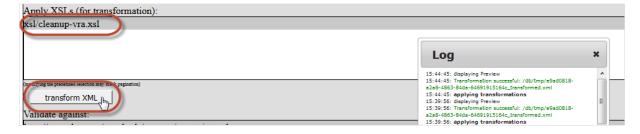


Figure 8: Starting the clean-up manually

```
1 / 80 ->
 1. <work xmlns="http://www.vraweb.org/vracore4.htm" id="w 1">
         <agentSet>
 <name type="personal" vocab="ULAN" refid="500125235">unknown French</name>
<role>architect</role>
 6.
7.
              </agent>
12. </culturalContextSet>
12. </aditable/
13. <adetsSet>
14. <adisplay>begun 1194 (creation); consecrated 1260 (other)</display>
15. <adet type="creation">
16. <aerliestDate>1194</earliestDate>
                   <latestDate>1194</latestDate>
18. </date>
19. <date type="other">
20. <earliestDate>1260</earliestDate>
                   <latestDate>1260</latestDate>
            </date>
24. <descriptionSet>
25. <display>The p
              <display>The present cathedral was constructed on the foundations of the earlier church; the oldest parts of the cathedral
<description source="CCO (Cataloging Cultural Objects) Catalog Examples; http://www.vrafoundation.org/ccoweb/cco/examplesin</p>
26.
27.
         </descriptionSet>
```

Figure 9: Preview of XML after clean-up

# 7.2. Validating the XML

Transforming data from .csv to XML does not *per se* guarantee a valid XML result. This is because the restricted VRA Core schema defines a number of fixed values. If the values in the .csv document do not match the values expected by the schema, validation will fail. We therefore strongly recommend validating the generated data in advanced mode.

## 7.2.1. Choosing a schema for validation

In the "Validate against" box you can choose a schema for validation. Depending on the selected mapping (see above) the respective default validation schema is automatically set.

```
Validate against:

(http://www.loc.gov/standards/vracore/vra-strict.xsd
http://www.loc.gov/standards/vracore/vra.xsd

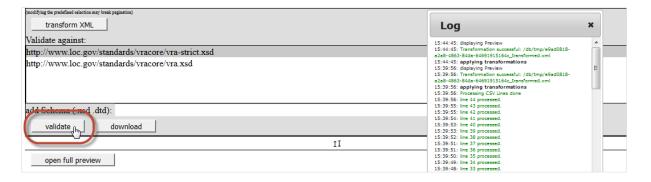
add Schema (.xsd .dtd):
```

For example, if you selected "VRA strict" as mapping, the corresponding schema <a href="http://www.loc.gov/standards/vracore/vra-strict.xsd">http://www.loc.gov/standards/vracore/vra-strict.xsd</a> will be set.

It is possible to choose different validation schemas. In a future version of the transform tool users may also upload their own schema.

#### 7.2.2. Starting the validation

To start the validation process, click "validate".



## 7.2.3. Getting the validation result

Once the validation is finished, a new browser tab will be opened with the validation result. In addition, the button's colour will change:



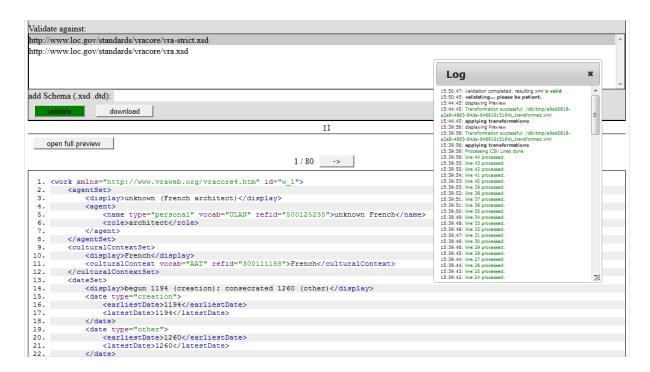


Figure 10: The tool generated valid VRA Core 4 XML.

# 8. If validation fails

VRA Core 4 in the restricted version uses controlled type lists and date formats. If your data does not match the prescribed rules, your XML will not validate.

