Cluster of Excellence
ASIA AND EUROPE
IN A GLOBAL CONTEXT

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

HRA – Visual Resources | Tutorial

Manual for VRA Core 4 Transform tool (DRAFT)

Cluster of Excellence “Asia and Europe in a Global Context”
Karl Jaspers Centre for Advanced Transcultural Studies
Voßstraße 2, Building 4400, 69115 Heidelberg, Germany

Phone: +49 (0) 6221 - 54 40 06
info@asia-europe.uni-heidelberg.de
www.asia-europe.uni-heidelberg.de

Version: 30.07.20152015 | Author: M A

Contents

1. Project history

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 Project Grant funded by the Visual Resources Association Foundation (VRAF) in Fall 2014. Additional support was given by the Heidelberg Research Architecture (HRA) at the Cluster of Excellence "Asia and Europe in a Global Context", Heidelberg University.

> "Matthias Arnold, Heidelberg Research Architecture, University of Heidelberg, will use the award to support the development of the VRA Core 4 XML Transform Tool. This tool will enable any user who can supply descriptive image metadata in a standardized CSV form (comma separated values, e.g. via Excel) to create validating VRA Core 4 XML. Useful on its own, this XML also represents an important step for further transformations to other XML schemas, like RDF to support output of data as LOD (Linked Open Data). The tool is being developed in consultation with both the VRA Data Standards and the Core OC Committees, will be shared with the image management community during beta-testing for feedback, and will be open-source and freely available upon completion. Susan Jane Williams (Independent Cataloging and Consulting Services) will serve as coordinator for the mapping templates and will help create documentation, demonstrations (sample records and mappings), and further mapping help in use of the tool. We are confident that this tool will have broad value and impact within the visual resources community, and for the VRA Core 4 standard."
> "2014-2015 VRA Foundation Project Grant Winners Named", VRAF News & Events, December 19, 2014

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

Please note:
The tool is still in BETA version, currently v.0.0.4. Any feedback on issues, bugs or problems encountered is very much appreciated and should be directed to the developers (see 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

2. Data preparation

Data transformation is based on templates. Two templates are being prepared:

- Minumum template
- Full template

Both templates use predefined headers for each column to process data.

The sequence of columns (headers) may be changed, 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”. Otherwise the transformation will not work.

The full template is already available and can be downloaded directly from GitHub:

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

There is also a .xslx version of the template available. However, only .csv templates can be transformed, so make sure you convert Excel or Calc spreadsheets 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 below.

To be covered:
Schema defined data, controlled terms, etc
Saving as csv (UTF-8)
Other data sources: EMWG export-import tool
Other data sources: local repository

3. Accessing the tool online

The transform tool can be accessed online for testing at
http://kjc-ws2.kjc.uni-heidelberg.de:8650/exist/apps/csv2xml/index.xq

Please note, this is a server dedicated to testing software and developments. It should be used with caution.

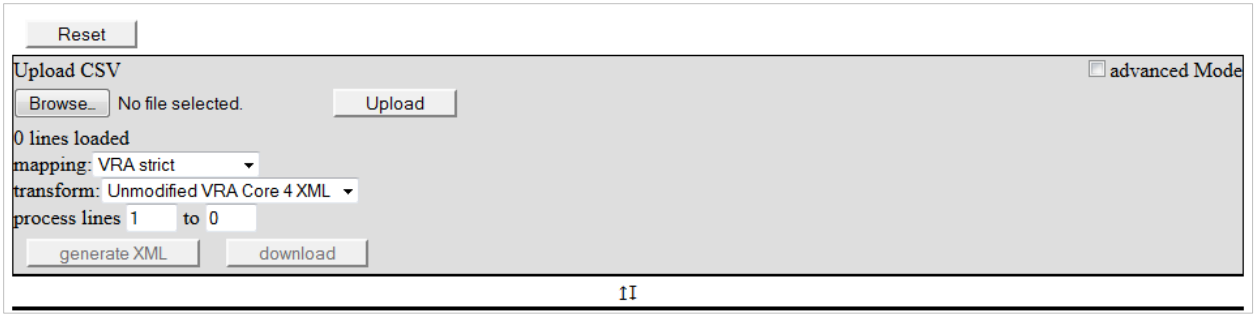


Figure 1: User interface - Simple mode.

