

# Allgemeine Beschreibung

---

Die von *FromThePage* exportierten TEI-XML-Dokumente sollen nun für die weitere Bearbeitung in *ediarum* an die Grundstruktur des DTA-Basisformat (DTABf) angepasst werden. Auch wenn für die Verarbeitung in *ediarum* kein valides DTABf-XML erzeugt werden muss, wird mit dem XSLT der vorliegenden Transition versucht, dem DTA-Basisformat in so vielen Bereichen wie möglich gerecht zu werden.

## Voraussetzungen

Die im DigEdTnT-Projekt vorgestellten Transitions setzen nicht nur bestimmte Kompetenzen der Benutzer:innen voraus, sondern stellen auch hinsichtlich der Software-Umgebung gewisse Anforderungen.

### Erforderliche Kenntnisse

- [EDV-Grundkenntnisse](#)
- Einrichten einer [Oxygen-Transformation](#)
- Grundlegende XSLT-Kenntnisse (für erweiterte Anpassungen)

### Benötigte Software

- Oxygen Editor

→ Hinweis: Natürlich wäre es auch ohne Oxygen Editor möglich, über diverse frei zugängliche Online-Plattformen eine XSL-Transformation durchzuführen. Da aber für *ediarum* in dieser Pipeline ohnehin Oxygen benötigt wird, wird nachfolgend das Transformationsszenario im Oxygen Editor beschrieben.

## Möglichkeiten & Grenzen

---

Der Übergang von einem Tool zu einem anderen lässt sich verschieden gestalten. Nachfolgend soll ein Überblick über die Vor- und Nachteile unserer Transition gegeben werden.

### Stärken

- Nicht valider TEI-XML-Export von *FromThePage* wird in anerkannten Standard (DTA-Basisformat) überführt

### Herausforderungen & Probleme

- Einschränkung hinsichtlich der Möglichkeiten an Metadaten-Elementen im DTA-Basisformat
- Entscheidung für einen DTA-konformen XML-Output führt zu Informationsverlust
- Verpflichtende Attribute bei Elementen gemäß DTABf, die zuvor in *FromThePage* nicht hinzugefügt wurden, führen zu einem nicht validen Dokument

## XSL Transformation

---

Für eine Transformation des aus *FromThePage* exportierten TEIs in ein für *ediarum* weiterverarbeitbares Dokument, haben wir für unser [Beispielprojekt](#) die Daten transformiert. Wie unser Transformationsszenario genau funktioniert, ist dem Punkt [XSLT Dokumentation](#) zu entnehmen. Dieses kann für andere Projekte angepasst und weiterverwendet werden.

Hier ist die [XSLT-Ressource](#) für unsere FromThePage-ediarum-Transformation.

Eine Kurzanleitung für das Einrichten eines Transformationsszenarios findet sich [hier](#).

# XSLT Dokumentation

---

## 1. Einbindung des DTA-Basisformat

Ein erster Schritt ist die Einbindung des [Schematron Regelsatzes](#) sowie des [RelaxNGs des Deutschen Textarchivs](#). Da es sich in unserem Fall um ein Manuskript handelt, wird in der XSL-Transformation neben dem [entsprechenden DTABf-Schematron](#) das [RNG-Schema des DTABf für Manuskripte](#) eingebunden.

```
<xsl:template match="/">
  <xsl:processing-instruction name="xml-
model">href="http://www.deutschestextarchiv.de/basisformat_ms.rng"
type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"
</xsl:processing-instruction>
  <xsl:processing-instruction name="xml-
model">href="http://www.deutschestextarchiv.de/basisformat.sch"
type="application/xml"
schematypens="http://purl.oclc.org/dsdl/schematron</xsl:processing-instruction>
  <xsl:apply-templates></xsl:apply-templates>
</xsl:template>
```

Dies erzeugt im transformierten XML die entsprechende Einbindung der XML-Modelle:

```
<?xml-model href="http://www.deutschestextarchiv.de/basisformat_ms.rng"
type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model href="http://www.deutschestextarchiv.de/basisformat.sch"
type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron?>
```