In case of problems with the validation, the error messages will be displayed in a separate tab. The messages will indicate both line and column of the respective generated XML document. This is displayed in the interface.

Error messages usually come in pairs. The first line will name the invalid value based on an enumeration in the schema (e.g. type values defined by vra-strict). The second line will indicate the parent element.

```
cepont>result>
caschittp://kjc=sv016.kjc.uni=heidelberg.de:5080/exist/apps/tamboti/resources/schemas/vra=strictCluster.xsdc/xsd>
cepont>
cepont>
cascaus>invalide/status>
cascaus>
cascaus
```

# 8.1. Example 1 (element type):

Error message (screenshot):

```
cduration units "esen" 362 (Auration)

cassage level "Error" line="114" column="51" over-enumeration-valid: Value "illustrative" is not facet-valid with respect to enumeration '[brandName, cited, creator, descriptive, former, generalView, inscribed, chessage level="Error" line="114" column="51" over-entribute. 5: The value 'illustrative' of attribute 'type' on element 'title' is not valid with respect to its type, 'titleTypeType'. (/message)

consists of the first of the column of the value 'illustrative' of attribute 'type' on element 'title' is not valid with respect to its type, 'titleTypeType'. (/message)

consists of the column of the value 'illustrative' of attribute 'type' on element 'title' is not valid with respect to its type, 'titleTypeType'. (/message)

consists of the column of the value 'illustrative' of attribute 'type' on element 'title' is not valid with respect to its type, 'titleTypeType'. (/message)

consists of the column of the value 'illustrative' of attribute 'type' on element 'title' is not valid with respect to its type, 'titleTypeType'. (/message)
```

Error message (transcribed):

```
<message level="Error" line="134" column="52">cvc-enumeration-valid: Value
'illustrative' is not facet-valid with respect to enumeration '[brandName, cited,
creator, descriptive, former, generalView, inscribed, owner, partialView, popular,
repository, translated, other]'. It must be a value from the enumeration.</message>
<message level="Error" line="134" column="52">cvc-attribute.3: The value
'illustrative' of attribute 'type' on element 'title' is not valid with respect to
its type, 'titleTypeType'.</message>
```

The error message refers to line="134" column="52" in the XML:

The first line of the message states that "illustrative" is not valid because it is not included in the list of possible values: "brandName, cited, creator, descriptive, former, generalView, inscribed, owner, partialView, popular, repository, translated, other".

The second line adds that the error occurred within attribute "type" of element <title>, and the provided value for "type" was not valid.

#### Result:

Values of <title type=""> are controlled by the schema.

You need to delete "illustrative" and use of one of the allowed values, i.e. "brandName, cited, creator, descriptive, former, generalView, inscribed, owner, partialView, popular, repository, translated, other".

# 8.2. Example 2 (element type):

Error message (screenshot):

```
Transage laval-"Error" line="72" column="C"hero-enumeration-valid: Value 'person' is not facet-valid with respect to enumeration ('personal, corporate, family, other)', It may be a value from the enumeration ('personal accordance) of the column of the enumeration of the enumerat
```

Error message (transcribed):

```
<message level="Error" line="247" column="67">cvc-enumeration-valid: Value 'person'
is not facet-valid with respect to enumeration '[personal, corporate, family,
other]'. It must be a value from the enumeration.</message>

<message level="Error" line="247" column="67">cvc-attribute.3: The value 'person'
of attribute 'type' on element 'name' is not valid with respect to its type,
'agentNameTypeType'.</message>
```

The error message refers to line="247" column="67" in the XML:

The first line of the message states that "person" is not valid because it is not included in the list of possible values: "personal, corporate, family, other".

The second line adds that the error occurred within attribute "type" of element <name>, and the provided value for "type" was not valid.

#### Result:

Values of <name type=""> are controlled by the schema.

You need to delete "person" and use of one of the allowed values, i.e. "personal, corporate, family, other".