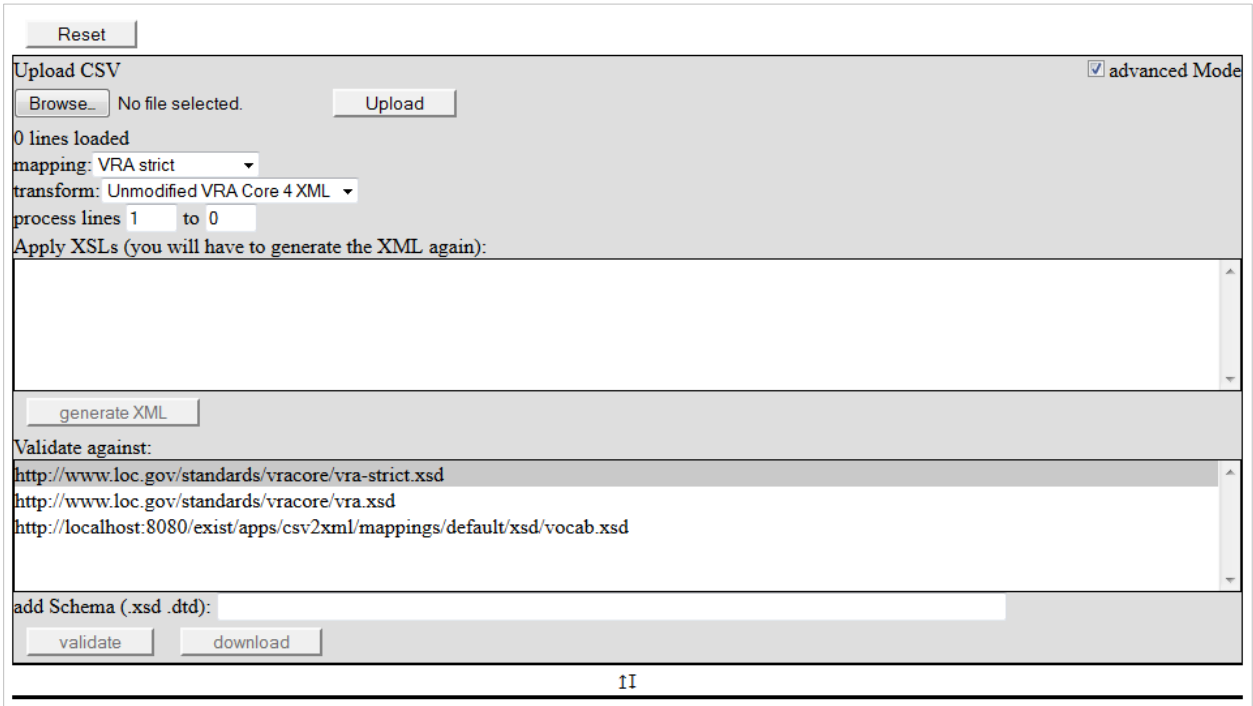
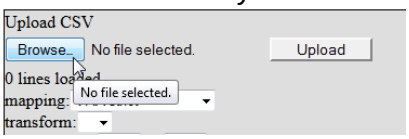


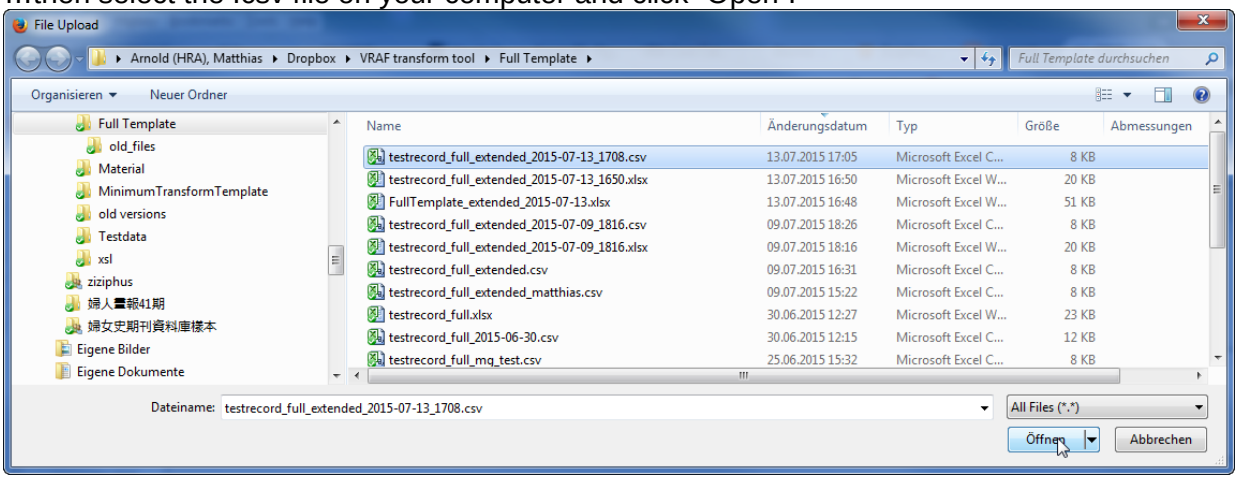
Figure 2: User interface - Advanced mode.

4. Uploading a file

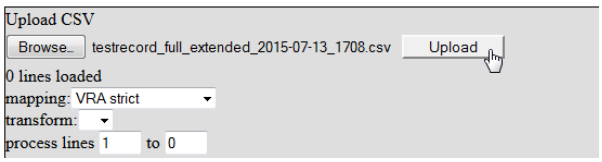
To transform data you need to upload it first. Click the "Browse" button:



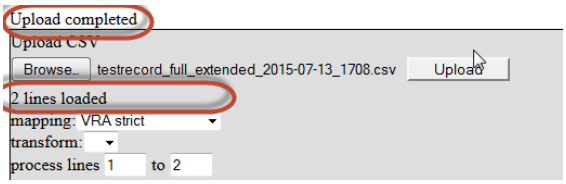
…then select the .csv file on your computer and click "Open".



The name of the .csv file will be displayed. Now click on "Upload".



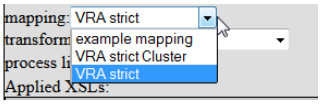
You will get an "Upload completed" message. Note that the number of lines in your csv will be shown in "x lines loaded" (in this example, the .csv only contains two lines).



5. Generating VRA Core 4 XML

5.1. Select a mapping

The tool supports generation of XML based on different mappings. For VRA Core, two variants are available: “VRA strict” and “VRA strict Cluster”.



5.1.1. *VRA strict*

This mapping is based on the VRA Core 4 restricted version http://www.loc.gov/standards/vracore/vra-strict.xsd. It uses controlled values for types and date formats. For more information please refer to http://www.loc.gov/standards/vracore/schemas.html.

This is the default mapping.

