

# XML Visualisierung im Translatio Nummorum Projekt per XSLT

Kern der Visualisierung ist die XSL Transformation, namens „trans\_num.xslt“. Wenn diese in den Header des zu visualisierenden XML Files eingetragen wird kann sie mit einem geeigneten Webbrowser wie Firefox (empfohlen) visualisiert werden. Alternativ kann die „trans\_num.xslt“ auch mit einem passenden XML Viewer verwendet werden. Das Script wird stetig weiterentwickelt und dient lediglich der projektinternen Visualisierung von Transkripten. Entsprechend ist es weder auf Geschwindigkeit noch auf einfaches Deployment optimiert.

## Abhängigkeiten

Damit die „trans\_num.xslt“ funktioniert, muss sie sich im gleichen Ordner wie die Transcript XMLs und die „monumentum.xml“ befinden. Im selben Verzeichnis müssen außerdem folgende Dateien liegen, deren sich „trans\_num.xslt“ bedient:

- arrow003.png
- arrow\_rs.png
- close.png
- container-back.png
- DejaVuSans-BoldOblique.ttf
- DejaVuSans-Bold.ttf
- DejaVuSans.eot
- DejaVuSans-KHI.ttf
- DejaVuSans-Oblique.ttf
- DejaVuSans.ttf
- lupe-small.png
- MKM.png
- monumentum.xml
- settingsback.png
- trans\_num.css
- trans\_num\_font.css
- trans\_num\_font-hq.css
- trans\_num.js

## Vorbereitung eines XML-Transkripts für die Visualisierung im Browser

Die Visualisierung der Transkripte im Webbrowser hat sich als die nutzerfreundlichste Herangehensweise entpuppt. Dazu muss das XML-Transkript mit der XSLT Visualisierung bekannt gemacht haben. Zu diesem Zweck trägt man im Header des XML Transkripts unter die erste Zeile folgendes ein:

```
<?xml-stylesheet type="text/xsl" href="trans_num.xslt" ?>
```

Der vollständige Header eines so manipulierten XML-Transkripts würde dann folgende Form annehmen