# 8.3. Example 3 (date pattern):

This example consists of two similar problems:

Error message (screenshot):

```
Commange level-Terre? line="510" column="50"-pre-pattern-valid; Value "01.05.18" is not facet-valid with respect to pattern "present!(-)=[0-9](1,22](-(0-9)[2])-)-) for type "dateValueType".

commange level="Errer" line="510" column="50"-pre-persen-valid; Value "01.05.18" is not facet-valid with respect to pattern "present!(-)=[0-9](1,22](-(0-9)[2])-)-) for type "dateValueType".

commange level="Errer" line="510" column="50"-pre-persen-valid; Value "01.05.18" is not facet-valid with respect to pattern "present!(-)=[0-9](2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0-9)[2](-0
```

#### Error message (transcribed):

```
<message level="Error" line="510" column="50">cvc-pattern-valid: Value '01.05.15'
is not facet-valid with respect to pattern 'present|(-)*[0-9]{1,12}(-[0-9]{2}(-[0-9]{2})*)*' for type 'dateValueType'.</message>
<message level="Error" line="510" column="50">cvc-complex-type.2.2: Element
'latestDate' must have no element [children], and the value must be valid.</message>
<message level="Error" line="513" column="53">cvc-pattern-valid: Value '2015/06' is
not facet-valid with respect to pattern 'present|(-)*[0-9]{1,12}(-[0-9]{2}(-[0-9]{2})*)*' for type 'dateValueType'.</message>
<message level="Error" line="513" column="53">cvc-complex-type.2.2: Element
'earliestDate' must have no element [children], and the value must be
valid.</message>
```

The error message refers to line="510" column="50" and line="513" column="53" in the XML:

```
506.
             <dateSet>
507.
                  <display>a predefined work date display</display>
508.
                  <date type="creation">
509.
                      <earliestDate>01.05.15</earliestDate>
510.
                      <latestDate>01.05.15</latestDate>
511.
                  </date>
512.
                  <date type="alteration">
513.
                      <earliestDate>2015/06</earliestDate>
514.
                      <latestDate>2015/06</latestDate>
515.
                  </date>
516.
              </dateSet>
```

The first lines of each message sat state that "01.05.15" and "2015/06" are not valid because they do not follow the predefined pattern for dateValues.

The second lines add that the error occurred within element <latestDate> (respective <earliestDate>), and the element must be valid.

#### Result:

Values of <earliestDate> and <latestDate> are pattern-controlled by the schema.

You need to write the dates in the correct format, i.e. "2015-05-01" or "2015-06".

# 8.4. Example 4 (work/image ID):

Error message (transcribed):

```
<message level="Error" line="239" column="18"> cvc-datatype-valid.1.2.1: '2' is not
a valid value for 'NCName'.</message>
<message level="Error" line="239" column="18"> cvc-attribute.3: The value '2' of
attribute 'id' on element 'work' is not valid with respect to its type,
'ID'.</message>
```

The error message refers to line="239" column="18" in the XML:

The first line of the message states that "2" is not a valid NCName, which stands for "Non-colonized Names" (cf. <a href="http://www.w3.org/TR/xmlschema-2/#NCName">http://www.w3.org/TR/xmlschema-2/#NCName</a>).

The second line adds that the error occurred within attribute "id" of element <work>, and the provided value was not valid.

#### Result:

Values of <work id=""> are regulated by the schema. The "VRA Core4 Element Description" explains in a note: "The XML id attribute must begin with a character. The convention used here prefixes a numeric value with w\_ for works, c\_ for collections, and i\_ for images."

You need to provide a work ID that begins with "w\_".

# 8.5. Example 5 (output empty):

If your data is processed without an error and was successfully validated, but the output does not show any of your data, you might need to check the encoding of the .csv file. We had a test case where data was by mistake encoded in "Unicode" instead of "UTF-8" and the result was this "empty" output. Note that some elements are there, but no descriptive metadata.



#### Result:

The .csv file must be encoded in "UTF-8".

See the document "Converting xslx to csv in Calc.pdf" also available in the doc folder on GitHub: <a href="https://github.com/exc-asia-and-europe/csv2xml/tree/master/doc">https://github.com/exc-asia-and-europe/csv2xml/tree/master/doc</a>.