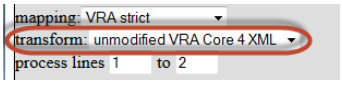
5.1.2. *4.1.2. VRA strict Cluster*

This mapping http://cluster-schemas.uni-hd.de/vra-strictCluster.xsd expands the restricted VRA Core 4 version by adding, for example, attributes for multilingual data and an element for geo-coordinates. For more information please refer to https://docs.google.com/document/d/11dFFCefjYSFX5LAQg9XK0NGQfuwUYIBIzYdZonGgfrs/edit?pli=1

5.2. Select a transformation

The tool supports different transformations. Additional transformations need to be installed. Currently, two transformations are implemented: “unmodified VRA Core 4 XML” and “transform into RDF” (for RDF transformations see below).

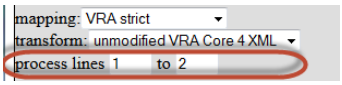
To transform the .csv data into VRA Core 4 XML choose “unmodified VRA Core 4 XML”.



5.3. Define the lines to be processed

In the “process lines” fields you can enter the number of the lines that will be processed. This feature allows to process smaller number of records or even individual ones and reduces processing time.

Per default, all lines will be processed.



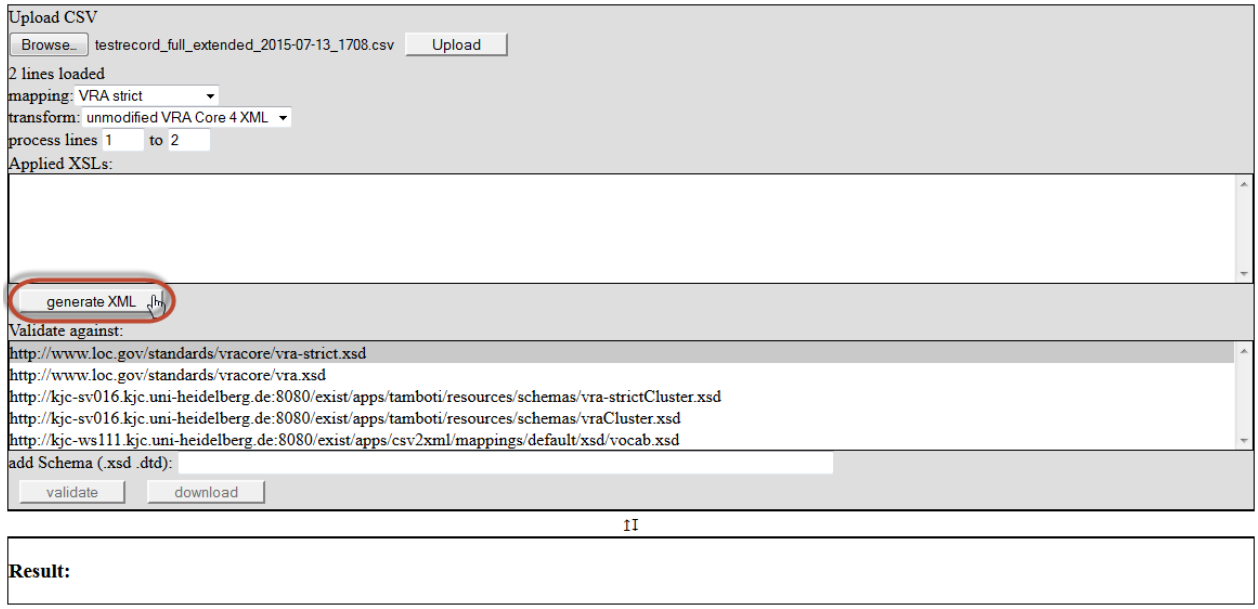
5.4. Applied XSLs

For the default transformation to VRA Core 4 XML this field will remain empty.

(See also “Generating RDF” below.)