## 2. Transformation der Metadaten

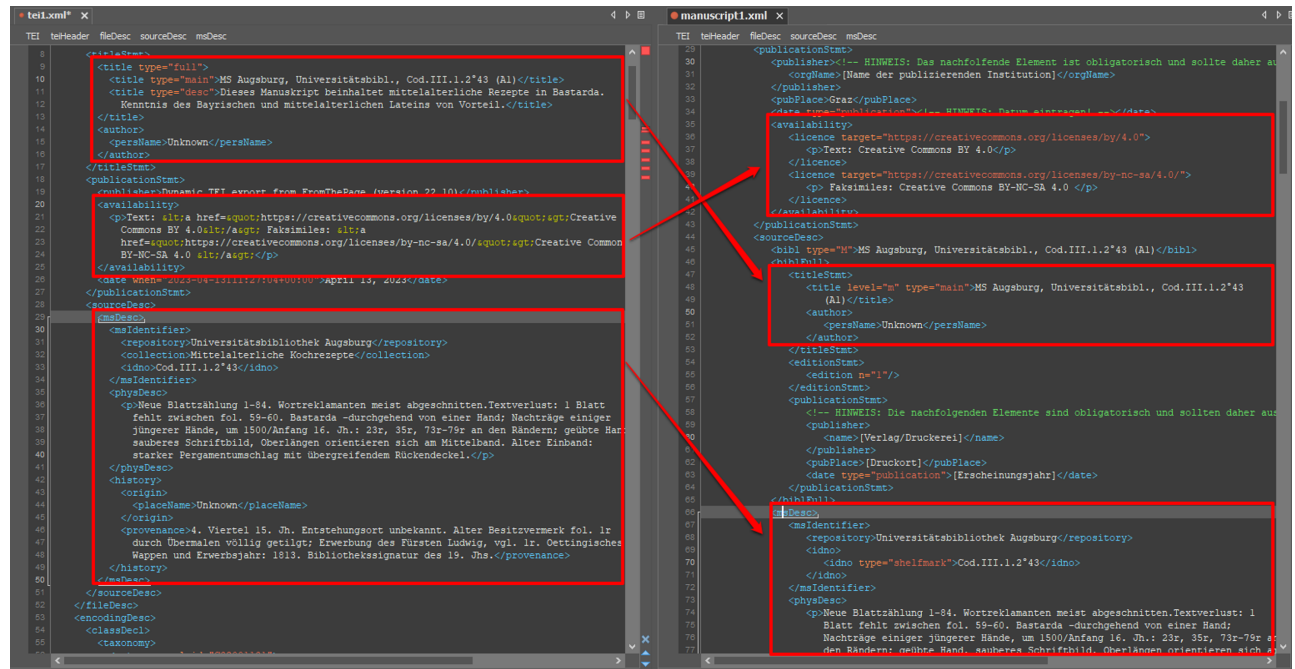
Auf Grundlage eines als gültig validierten [DTA-Metadatenbeispiels](#) wurden schließlich die Metadaten des *FromThePage*-Exports in eine DTABf-konforme Struktur übertragen.

### Transformationsarten

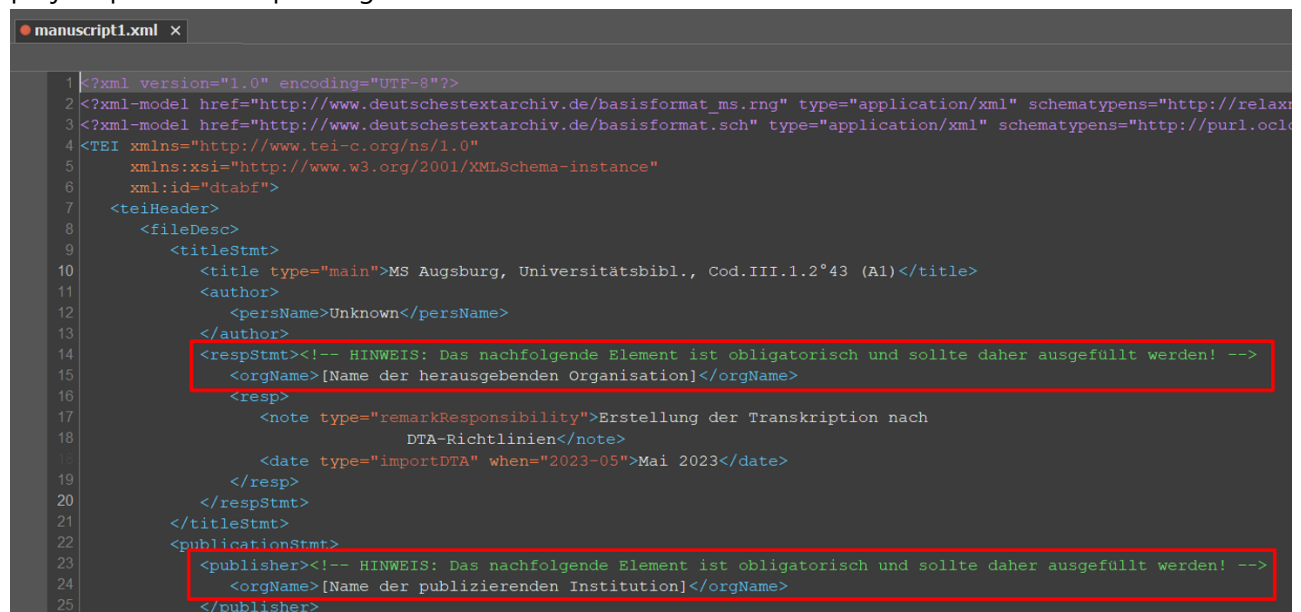
Bei der Übertragung der Metadaten wurden verschiedene Maßnahmen gesetzt:

- **Übertragung von Inhalten:** Nach einer Gegenüberstellung des exportierten TEIs und dem DTABf-Beispiel wurden zuerst Inhalte wie Titel, Autor, Lizenzen sowie die Beschreibung des Manuskripts

übernommen und in entsprechende zulässige bzw. obligatorische Elemente im Header-Bereich des überführt.



- **Einfügen von Platzhaltern:** Einige obligatorische Felder im DTABf-XML, für die es keine entsprechenden Metadatenbeschreibungen im TEI-Export von *FromThePage* gibt, wurden schließlich mit beschreibenden Platzhaltern in eckigen Klammern gefüllt. Diese erfordern noch eine nachträgliche projektspezifische Anpassung in *ediarum*.



- **Entfernen von (vorerst) irrelevanten Elementen:** Einige der Elemente im *FromThePage*-Export wurden außerdem ignoriert und nicht in das neue DTA-XML überführt, da diese im Zuge der Edition eine unwesentliche Rolle spielen oder ohnehin am Ende des Projekts noch ergänzt werden müssen. Als nicht weiter relevante Information im TEI-Export von *FromThePage* wurden zum Beispiel Angaben zu Änderungen während der Transkription identifiziert. Aber auch die in den <respStmt> enthaltenen Mitarbeitenden wurden nicht übernommen, da der *teiHeader* ohnehin am Ende des Projektes noch einmal überarbeitet werden muss.
- **Hinzufügen von speziellen Inhalten für ediarum:** Damit es beim Importieren der XML-Dokumente in *ediarum* zu keinen Schwierigkeiten kommt, ist es außerdem notwendig, dass mit dem XSLT im <TEI>-Element ein Attribut zur Namespace-Deklaration von *telota* sowie ein weiteres Attribut zur

Dokumententypbestimmung eingefügt werden. In den für den *ediarum*-Import transformierten XML-Dokumenten sieht das <TEI>-Element letztlich folgend aus:

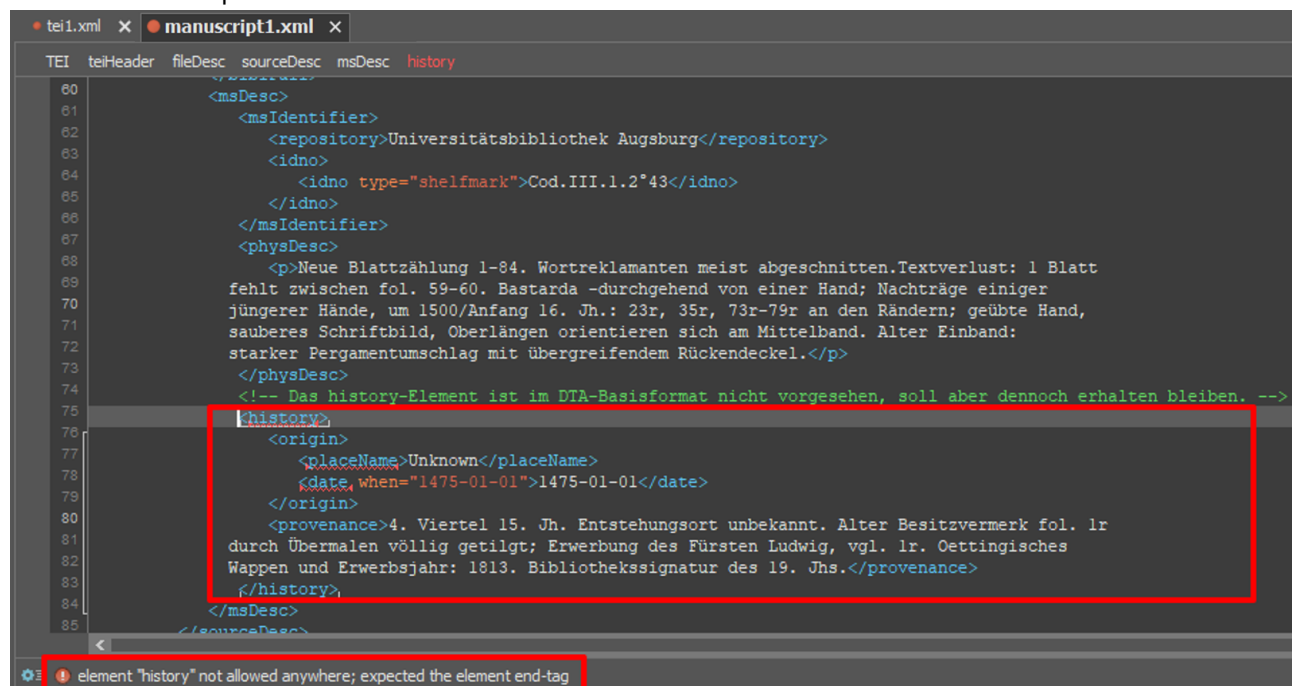
```
<TEI xmlns="http://www.tei-c.org/ns/1.0"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:telota="http://www.telota.de" xml:id="dtabf"
telota:doctype="document">
```