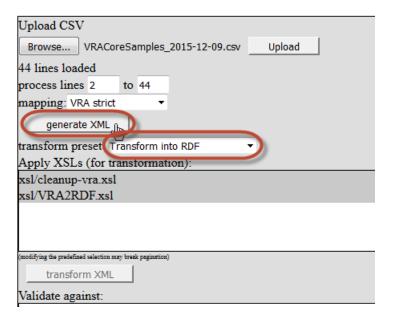
# 9. Generating RDF XML

In addition to generating VRA Core 4 XML it is possible to transform the data to RDF. At the moment, this feature is still experimental, meaning that only basic testing was performed and no output optimisation took place.

The tool makes use of the XSLT stylesheet as provided by the VRA-RDF-Project <a href="https://github.com/mixterj/VRA-RDF-Project">https://github.com/mixterj/VRA-RDF-Project</a>.

To transform your data to RDF first make sure the .csv data will validate in VRA Core.

Then go back and select "transform into RDF" in the "transform" line. Note that the entry "xsl/VRA2RDF.xsl" appears in the "Applied XSL's" box.



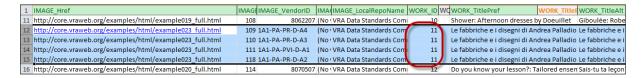
Now click "generate XML".

Once generated, the RDF XML will be displayed in the "Result" box and can be downloaded.

1/80 1. <rdf:Description xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-nsf" rdf:about="fw\_1">
2. <rdf:type rdf:resource="http://purl.org/vra/CreativeWork"></rdf:type>
3. <rdf:type rdf:resource="http://vocab.getty.edu/aat/300007501"></rdf:type>
4. <rdf:type rdf:resource="http://vocab.getty.edu/aat/300007501"></rdf:type>
5. <rdf:type rdf:resource="http://vocab.getty.edu/aat/300007466"></rdf:type></rdf:type></rdf:type rdf:resource="http://vocab.getty.edu/aat/300007466"></rdf:type></rdf:type></rdf:type rdf:resource="http://purl.org/vza/">
6. <vra:creator xmlns:vra="http://purl.org/vza/">
7. <rdf:Description rdf:about="http://purl.org/vza/">
8. <rdf:type rdf:resource="http://purl.org/vza/Agent"></rdf:type>
9. <rdf:type rdf:resource="http://purl.org/vza/Agent"></rdf:type>
9. <rdf:type rdf:resource="http://purl.org/vza/Person"></rdf:type>
9. </rdf:Description>
10. </rdf:Description>
11. </rdf:Description>
12. </rdf:Description over culturalContextSet--><!--Skipping over display--><!--Skipping over French-->
13. </rdf:Description xdf:about="http://purl.org/vza/">
14. </rdf:Description xdf:about="http://purl.org/vza/">
15. </rdf:Description>
16. </rdf:Description>
17. </rdf:Description>
18. </rdf:Description>
19. </rdf:Description < 18. 19. 20. 21. 22. 23. <vra:endDate>1194 24. </rdf:Description>
</rdf:Description>
</wra:WasCreated><!--Skipping over descriptionSet--><!--Skipping over display--><!--Skipping over The present cathedral was consequenced by the present cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundations of the earlier church cathedral was constructed on the foundation of the earlier church cathedral was constructed on the foundation of the earlier church cathedral was constructed on the foundation of the earlier church cathedral was constructed on the foundation of the earlier church cathedral was constructed on the foundation of the earlier church cathedral was constructed on the foundation of the earlier church cathedral was constructed on the foundation of the earlier church cathedral was constructed on the earlier church cathedral was constructed on the earlier church cathedral was constr 25. 26. 27. 28. 29. <rdf:Description> r:Description>
cynd:rype raf:resource="http://purl.org/vra/Place"></raf:type>
<vra:containedIn>
<raf:tpescription raf:about="http://vocab.getty.edu/tgn/7033010">
<raf:type raf:resource="http://purl.org/vra/Place"></raf:type>
<vra:name>Chartres Cathedral</vra:name> 30. 31. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. </rdf:Description>

# 10. Processing works with multiple image records

Sometimes work records can have multiple image records assigned to it. For example, in the VRA Core sample records the work in Example 23 has four images:



Therefore, the VRA Core 4 XML Transform Tool can only use one data set to create the work xml from. The decision will be made by looking at the contents of "WORK\_ID" and using just the first set (for the work).