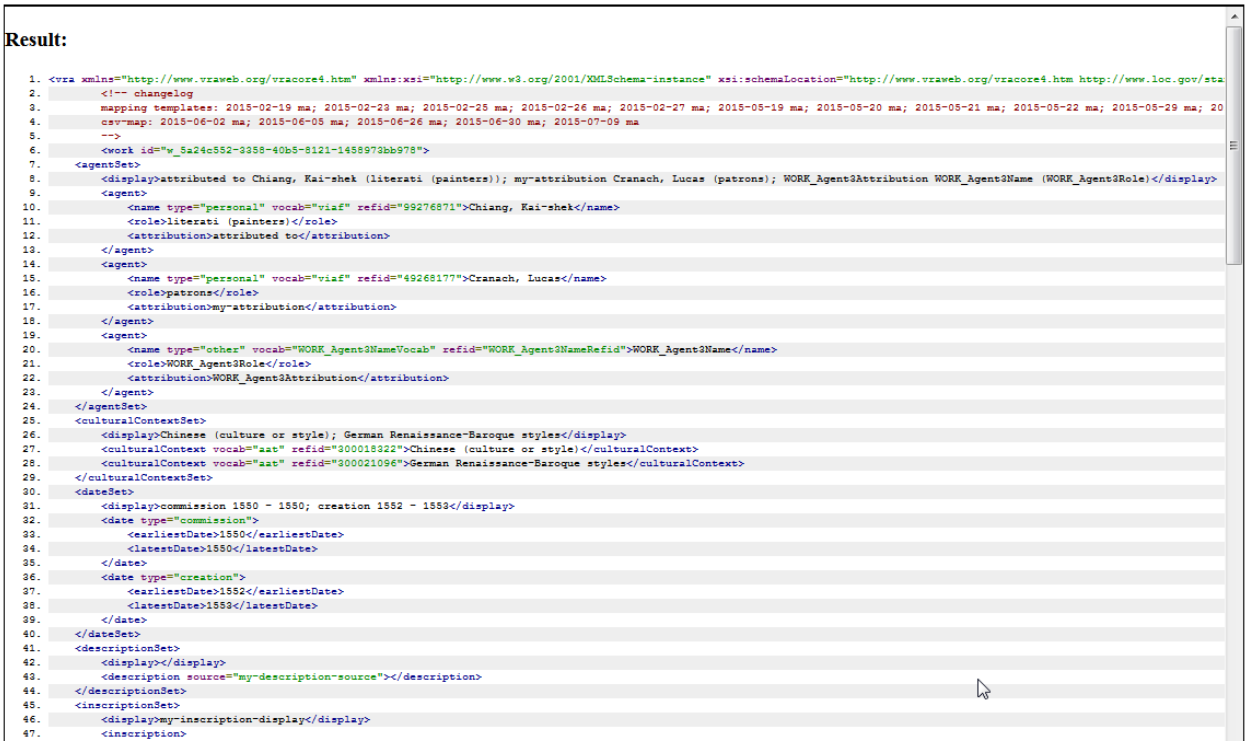
5.5. Generating XML

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



Note: Processing a large number of records may take some time, please be patient.

Once the XML is generated, a preview will be displayed in the “Result” box.

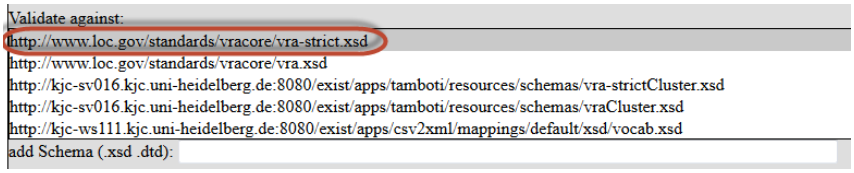


6. Validating the XML

Just by applying a transformation the tool can not be sure the resulting xml files are valid. The restricted VRA Core schema defines a number of 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.

6.1. 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.



For example, if you selected "VRA strict" as mapping, the corresponding schema http://www.loc.gov/standards/vracore/vra-strict.xsd is automatically set.

It is possible to chose different validation schemas. Users may also upload their own schema (ALPHA!).

6.2. Starting the validation

To start the validation process, click “validate”.

Note: Validating a large XML file may take some time, please be patient.

6.3. Getting the validation result

Once the validation is finished, a new browser tab will be opende with the validation result. In addition, the button’s color will change:

Result NOT VALID

Resul VALID

6.4. Interpreting the validation

A valid result will look like this:

```
1.
2.              <report-result>
3.                  <xsd>http://www.loc.gov/standards/vracore/vra-strict.xsd</xsd>
4.                  <report>
5.                      <status>valid</status>
6.                      <namespace>http://www.vraweb.org/vracore4.htm</namespace>
7.                      <duration unit="msec">17</duration>
8.                  </report>
9.              </report-result>
10.
```

6.5. Validation errors

Example:

```
<report-result>
    <xsd>http://kjc-sv016.kjc.uni-heidelberg.de:8080/exist/apps/tamboti/resources/schemas/vra-strictCluster.xsd</xsd>
    <report>
        <status>invalid</status>
        <namespace>http://www.vraweb.org/vracore4.htm</namespace>
        <duration unit="msec">26</duration>
        <message level="Error" line="30" column="36">cvc-pattern-valid: Value 'early1' 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="30" column="36">cvc-complex-type.2.2: Element 'date' must have no element [children], and the value must be valid.</message>
    </report>
</report-result>
```

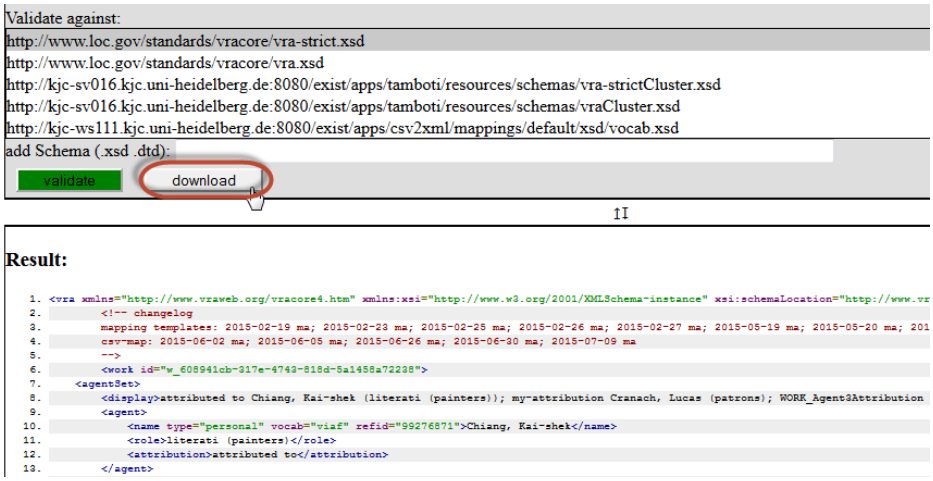
<message level="Error" line="587" column="24">cvc-pattern-valid: Value 'early1' is not facet-valid with respect to pattern 'present|(-)*[0-9]{1,12}(-[0-9]{2}(-[0-9]{2})*)*' for type 'dateValueType'.</message>

<message level="Error" line="587" column="24">cvc-complex-type.2.2: Element 'date' must have no element [children], and the value must be valid.</message>