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-stylesheet type="text/xsl" href="trans_num.xslt" ?>
<content>
```

Wichtig ist, dass das Transkript-XML im UTF-8 Format ohne BOM vorliegt bzw. gespeichert wird.

## Konfiguration

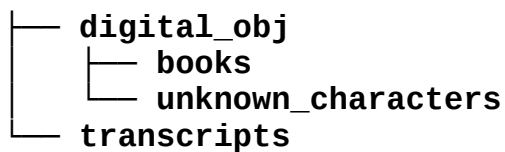
Bestimmte Parameter des XSLT Scriptes lassen sich ändern und auf die eigene Arbeitsumgebung anpassen. All diese Parameter sind im oberen Abschnitt des Scriptes definiert. Am wichtigsten sind die Pfadangaben zu den Digitalisaten und zu den „Unknown Characters“ (das sind Bitmaps von Buchstaben, die nicht im UFT-8 Standard bzw. gängigen Fonts enthalten sind). Pfadangaben sollten relativ gemacht werden, also ausgehend von der Lage der XSLT in der Ordnerstruktur. So ist eine komplikationslose Portabilität des Projektordners sichergestellt.

Es folgt eine Auflistung aller Konfigurationsparameter:

Setup-Code	Erläuterung
<pre>&lt;!-- Pfad zu ausgeschnittenen Bildern und Buchseitenscans --&gt; &lt;xsl:variable name="IMG_PATH"&gt;   &lt;xsl:text&gt;../digital_obj/books/&lt;/xsl:text&gt;   &lt;xsl:value-of select="//opacid"/&gt;   &lt;xsl:text&gt;&lt;/xsl:text&gt; &lt;/xsl:variable&gt;</pre>	<p>Anpassungen u.U. sinnvoll. Pfadeinstellung hier: <code>&lt;xsl:text&gt;[IHR PFAD]&lt;/xsl:text&gt;</code> Das ist der Pfad zu den Digitalisaten (ideraler Weise relativ, wie hier), die unterhalb dieses Pfades in Ordnern liegen, deren Name exakt der im Transkript XML angegebenen Opac-ID entspricht. Abschließend muss ein „/“ stehen.</p>
<pre>&lt;!-- Pfad zu unbekannten Buchstaben (als Bild ausgeschnitten vorliegend) --&gt; &lt;xsl:variable name="UC_PATH"&gt;   &lt;xsl:text&gt;../digital_obj/unknown_characters/&lt;/xsl:text&gt; &lt;/xsl:variable&gt;</pre>	<p>Anpassungen u.U. sinnvoll. Pfadeinstellung hier: <code>&lt;xsl:text&gt;[PFAD zum Ordner mit den Unknown Characters]&lt;/xsl:text&gt;</code> Das ist der Pfad zum Ordner, der die Unknown Character Bitmaps enthält (ideraler Weise relativ, wie hier). Abschließend muss ein „/“ stehen.</p>
<pre>&lt;!-- Skalierung von in den Text eingeblendeten Symbolen wie Links in px --&gt; &lt;xsl:variable name="SymbolScale"&gt;   &lt;xsl:text&gt;15&lt;/xsl:text&gt; &lt;/xsl:variable&gt;</pre>	<p>Anpassung nicht empfehlenswert. Anpassung der Icon Größe.</p>
<pre>&lt;!-- Höhen-Skalierung von Unknown Charakters ("UC") --&gt; &lt;xsl:variable name="UCScale"&gt;   &lt;xsl:text&gt;20&lt;/xsl:text&gt;   &lt;!-- Ganz guter Kompromiss zw. Size und Lesbarkeit. Zwar größer als Text, allerdings sind auch im Original einige der UCs deutlich größer gedruckt als der restliche Text; mehr als 50px ist quatsch, denn auf diese Height sind die Buchstabenbilder physisch gerechnet --&gt; &lt;/xsl:variable&gt;</pre>	<p>Anpassung nicht empfehlenswert. Anpassung der Größe von Unknown Characters,</p>

## Empfohlene Ordnerstruktur

Wenn Sie unterhalb eines beliebigen Translatio Nummorum Projektordners folgende Ordnerstruktur einrichten, brauchen Sie keine Änderungen am Setup des XSLT Scriptes vorzunehmen:



Wenn Sie das gelieferte 7z File auspacken, sollte exakt diese Struktur entstehen. Bitte beachten Sie die originalgetreue Gross/Kleinschreibung.

## Erläuterung der Struktur

Ordnername	Bedeutung
digital_obj	Hier befindet sich alles, was nicht direkt mit XML und dessen Darstellung zu tun hat, in der Regel Bilddateien. Auf dieser Ebene befinden sich keine einzelne Dateien, nur weitere Unterordner.
books	Unterordner von digital_obj. Hier befinden sich in Unterordnern die Buchdigitalisate. Diese Buchdigitalisats-Unterordner tragen den Namen der Opac-ID, wie er im dazu passenden Transkript-XML eingetragen ist. Zum Beispiel „b304254f“ (dieses Buch wird als Beispiel mitgeliefert).
unknown_characters	In diesem Ordner befinden sich Bitmaps von schwierig darstellbaren Buchstaben.
transcripts	In diesem Ordner befinden sich die XML Transkripte sowohl die XSLT-Datei mitsamt aller abhängigen Dateien.

## Einschränkungen

Erst nach Abschluß aller Änderungen wird am KHI das XML-Transkript finalisiert (derart finalisierte XMLs dürfen dann nicht mehr bearbeitet werden). Dabei werden bestimmte ID Nummern automatisch innerhalb des Transkripts generiert. Erst mit diesen automatisch generierten Nummern ist es möglich, eingebettete Images (<img>) darzustellen, ansonsten wird ein Platzhalter angezeigt. Genau dies ist in der vorliegenden Version der Fall – eingebettete <img> Bilder werden nicht angezeigt. Die Anzeige von kompletten Buchseiten wird hingegen unterstützt.