In the example, only the information from line 12 will be used to create the WORK record.



You need to make sure that different lines containing information about the same work are identical. In addition, you need to make sure that your WORK ID's are uniqe.

In the XML output, the VRA Core 4 XML Transform Tool will not only relate the image records to the respective WORK, but it will also relate the work record to all IMAGE records.

```
<work id="w 11">
                <agentSet>
                                                                                                                                                                            [15 lines]
1303 ▶
                <culturalContextSet>
                                                                                                                                                                             [3 lines]
1307
                <dateSet>
                                                                                                                                                                              [6 lines]
1314 ▶
                                                                                                                                                                             [3 lines]
1318
                <locationSet>
1326
                                                                                                                                                                              [4 lines]
1331
                 <measurementsSet
1335
1336
                     <display>depicts Palazzo della Ragione; </display>
1337
                     <relation type="depicts">Palazzo della Ragione</relation>
1338
                     <relation type="imageIs" relids="i_7f164d5d-7b9d-4825-86d3-483e8b9d782d">Details, Palazzo della Ragione (plate I XXXI)</relation>
<relation type="imageIs" relids="i_c23f56e6-7b0a-40e3-b829-5ec574101dc1">Section, Palazzo della Ragione (plate I XXXI)</relation>
1339
                     <relation type="imageIs" relids="i_7dfcbe6b-a4bd-4ec7-8e2a-d30c0c00150f">Palazzo Vicentino; plan (plate T XXVII)</rela
                     <relation type="imageIs" relids="i_27469e50-189b-4c4a-8a57-93c6ac269552">Elevation, Palazzo della Ragione (plate T XXIX)/relation:
1341
1342
1343
                     <display>Eighteenth century; Palladian</display>
```

Figure 11: XML output of Example 23, relationSet, showing relations to all four IMAGE records

# 11. Contact

If you have problems or questions please contact Matthias Arnold at <a href="matthias.arnold@uni-hd.de"><u>matthias.arnold@uni-hd.de</u></a>.

#### **Contact details**

<u>matthias.arnold@uni-hd.de</u> or <u>arnold@asia-europe.uni-heidelberg.de</u>

Phone: +49 (0) 6221 - 54 4094 Fax: +49 (0) 6221 - 54 4012

Skype: matz-skype

#### **Further links**

Web-Portal of Heidelberg Research Architecture <a href="http://hra.uni-hd.de/">http://hra.uni-hd.de/</a>

Ziziphus VRA Core 4 XML editor:

https://github.com/exc-asia-and-europe/ziziphus Ziziphus - guidelines and help files: http://ziziphus-help.uni-hd.de

#### Office address

Karl Jaspers Centre, Room 005b

#### MediaLab

Karl Jaspers Centre, Room 005c

# 12. Appendix: Elements, displays and repetitions

## WORK

[workID]

#### **Agent** display

3x nameType - name - nameVocab - nameRefid - role

## **CulturalContext** display

2x culturalContext - vocab - refid

#### Date display

2x type – earliestDate – latestDate

## **Description** display [display = description]

descriptionSource

Inscription display [display = text]

#### **Location** display

Location notes

2x type - Name - NameVocab - NameRefID - ObjRefID - Geo - GeoVocab - GeoRefid

#### **Material** display

4x material - vocab - refid

# Measurements display

2x extent - unit - value1-4 - type1-4

## **Relation** display

2x relatedWork - type

[transform adds work-image link(s)]

**Rights** display [display = text]

**Source** display

StateEdition display

## Style/Period display

4x stylePeriod - vocab - refid

#### **Subject** display

8x subject - vocab - refid - type

#### Technique display

4x technique – vocab – refid

#### TextRef display

Title display

# titlePref, titlePrefType, titleAlt, titleAltType

#### Worktype display

3x worktype - vocab - refid

# **IMAGE**

[IMAGE\_Href (-> href); Filename (-> refid)]

## **Agent** display

1x name – nameVocab – nameRefid – role

Date display

**Description** display [display = description]

Inscription display [display = text]

