

Grundlagen der Datenmodellierung mit TEI/XML. Dokumentstruktur und elementare Tags für Handschriften

Summer School “Digitale Editorik”, Zentralbibliothek Zürich

Silvio Raciti
silvio.raciti@unibe.ch

Forschungsstelle Jeremias Gotthelf, Universität Bern

30. August 2023

Inhalt

- 1 Organisatorisches
- 2 Digitale Repräsentation eines Manuskripts
- 3 Markup, XML und TEI
- 4 TEI
- 5 Workflow
- 6 Analysieren und Modellieren
- 7 Codieren
- 8 Visualisieren
- 9 Publizieren
- 10 Schluss

- Historiker (kein Informatiker, kein Programmierer)
- erster Kontakt mit Markup durch \LaTeX
- seit 2015 in der Gotthelfedition tätig
- Umstellung auf digitale Arbeitsweisen miterlebt
- 2019: ein Jahr in der Zentralbibliothek für die Lavater-Briefedition tätig
- Historiker und Anwender, vor allem im 19. Jh. unterwegs

Organisatorisches

- Vormittags: Einführung in die Grundlagen von TEI
- Nachmittags: Praktische Anwendung anhand eines Arbeitsblattes
- jetzt schon Computer bereithalten für kleine Übungen zum Kennenlernen von XML und oXygen
- Empfehlung: oXygen XML Author (rotes Icon)
- Materialien aus der Dropbox beziehen und lokal auf dem eigenen Gerät speichern
- **TEI-Website** im Blick
- **Dropbox: Materialien**



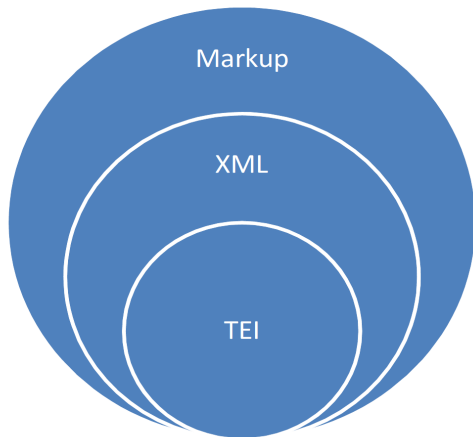
- Ausgestaltung der Repräsentation abhängig von editorischen Zielen
- unabdingbar Bedingungen:
 - Wiedergabe der Zeichensequenz
 - Strukturen/Layout
 - Resultat soll weiterverarbeitet werden können
 - Trennung von Inhalt und Darstellung
 - Editions wissenschaftliche Phänomene
- weitere Phänomene, von Editionsrichtlinien vorgegeben: z.B. Linguistische Aspekte
- Resultat: *eine* Repräsentation (Interpretation) des Textes

- Konventionen zur Erfassung von Text und Metadaten
 - neben der Zeichensequenz und Textstrukturen/Layout: auch Kontext und Materialität einer Handschrift
 - Editionsgegenstand Brief: mehr als ein Text oder eine Handschrift: ein Ereignis
- basiert auf XML: eXtensible Markup Language
- offener Standard; kann weiterentwickelt und angepasst werden
- als reiner Text gespeichert: Langzeitspeicherung und Nachnutzung
- Menschen- und maschinenlesbar (sprechende Tags)

XML-Editor: oXygen

- Editor-Modus: XML kann direkt eingegeben und bearbeitet werden
- Autor-Modus: grafische Benutzeroberfläche – fast wysiwyg
- mit Frameworks erweiterbar (diy oder extern zukaufbar)
- unterstützt eine breite Palette an zusätzlichen Formaten und Sprachen
- tolle Software, aber was wir heute machen, lässt sich mit jedem XML-Editor erledigen