Dass das Attribut @telota:doctype einen Validierungsfehler hervorruft, muss vorerst ignoriert werden.

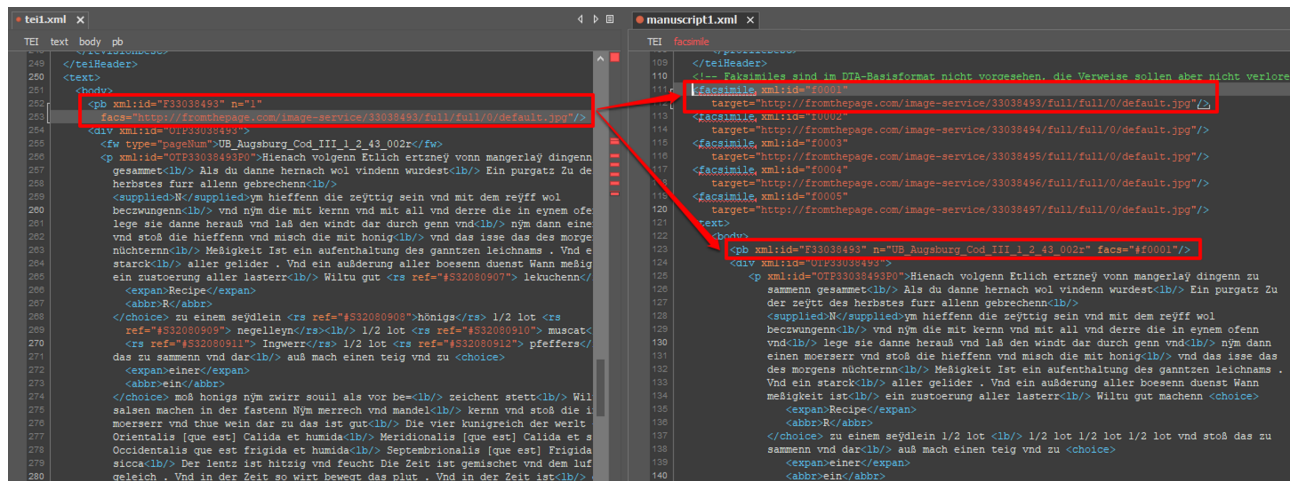
## Einschränkungen

Die Überführung der Metadaten in das Schema des DTA-Basisformat ergab außerdem Schwierigkeiten, deren Lösung nicht zu einem validen Output führt. Diesen Kompromiss muss man aber mitunter eingehen, um keine bereits vorhandenen Metadaten zu verlieren.

- **Fehlende Elemente:** Würde man sich strikt an das DTABf-Schema halten, müsste man auf bereits in *FromThePage* hinzugefügte Metadaten verzichten, was zu einem Datenverlust führen würde. So gibt es beispielsweise weder das <history>-Element zur Beschreibung der geschichtlichen Hintergründe wie Herkunft und Entstehung eines Manuskripts im DTA-Basisformat, noch gibt es ein anderes passendes Element, in das die historischen Metadaten übertragen werden können. Im Rahmen dieses Projekts wurde daher entschieden, das <history>-Element dennoch beizubehalten, auch wenn dies zu einem nicht validen Output führt.