#### **Location** display

[localRepoName, accessionID]

**Material** display

**Measurements** display

#### **Relation** display

[transform adds image-work link(s)]

# **Rights** display

Rights notes

1x type – holder – text

# **Source** display

[VendorName, VendorID]

1x sourceType, sourceValue

#### **Subject** display

3x subject – vocab – refid – type

# **Technique** display

1x technique – vocab – refid

[if not provided: <technique vocab="AAT" refid="300237903">digital imaging</technique>]

## **Title** display

1x type, title

## WorkType display

1x technique – vocab – refid

[if not provided: <worktype vocab="AAT" refid="300215302">digital images</worktype>]

# 13. Appendix: Columns Full Template

IMAGE\_Filename IMAGE\_Href IMAGE\_AccessionID IMAGE\_VendorID

IMAGE\_VendorName
IMAGE\_LocalRepoName

WORK ID

WORK\_TitleDisplay
WORK\_TitlePref
WORK\_TitlePrefType
WORK\_TitleAlt
WORK\_TitleAltType
WORK\_AgentDisplay
WORK\_Agent1NameType
WORK\_Agent1Name

WORK\_Agent1NameVocab WORK\_Agent1NameRefid WORK\_Agent1Role

WORK\_Agent1Attribution WORK\_Agent2NameType WORK\_Agent2Name

WORK\_Agent2NameVocab WORK\_Agent2NameRefid WORK\_Agent2Role

WORK\_Agent2Attribution WORK\_Agent3NameType WORK\_Agent3Name

WORK\_Agent3NameVocab WORK\_Agent3NameRefid

WORK\_Agent3Role

WORK\_Agent3Attribution
WORK\_CulturalContextDisplay
WORK CulturalContext1

WORK\_CulturalContext1Vocab WORK\_CulturalContext1Refid WORK\_CulturalContext2

WORK\_CulturalContext2Vocab WORK\_CulturalContext2Refid

WORK\_DateDisplay
WORK\_Date1Type
WORK\_EarliestDate1
WORK\_LatestDate1
WORK\_Date2Type
WORK\_EarliestDate2
WORK\_LatestDate2
WORK\_LatestDate2
WORK\_DescriptionDisplay
WORK\_DescriptionSource

WORK Inscription Display

WORK\_LocationDisplay WORK\_LocationNotes WORK\_Location1Type

WORK\_Location1Name
WORK\_Location1NameVocab
WORK\_Location1NameRefID

WORK\_Location1ObjRefID WORK Location1Geo

WORK\_Location1GeoVocab WORK\_Location1GeoRefid WORK\_Location2Type

WORK\_Location2Name
WORK\_Location2NameVocab
WORK\_Location2NameRefID
WORK\_Location2ObjRefID
WORK\_Location2Geo

WORK\_Location2GeoVocab WORK\_Location2GeoRefid WORK\_MaterialDisplay WORK\_Material1 WORK MaterialVocab1

WORK\_MaterialRefid1
WORK\_Material2
WORK\_MaterialVocab2
WORK\_MaterialRefid2
WORK\_Material3
WORK\_MaterialVocab3

WORK\_MaterialVocab4 WORK\_MaterialRefid4

WORK\_MaterialRefid3

WORK\_MeasurementsDisplay WORK\_Measurements1Extent WORK\_Measurements1Unit WORK\_Measurements1Value1 WORK\_Measurements1Type1 WORK\_Measurements1Value2 WORK\_Measurements1Type2 WORK\_Measurements1Value3 WORK\_Measurements1Type3 WORK\_Measurements1Value4

WORK\_Measurements1Type3
WORK\_Measurements1Value4
WORK\_Measurements1Type4
WORK\_Measurements2Extent
WORK\_Measurements2Unit
WORK\_Measurements2Value1
WORK\_Measurements2Type1
WORK\_Measurements2Value2
WORK\_Measurements2Type2