Verhältnis von Markup, XML und TEI



Markup-Sprachen

- Ursprung in Druckwesen. Anweisungen für Setzer, vgl. z.B. \LaTeX
- Auszeichnung, Markierung; Verb auszeichnen, markieren, codieren
- benennt / charakterisiert Teile einer Zeichenkette in formalisierter Weise
- Codieren: implizite Eigenschaften der Teile der Zeichenkette explizit machen
- prozedurales vs deskriptives Markup

prozedurales vs. deskriptives Markup

```
<markup>
  <fett>Woody Allens Urteil</fett>
  <umbruch/>Die <kursiv>Recherche</kursiv> ist ein
  tolles, aber <kursiv>viel</kursiv> zu langes Buch.
</markup>
```

Markup

```
<markup>
  <überschrift>Woody Allens Urteil</überschrift>
  <absatz>Die <titel>Recherche</titel> ist ein
  tolles, aber <emphase>viel</emphase> zu langes
  Buch.</absatz>
</markup>
```

+

Stylesheet (Pseudo-Code)

```
überschrift = 14pt, fett, eigener Absatz
absatz      = 12pt, Blocksatz, eigener Absatz
emphase     = 12pt, kursiv, inline
titel      = 12pt, kursiv, inline
```

prozedurales vs. deskriptives Markup

- visuell, typographisch vs. semantisch, funktional, strukturell
- Anweisung wie Text dargestellt werden soll vs. Markierung der Funktion des Textstückes, Struktur und Bedeutung
- Fokus auf Darstellung vs. Darstellung ausgelagert, getrennt von Inhalt
- in beiden Fällen kein wysiwyg
- Visualisierung muss separat bewerkstelligt werden

XML

- eXtensible Markup Language
- Standard zur Darstellung von hierarchisch strukturierten Daten
- Metasprache zur Definition von XML-Formaten
- deshalb anpassbar und erweiterbar
- Prinzipien + Syntax
- wenige, mächtige Mechanismen
- “sprechende” Tags
- lizenzfrei, plattform- und anwendungsunabhängig

Beispiel für XML-Dokument

```

1837-01-08_AB-EDBern_G0370_H.xml x
TEI text body
218 <text>
219   <body>
220     <div xml:id="G0370_H_gbn_wdd_5z" type="brf_text" subtype="A">
221       <seg>
222         <pb next="1"/>
223       </seg>
224       <head>
225         <supplied>1837/01/08 – An das Erziehungsdepartement</supplied>
226       </head>
227       <opener>
228         <salute xml:id="G0370_H_cck_3qc_wgb" rend="text-align(indent1)">Hochgeachtete
229           Herren!</salute>
230       </opener>
231       <p xml:id="G0370_H_dck_3qc_wgb">Jacob Theiler von Spiez,
232         <subst n="ueberschreibung">
233           <del><unclear>provisi</unclear></del>
234           <add>provisorischer</add>
235         </subst> Lehrer <unclear>in</unclear><lb/> Lauterbach, Gemeinde Lüzelflüh, hat
236         <ex>en</ex> 2<g type="ozsuffix" rend="ten"/> Jener seine Schule<lb/>
237         <subst n="ueberschreibung">
238           <del>verlassend</del>
239           <add>verlassen</add>
240         </subst>, seinen Kostleuten abend, er gehe zu seinen Eltern<lb/> und werde
  
```

Womit arbeiten wir?

- Tags: `<head>`
- Elemente: `<head>Xyz</head>`
- Attribute: `<head n="">`; (oft wiedergegeben als @n)
- Werte `<head n="1">`

weitere Regeln und Möglichkeiten

- Kommentare: `<!-- Kommentar -->`
- processing instructions:
`<?xml-stylesheet href="style.xsl" type="text/xsl"?>`
- Tags sind case sensitive: `<name>` entspricht nicht `<Name>`
- Leere Elemente abkürzen: `<pb></pb>` entspricht `<pb/>`

Zeichen, die XML verwendet

- z.B. & oder < >
- Lösung: Entitäten
- & wird als & eingegeben
- < wird als <;, > als >; eingegeben und dann korrekt dargestellt
- in beschränktem Masse auch für geraden Apostroph und gerade Anführungszeichen ' ";