- **Einbindung der Faksimiles:** Im DTA-Basisformat sind keine <facsimile>-Elemente zulässig. Es gibt zwar die Möglichkeit, die <pb>-Elemente zu Beginn jeder transkribierten Seite mit einem @facs-Attribut zu ergänzen, der Wert dieses Attributs darf jedoch keine URL sein. Der Wert des ersten @facs in einem <pb> muss gemäß Schematron "#f0001" entsprechen. Da es für unser Projekt aber im weiteren Verlauf wichtig ist, die URLs, die zu den Bildressourcen führen, beizubehalten, sieht unsere Transformation die Aufnahme von <facsimile>-Elementen vor. Die im *FromThePage*-Export enthaltenen Links auf die Bilddateien (dort im @facs-Attribut des <pb>-Elements) finden sich im transformierten DTABf-XML im @target-Attribut der <facsimile>-Elemente wieder.



→ Auch wenn das Einbinden der **<facsimile>**-Elemente zu einem nicht validen XML führt, lässt sich das Dokument ohne Probleme mit *ediarum* weiterverarbeiten. Es wäre aber auch möglich, das DTABf-Schema um **<facsimile>**-Elemente oder andere projektspezifisch relevante Elemente zu erweitern.

### 3. Transformation des transkribierten Manuskripts

Der wichtigste Abschnitt unserer Transformation besteht letztlich darin, die in *FromThePage* teilweise ausgezeichneten Manuskript-Transkriptionen in eine DTABf-konforme Struktur zu übertragen. Dafür soll nun zuerst eine kurze Übersicht darüber gegeben werden, welche Annotationen bei unserem [Beispielprojekt](#) bei der Transkription über *FromThePage* bereits erfolgt sind, ob bzw. welche Entsprechung es dafür im DTABf-Basisformat gibt, und mit welchen Problemen man bei der Transformation konfrontiert ist.

| Annotation                                   | FromThePage Export                                         | DTABf                                                          | Transformation & Probleme               |
|----------------------------------------------|------------------------------------------------------------|----------------------------------------------------------------|-----------------------------------------|
| <b>Button und Darstellung in Webeditor</b>   | <b>Element in TEI-Export</b>                               | <b>Zulässiges Äquivalent</b>                                   |                                         |
| Abkürzung (durch Schreiber:in des Originals) | <b>abbr:</b><br><abbr<br>expan=""><br></abbr>              | <choice><br><expan><br></expan><br><abbr> </abbr><br></choice> | Keine Anpassungen nötig                 |
| Ergänzung (durch Schreiber:in des Originals) | <b>add:</b><br><add><br></add>                             | <add<br>place="XXX"><br></add>                                 | Obligatorisches Attribut:<br>@place     |
| Tilgung (durch Schreiber:in des Originals)   | <b>del:</b><br><del><br></del>                             | <del<br>rendition="XXX"><br></del>                             | Obligatorisches Attribut:<br>@rendition |
| Fußnote (durch Editor:in)                    | <b>footnote:</b><br><footnote<br>marker=""><br></footnote> | <note n=""<br>type="footnote"><br></note>                      | Anpassung unproblematisch               |

|                                    |                                            |                                                         |                                                         |                                                                                                                                                                   |
|------------------------------------|--------------------------------------------|---------------------------------------------------------|---------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Auslassung<br>(durch<br>Editor:in) | <b>gap:</b><br><gap><br></gap>             | <gap> </gap>                                            | <gap> </gap>                                            | Keine Anpassungen nötig                                                                                                                                           |
| Überschrift                        | <b>head:</b><br><head><br></head>          | <head> </head>                                          | <fw> </fw>                                              | Anpassung unproblematisch<br>Hinweis: Das <head>-<br>Element wurde in unserem<br>Beispielprojekt<br>zweckentfremdet und für eine<br>laufende Kopfzeile verwendet. |
| Zeilenumbruch                      | <b>lb:</b><br><lb> </lb>                   | <lb> </lb>                                              | <lb> </lb>                                              | Keine Anpassungen nötig                                                                                                                                           |
| Regularisierte<br>Form             | <b>reg:</b><br><reg<br>orig=""><br></reg>  | <choice><br><reg> </reg><br><orig> </orig><br></choice> | <choice><br><reg> </reg><br><orig> </orig><br></choice> | Keine Anpassungen nötig                                                                                                                                           |
| Unklare Stelle                     | <b>unclear:</b><br><unclear><br></unclear> | <unclear><br></unclear>                                 | <supplied<br>cert="low"><br></supplied>                 | Anpassung unproblematisch                                                                                                                                         |