WORK Measurements2Value3

WORK\_Measurements2Type3 WORK Measurements2Value4 WORK Measurements2Type4 WORK Relation Display WORK\_RelatedWork1 WORK\_RelationType1 WORK RelatedWork2 WORK RelationType2 WORK\_RightsDisplay WORK\_SourceDisplay WORK\_StateEditionDisplay WORK StylePeriodDisplay WORK StylePeriod1 WORK StylePeriodVocab1 WORK StylePeriodRefid1 WORK\_StylePeriod2 WORK StylePeriodVocab2 WORK\_StylePeriodRefid2 WORK\_StylePeriod3 WORK\_StylePeriodVocab3 WORK StylePeriodRefid3 WORK StylePeriod4 WORK\_StylePeriodVocab4 WORK StylePeriodRefid4 WORK SubjectDisplay WORK\_Subject1 WORK\_SubjectType1 WORK\_SubjectVocab1 WORK\_SubjectRefid1 WORK\_Subject2 WORK\_SubjectType2 WORK\_SubjectVocab2 WORK SubjectRefid2 WORK\_Subject3 WORK\_SubjectType3 WORK\_SubjectVocab3 WORK SubjectRefid3 WORK\_Subject4 WORK\_SubjectType4 WORK\_SubjectVocab4 WORK\_SubjectRefid4 WORK Subject5 WORK\_SubjectType5 WORK\_SubjectVocab5 WORK SubjectRefid5 WORK Subject6 WORK SubjectType6 WORK\_SubjectVocab6 WORK\_SubjectRefid6 WORK\_Subject7

WORK\_SubjectType7

WORK\_SubjectVocab7 WORK SubjectRefid7 **WORK Subject8** WORK\_SubjectType8 WORK\_SubjectVocab8 WORK\_SubjectRefid8 WORK TechniqueDisplay WORK Technique1 WORK\_TechniqueVocab1 WORK\_TechniqueRefid1 WORK Technique2 WORK TechniqueVocab2 WORK\_TechniqueRefid2 WORK\_Technique3 WORK\_TechniqueVocab3 WORK TechniqueRefid3 WORK Technique4 WORK\_TechniqueVocab4 WORK\_TechniqueRefid4 WORK TextrefDisplay WORK TextrefName WORK TextrefNameType WORK TextrefRefid WORK\_TextrefRefidType WORK\_WorktypeDisplay WORK\_Worktype1 WORK\_WorktypeVocab1 WORK\_WorktypeRefid1 WORK\_Worktype2 WORK\_WorktypeVocab2 WORK WorktypeRefid2 WORK\_Worktype3 WORK WorktypeVocab3 WORK\_WorktypeRefid3

IMAGE\_AgentDisplay IMAGE\_Agent1Name IMAGE\_Agent1NameVocab IMAGE\_Agent1NameRefid IMAGE\_Agent1Role IMAGE\_DateDisplay IMAGE DescriptionDisplay IMAGE\_InscriptionDisplay IMAGE\_LocationDisplay IMAGE MaterialDisplay IMAGE MeasurementsDisplay IMAGE\_RelationDisplay IMAGE\_RightsDisplay IMAGE\_RightsNotes IMAGE RightsType IMAGE\_RightsHolder

IMAGE\_RightsText
IMAGE\_SourceDisplay
IMAGE\_SourceValue
IMAGE\_SourceType
IMAGE\_SubjectDisplay
IMAGE\_Subject1
IMAGE\_SubjectType1
IMAGE\_SubjectVocab1
IMAGE\_SubjectRefid1
IMAGE\_Subject2
IMAGE\_SubjectType2
IMAGE\_SubjectType2
IMAGE\_SubjectVocab2
IMAGE\_SubjectRefid2
IMAGE\_SubjectRefid2
IMAGE\_Subject3

IMAGE\_SubjectType3
IMAGE\_SubjectVocab3
IMAGE\_SubjectRefid3
IMAGE\_TechniqueDisplay
IMAGE\_Technique
IMAGE\_TechniqueVocab
IMAGE\_TechniqueRefid
IMAGE\_TitleDisplay
IMAGE\_Title
IMAGE\_TitleType
IMAGE\_WorktypeDisplay
IMAGE\_Worktype
IMAGE\_WorktypeVocab

IMAGE\_WorktypeRefid