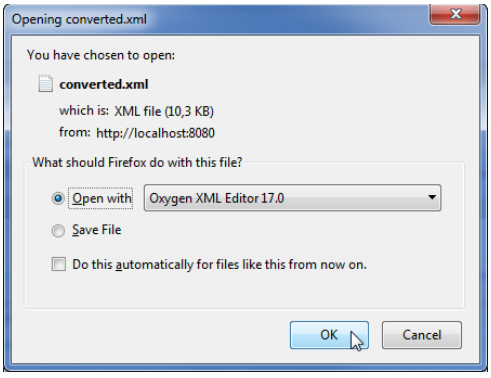
In this case the …

7. Downloading the generated XML

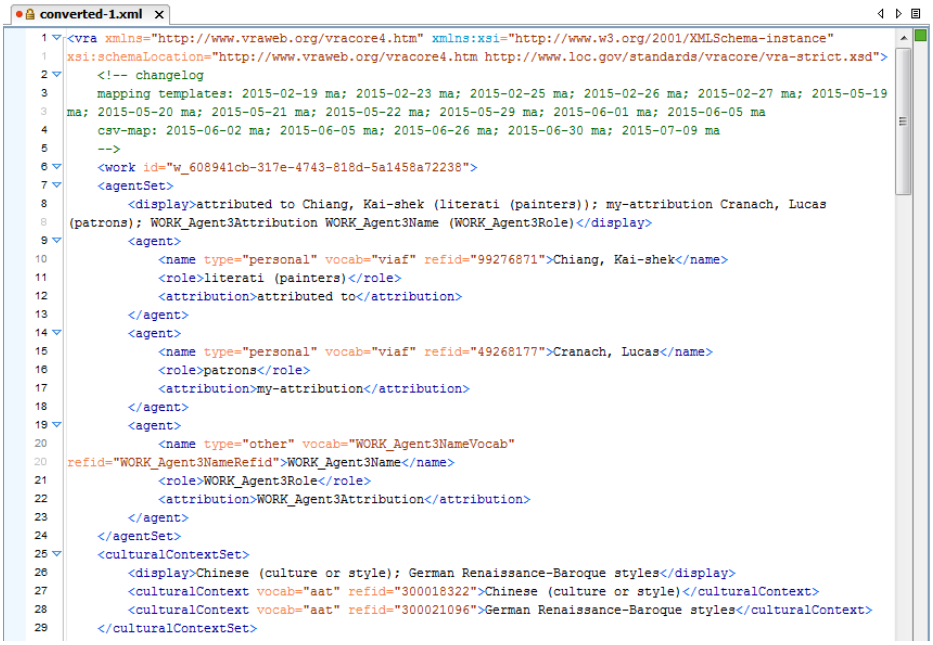
After successful validation click the “download” button to copy the valid XML to your computer.



The default file name will be “converted.xml”.



Display of the generated file in Oxygen.



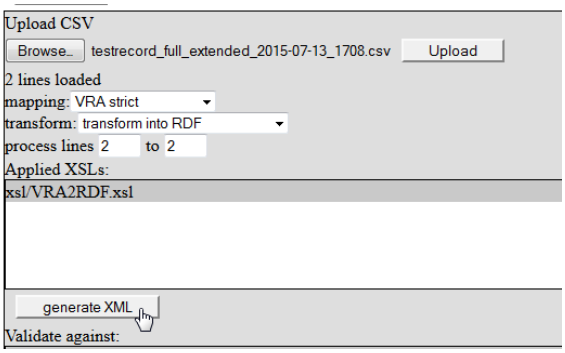
8. Generating RDF XML

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.

The tool makes use of the XSLT stylesheet as provided by the VRA-RDF-Project https://github.com/mixterj/VRA-RDF-Project.

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.

Result:

```
<rdf:RDF xmlns:library="http://purl.org/library/" xmlns:void="http://rdfs.org/ns/void#" xmlns:vra="http://purl.org/vra/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:rdfs="http
    --><!--Skipping over
    -->
  <rdf:Description rdf:about="#">
    <rdf:type rdf:resource="http://purl.org/vra/CreativeWork"></rdf:type><!--Skipping over
  --><!--Skipping over agentSet--><!--Skipping over
    --><!--Skipping over display--><!--Skipping over attributed to Chiang, Kai-shek (literati (painters)); my-attribution Cranach, Lucas (patrons); WORK_Agent3Attribution WORK
    -->
    <vra:creator>
      <rdf:Description>
        <rdf:type rdf:resource="http://purl.org/vra/Agent"></rdf:type>
        <rdf:type rdf:resource="http://purl.org/vra/Person"></rdf:type>
        <vra:name>Chiang, Kai-shek</vra:name>
      </rdf:Description>
    </vra:creator><!--Skipping over
    -->
    <vra:creator>
      <rdf:Description>
        <rdf:type rdf:resource="http://purl.org/vra/Agent"></rdf:type>
        <rdf:type rdf:resource="http://purl.org/vra/Person"></rdf:type>
        <vra:name>Cranach, Lucas</vra:name>
      </rdf:Description>
    </vra:creator><!--Skipping over
    -->
    <vra:creator>
      <rdf:Description>
        <rdf:type rdf:resource="http://purl.org/vra/Agent"></rdf:type>
        <vra:name>WORK_Agent3Name</vra:name>
      </rdf:Description>
    </vra:creator><!--Skipping over
  --><!--Skipping over
  --><!--Skipping over culturalContextSet--><!--Skipping over
    --><!--Skipping over display--><!--Skipping over Chinese (culture or style); German Renaissance-Baroque styles--><!--Skipping over
    -->
    <vra:culturalContext>
      <rdf:Description>
        <vra:name>Chinese (culture or style)</vra:name>
      </rdf:Description>
```