Wie aus der Tabelle ersichtlich wird, können die meisten Elemente entweder ohne weitere Bearbeitung übernommen werden oder benötigen nur geringe Anpassungen, um dem DTA-Basisformat zu entsprechen. Nachfolgend eine transkribierte und von *FromThePage* als TEI exportierte Seite von einem der Manuskripte:

```
<pb xml:id="F33038495" n="3"
facs="http://fromthepage.com/image-service/33038495/full/full/0/default.jpg"/>
<div xml:id="OTP33038495">
<fw type="pageNum">UB_Augsburg_Cod_III_1_2_43_059v</fw>
  <p xml:id="OTP33038495P0">nach an der andernn seittenn mit grunen totternn vnd
peterlein vnd <choice>
  <expan>versaltz</expan>
  <abbr>vsaltz</abbr>
</choice> es nit<lb/> Wiltu machen ein meyschen kuchenn So nym auff zehen eÿer
vnd
zuslach<lb/> sie wol vnd nym darzu peterlein vnd rurr es <choice>
  <expan>vndereinander</expan>
  <abbr>vndeinand</abbr>
</choice> vnd nym einen<lb/> morserr vnd setz auff ein kolenn vnd thue dar ein
einen
loeffel vol schmaltz<lb/> vnd laß es heÿß werdenn vnd geuß die eÿerr darein
vnd laß es
kul pachen<lb/> vnd thu es also gantz auff ein schussel <choice>
  <expan>versaltz</expan>
  <abbr>vsaltz</abbr>
</choice> es nit<lb/> Item Wiltu machen ein essen in dem meÿen das heÿßt ein
gespoet So
nym<lb/> einen fliessendenn keß vnd schneid den in vil schnittenn die duenn
sein<lb/> vnd
```



nÿm darzu sechs eÿer vnd slach die auff den keß vnd nÿm<lb/> meichßige putternn in ein pfannen vnd thue den keß mit den eÿernn<lb/> vber das fewrr vnd zeuch es eÿ dar mit auff das es schlecht werr vnd<lb/> richte es an vnd <choice>  
 <expansion>versaltz</expansion>  
 <abbr>vsaltz</abbr>  
 </choice> es nicht<lb/> Item einen fladenn zu machenn von fischenn velcherlej sie sind So<lb/> nÿm ein dicke mandelmilch wol gemengt mit reiß mel vnd thue<lb/> dar ein einen apffel oderr zwenn vnd ein wenig wurtz vnd seud<lb/> es in einem ofen vnd laß es pachenn vnd <choice>  
 <expansion>versalcz</expansion>  
 <abbr>vsalcz</abbr>  
 </choice> es nit<lb/> Wiltu machen morchen vmb weihennachtenn So nÿm ein teick<lb/> auß weissem brot vnd auß ein wenig melbs vnd schla eÿer dar an<lb/> vnd mach zwen knebel vnd wurff die in den teick vnd zeuch<lb/> sie darInnen vmb vnd leg sie in ein schmaltz das nit zu heÿß<lb/> sej vnd wenn es ein wenig gepack So nÿm es her wider auß<lb/> vnd schneÿd es dann mitten auff dem knebel auff von <choice>  
 <expansion>einander</expansion>  
 <abbr>einand</abbr>  
 </choice><lb/> vnd full es danne mit ein geruntenn eÿernn vnd zeuch es durch<lb/> einen lindenn straubenn teick leg es in ein schmaltz vnd laß es<lb/> pachenn vnd setz die morchen dar ein vnd laß sie pachenn<lb/> Wiltu machen ein gestroczttes gepachens So mach ein teÿck von<lb/> eÿtell eÿernn vnd wurtz in wol vnd mach in gel vnd warmm <unclear>du</unclear><note n="\*" type="footnote">Lesart nach Feyl: dutzend</note><lb/>  
 <choice>  
 <expansion>gutter</expansion>  
 <abbr>gutt</abbr>  
 </choice> holmm in den teick das sie naß werdenn vnd nÿm sie dann<lb/> her auß vnd pack es in einem schmaltz vnd <choice>  
 <expansion>versaltz</expansion>  
 <abbr>vsaltz</abbr>  
 </choice> nit<lb/> Wiltu machen gut kuchenn vonn eÿernn So nÿm eÿer wie vil<lb/> du wilt vnd zu slach die wol vnd schneid semel funf lot dar<lb/>  
 <choice>  
 <expansion>vnter</expansion>  
 <abbr>vnt</abbr>  
 </choice> vnd thue dar ein weinperr vnd schmaltz in ein pfannen des<lb/> genug sej vnd geuß die eÿer dar ein vnd laß es packenn ÿnnenn<lb/> vnd aussenn do mit slach es auff

```

    ein panck vnd hack dar <choice>
      <expan>vnter</expan>
      <abbr>vnt</abbr>
    </choice><lb/> gut wurtz vnd schneid es zu scheubenn vnd richt es an<lb/> Item
Ein essen
von milch kuchenn So soltu sie clein <choice>
  <expan>schnëyden</expan>
  <abbr>schnëy</abbr>
</choice><note n="*" type="footnote">Rezept endet hier</note>
</p>
</div>

```

Bei der Transformation können einige Elemente in ihrer ursprünglichen Form erhalten bleiben, während andere an die Vorgaben des DTA-Basisformats angepasst wurden.

```

<pb xml:id="F33038495" n="UB_Augsburg_Cod_III_1_2_43_059v" facs="#f0003"/>
<div xml:id="OTP33038495">
  <p xml:id="OTP33038495P0">nach an der andernn seittenn mit grunen totternn vnd
peterlein
  vnd <choice>
    <expan>versaltz</expan>
    <abbr>vsaltz</abbr>
  </choice> es nit<lb/> Wiltu machen ein meyschen kuchenn So nÿm auff zehen
eÿer vnd
  zuslach<lb/> sie wol vnd nÿm darzu peterlein vnd rurr es <choice>
    <expan>vndereinander</expan>
    <abbr>vndeinand</abbr>
  </choice> vnd nÿm einen<lb/> morserr vnd setz auff ein kolenn vnd thue dar
ein einen
  loeffel vol schmaltz<lb/> vnd laß es heÿß werdenn vnd geuß die eÿerr
darein vnd laß
  es kul pachen<lb/> vnd thu es also gantz auff ein schussel <choice>
    <expan>versaltz</expan>
    <abbr>vsaltz</abbr>
  </choice> es nit<lb/> Item Wiltu machen ein essen in dem meÿen das heÿßt
ein
  gespoet So nÿm<lb/> einen fliessendenn keß vnd schneid den in vil
schnittenn die
  duenn sein<lb/> vnd nÿm darzu sechs eÿer vnd slach die auff den keß vnd
nÿm<lb/>
  meichßige putternn in ein pfannen vnd thue den keß mit den eÿernn<lb/>
vber das
  fewrr vnd zeuch es eÿ dar mit auff das es schlecht werr vnd<lb/> richte es
an vnd <choice>
    <expan>versaltz</expan>
    <abbr>vsaltz</abbr>
  </choice> es nicht<lb/> Item einen fladenn zu machenn von fischenn
velcherlej sie
  sind So<lb/> nÿm ein dicke mandelmilch wol gemengt mit reiß mel vnd
thue<lb/> dar
  ein einen apffel oderr zwenn vnd ein wenig wurtz vnd seud<lb/> es in einem