9. Contact

If you have problems or questions please contact Matthias Arnold at marnold@asia-europe.uni-heidelberg.de

Contact details
arnold@asia-europe.uni-heidelberg.de
Phone: +49 (0) 6221 - 54 4094
Fax: +49 (0) 6221 - 54 4012
Skype: matz-skype

Office address
Karl Jaspers Centre, Room 005b

MediaLab
Karl Jaspers Centre, Room 005c

Further links
HRA Portal
http://hra.uni-hd.de/

MediaLab Tutorials in Sharepoint
https://sharepoint.urz.uni-heidelberg.de/vjc/kjc/hra/medialab/

10. 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 – NameType – NameVocab – NameRefID – ObjType – 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]

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
[Filename (-> href); accessionID (-> refid)]

Agent display
1x nameType – name – nameVocab – nameRefid – role

CulturalContext display

Date display

Description display [display = description]

Inscription display [display = text]

Location display
[localRepoName, accession-id]

Material display

Measurements display

Relation display
[transform adds image-work link]

Rights display
Rights notes
1x type – holder – text

Source display
[VendorName, VendorID]
1x sourceType, sourceValue

StateEdition display

Style/Period display

Subject display
3x subject – vocab – refid – type

Technique display
[Technique: <technique vocab="AAT" refid="300237903">digital imaging</technique>]

TextRef display

Title display
1x type, title

WorkType display
[Worktype: <worktype vocab="AAT" refid="300215302">digital images</worktype>]

11. Appendix: Columns Full Template

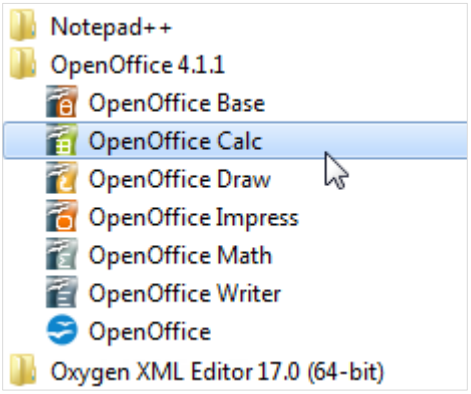
IMAGE_Filename
IMAGE_AccessionID
IMAGE_VendorID
IMAGE_VendorName
WORK_ID
WORK_AgentDisplay
WORK_Agent1NameType
WORK_Agent1Name
WORK_Agent1NameVocab
WORK_Agent1NameRefid
WORK_Agent1Role
WORK_Agent1Attribution
WORK_Agent2NameType
WORK_Agent2Name
WORK_Agent2NameVocab
WORK_Agent2NameRefid
WORK_Agent2Role
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_DescriptionDisplay
WORK_DescriptionSource
WORK_InscriptionDisplay
WORK_LocationDisplay
WORK_LocationNotes
WORK_Location1Type
WORK_Location1Name
WORK_Location1NameType
WORK_Location1NameVocab
WORK_Location1NameRefID
WORK_Location1ObjType
WORK_Location1ObjRefID
WORK_Location1Geo
WORK_Location1GeoVocab
WORK_Location1GeoRefid
WORK_Location2Type
WORK_Location2Name
WORK_Location2NameType
WORK_Location2NameVocab
WORK_Location2NameRefID
WORK_Location2ObjType
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_MaterialRefid3
WORK_Material4
WORK_MaterialVocab4
WORK_MaterialRefid4
WORK_MeasurementsDisplay
WORK_Measurements1Extent
WORK_Measurements1Unit
WORK_Measurements1Value1
WORK_Measurements1Type1
WORK_Measurements1Value2
WORK_Measurements1Type2
WORK_Measurements1Value3
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_RelationDisplay
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_TitleDisplay
WORK_TitlePref
WORK_TitlePrefType
WORK_TitleAlt
WORK_TitleAltType
WORK_WorktypeDisplay
WORK_Worktype1
WORK_WorktypeVocab1
WORK_WorktypeRefid1
WORK_Worktype2
WORK_WorktypeVocab2
WORK_WorktypeRefid2
WORK_Worktype3
WORK_WorktypeVocab3
WORK_WorktypeRefid3
IMAGE_AgentDisplay
IMAGE_Agent1NameType
IMAGE_Agent1Name
IMAGE_Agent1NameVocab
IMAGE_Agent1NameRefid
IMAGE_Agent1Role
IMAGE_CulturalContextDisplay
IMAGE_DateDisplay
IMAGE_DescriptionDisplay
IMAGE_InscriptionDisplay
IMGAE_LocationDisplay
IMAGE_MeasurementsDisplay
IMAGE_RelationDisplay
IMAGE_RightsDisplay
IMAGE_RightsNotes
IMAGE_RightsType
IMAGE_RightsHolder
IMAGE_RightsText
IMAGE_SourceDisplay
IMAGE_SourceValue
IMAGE_SourceType
IMAGE_StateEditionDisplay
IMAGE_StylePeriodDisplay
IMAGE_SubjectDisplay
IMAGE_Subject1
IMAGE_SubjectRefid1
IMAGE_SubjectVocab1
IMAGE_SubjectType1
IMAGE_Subject2
IMAGE_SubjectRefid2
IMAGE_SubjectVocab2
IMAGE_SubjectType2
IMAGE_Subject3
IMAGE_SubjectRefid3
IMAGE_SubjectVocab3
IMAGE_SubjectType3
IMAGE_TechniqueDisplay
IMAGE_TextrefDisplay
IMAGE_TitleDisplay
IMAGE_Title
IMAGE_TitleType
IMAGE_WorktypeDisplay