```



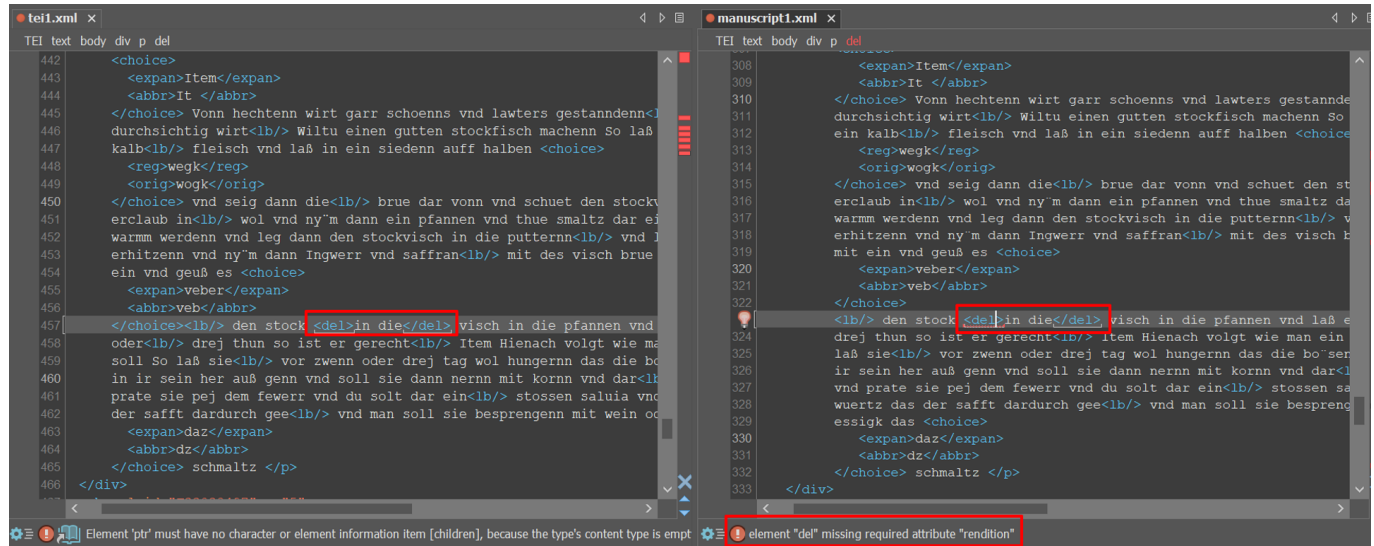
ofen vnd  
 laß es pachenn vnd **<choice>**  
     **<expan>**versalcz**</expan>**  
     **<abbr>**vsalcz**</abbr>**  
**</choice>** es nit**<lb/>** Wiltu machen morchen vmb weihennachtenn So nym ein  
 teick**<lb/>**  
 auß weissem brot vnd auß ein wenig melbs vnd schla eyer dar an**<lb/>** vnd  
 mach zwen  
 knebel vnd wurff die in den teick vnd zeuch**<lb/>** sie darInnen vmb vnd leg  
 sie in ein  
 schmaltz das nit zu heiß**<lb/>** sej vnd wenn es ein wenig gepack So nym es  
 her wider  
 auß**<lb/>** vnd schneyd es dann mitten auff dem knebel auff von **<choice>**  
     **<expan>**einander**</expan>**  
     **<abbr>**einand**</abbr>**  
**</choice>**  
**<lb/>** vnd full es danne mit ein geruntenn eyernn vnd zeuch es durch**<lb/>**  
 einen  
 lindenn straubenn teick leg es in ein schmaltz vnd laß es**<lb/>** pachenn vnd  
 setz die  
 morchen dar ein vnd laß sie pachenn**<lb/>** Wiltu machen ein gestroczt  
 gepachens So  
 mach ein teyck von**<lb/>** eytell eyernn vnd wurtz in wol vnd mach in gel vnd  
 warmm  
     **<supplied cert="low">**du**</supplied>**  
     **<note place="foot" n="\*">**Lesart nach Feyl: dutzend**</note>**  
**<lb/>**  
**<choice>**  
     **<expan>**gutter**</expan>**  
     **<abbr>**gutt**</abbr>**  
**</choice>** holmm in den teick das sie naß werdenn vnd nym sie dann**<lb/>** her  
 auß vnd  
 pack es in einem schmaltz vnd **<choice>**  
     **<expan>**versaltz**</expan>**  
     **<abbr>**vsaltz**</abbr>**  
**</choice>** nit**<lb/>** Wiltu machen gut kuchenn vonn eyernn So nym eyer wie  
 vil**<lb/>**  
 du wilt vnd zu slach die wol vnd schneid semel funf lot dar**<lb/>**  
**<choice>**  
     **<expan>**vnter**</expan>**  
     **<abbr>**vnt**</abbr>**  
**</choice>** vnd thue dar ein weinperr vnd schmaltz in ein pfannen des**<lb/>**  
 genug sej  
 vnd geuß die eyer dar ein vnd laß es packenn ynnenn**<lb/>** vnd aussenn do  
 mit slach  
 es auff ein panck vnd hack dar **<choice>**  
     **<expan>**vnter**</expan>**  
     **<abbr>**vnt**</abbr>**  
**</choice>**  
**<lb/>** gut wurtz vnd schneid es zu scheubenn vnd richt es an**<lb/>** Item Ein  
 essen von  
 milch kuchenn So soltu sie clein **<choice>**  
     **<expan>**schneyden**</expan>**  
     **<abbr>**schney**</abbr>**

```

    </choice>
    <note place="foot" n="*">Rezept endet hier</note>
  </p>
</div>

```

Die einzigen beiden Elemente, die in unserem Projekt im Textbereich für Probleme sorgen, sind Hinzufügungen (<add>) sowie Tilgungen (<del>) durch den/die ursprüngliche:n Schreiber:in. Denn während diese bei der Transkription in *FromThePage* keine Attribute erhalten haben, sind für das DTA-Basisformat Angaben zur Lokalisierung der Hinzufügung bzw. zur Art der Tilgung (Durchstreichen, Überschreiben, Radieren oder Auskratzen etc.) obligatorisch.



Da wir über diese Informationen aber nicht verfügen und daher die Manuskripte erneut durchgehen müssten, haben wir hier im transformierten XML vorerst nur die Attribute ohne Werte hinzugefügt und widmen uns diesem Problem bei der [Annotation in ediarum](#).

→ Hinweis: Die semantischen Tags, die im Rahmen der [Subject-Linking-Exploration in FromThePage](#) hinzugefügt wurden, haben wir im Zuge der Transformation wieder entfernt, da wir einerseits die semantische Annotation ausschließlich über *ediarum* vornehmen wollen und in unserem Projekt zudem nicht die in *FromThePage* dafür angelegten <rs>-Elemente genutzt werden sollen.