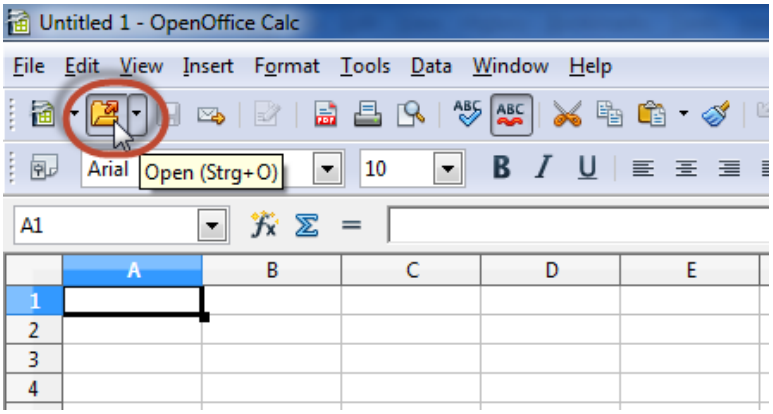
12. Appendix: Convert .xslx file to .csv using OpenOffice Calc

Save your Excel spreadsheet as usual.

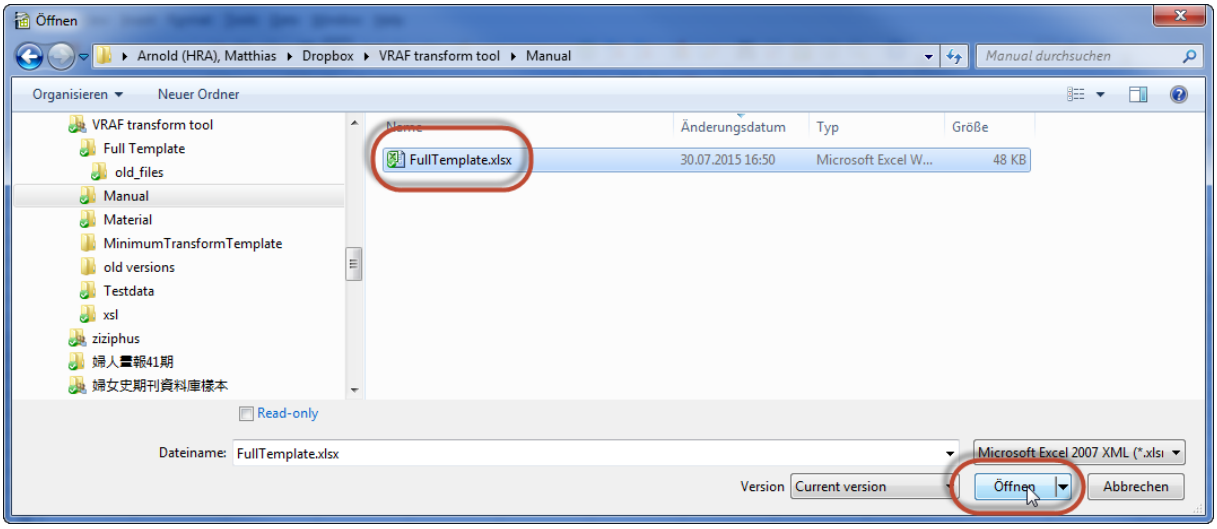
Start OpenOffice Calc

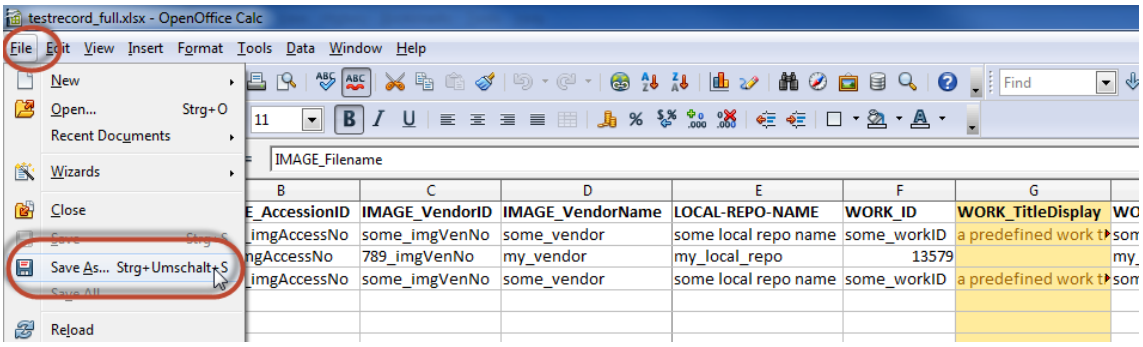


Click “Open”…

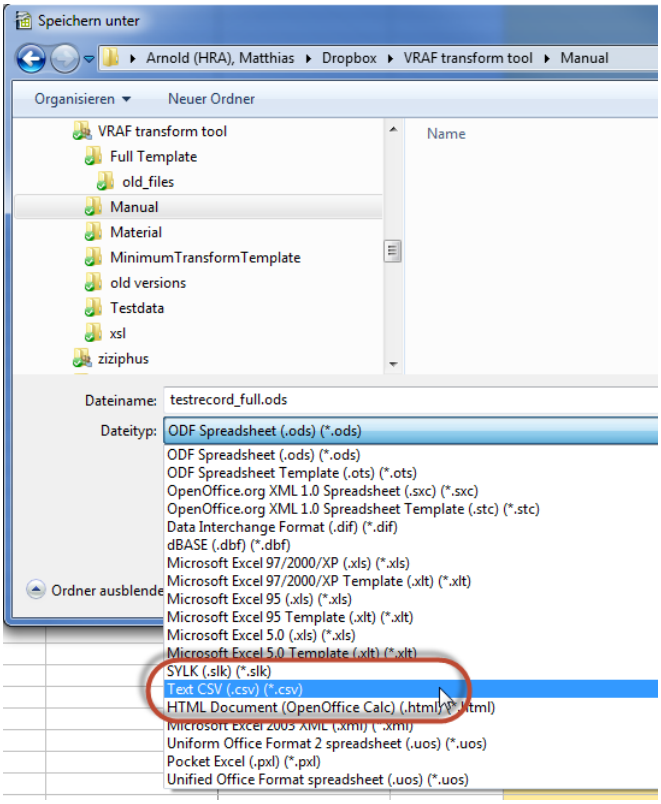


…and select your spreadsheet.



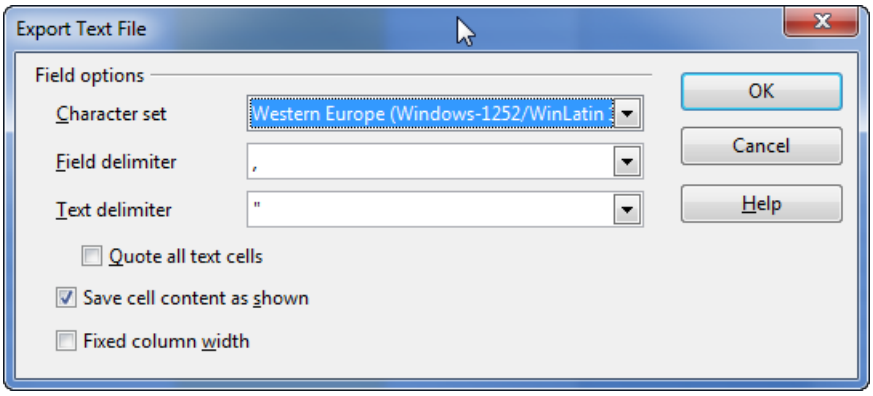


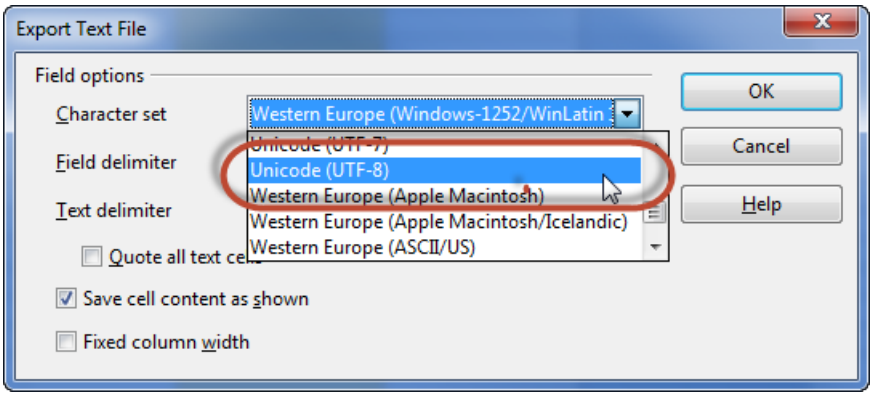
Once the document is opened go to the “File menu” and select “Save As…”



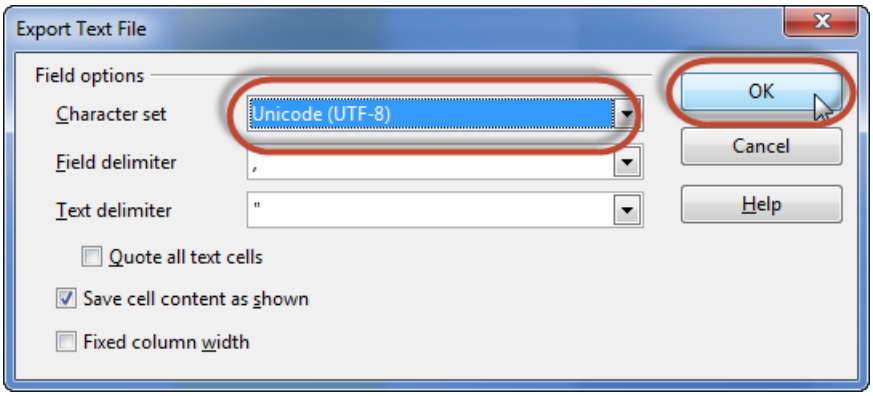
Chose “Text CSV (.csv) (*.csv)” and hit the “Save” button.

This will open the “Export Text File” dialog.

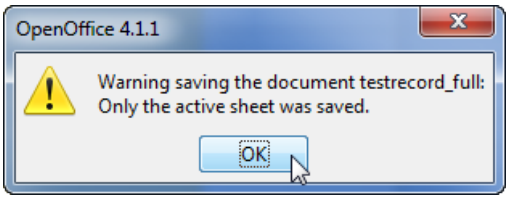




Change the “Character set” to “Unicode (UTF-8)”.



Then click “OK”.



A message will be displayed, informing you that only the active sheet was saved to .csv format. Click “OK” to confirm.

You can now close Calc.

To double-check if your data is correctly encoded, open the .csv file in an editor like NotePad++. If it displays correctly, you can proceed to XML transformatoin.