Wohlgeformtheit eines XML-Dokumentes

- Es entspricht den allgemeinen, immer gleichen Grundprinzipien von XML
- Es enthält einen Prolog, der XML-Version und Zeichensatz definiert, dem alle Zeichen im Dokument entsprechen
- Hierarchische Struktur, nur ein Element auf oberster Ebene (Wurzelement), keine Überlappungen
- Elemente können Unterelemente (“Kinder”) enthalten
- Elemente können mit Attributen näher spezifiziert werden
- Attribute können Werte haben, die in Anführungszeichen gesetzt werden
- ein Element darf nicht mehrere Attribute mit dem gleichen Namen aufweisen

Kleine Übung

- Öffnen Sie die Datei `Test_XML.xml` im Ordner `Material_TEI_Einfuehrung`
- Geben Sie eine öffnende Spitzklammer ein: `<`. Was passiert?
- Geben Sie einen Namen für einen Tag ein und eine schliessende Spitzklammer: `>`. Was passiert?
- Manipulieren Sie den Anfangstag und dann den Schlusstag und beobachten Sie was geschieht.
- Lassen Sie den öffnenden Tag mit einer Ziffer beginnen.
- Versuchen Sie ein Element um den Schlusstag zu setzen.
- Schreiben Sie ein verbotenes Zeichen. Unter `Fenster > Ansicht` zeigen können Sie das Fenster mit den Entitäten einblenden lassen.

weitere Technologien zur Verarbeitung von XML

- Stylesheet (**C**ascading **S**tylesheets: CSS, **eX**tensible **S**tylesheet **L**anguage: XSL)
- Visualisierung für unterschiedliche Ansprüche
- PDF für Druckvorlage
- HTML für Webdarstellung
- Ausgabe für unterschiedliche Textfassungen
- Tools zur Weiterverarbeitung: Anreicherung mit linguistischen Infos
- X-Technologien: z.B. XPath, XQuery, XSLT: Tools zur Abfrage, Manipulation und Weiterverarbeitung von XML-Daten

Validität

- Es entspricht der Sprache eines definierten XML-Formates
- Schema: stellt Syntax und Lexikon des XML-Formates zur Verfügung
- Hilfe beim Auszeichnen von Texten: zeigt Fehler an, schreibt oder schlägt Elemente, Attribute und/oder Werte vor
- TEI ist ein solche Sprache, die durch XML definiert wird

Von XML zu TEI

- `<TEI xmlns="http://www.tei-c.org/ns/1.0">` als Wurzelement
- zeigt an, dass hier die Tags aus den TEI-Richtlinien gelten

- Standard zur Codierung von Texten
- Institution zur Standardpflege
- Community, die den Standard weiterentwickelt

Ziele der Text Encoding Initiative

- Ermöglichung von Datenaustausch
- universelle Einsetzbarkeit: egal welche Sprache, egal welche Epoche!
- Hilfestellung für alle Typen von Anwender
- Resultat: flexibler, modularer, vielschichtiger Standard

Geschichte der TEI

- Geburtsstunde TEI: 1987
- 1988-1999: P1-P3: Entwicklung ohne Institution und Budget
- 1999: Gründung als Konsortium
- 2001-2003: von SGML (P3) zu XML (P4)
- 2003: TEI P5, seither regelmässige Überarbeitungen
- [TEI-Website](#)

- mehr als nur ein Schema
- Begriffe und Definitionen für hunderte textkritische Phänomene
- Nutzungshinweise für alle Elemente der Module
- System von 23 Modulen, um angepasste Codierungs-Schemata zu generieren
- Neben Modulen: Elemente auch in Klassen organisiert
- Bei 500 Elementen: Übersicht, Modularität, Bearbeitung vereinfacht
 1. Attribut-Klassen: alle Elemente der Klasse haben die gleichen Attribute
 2. Modell-Klassen: alle Elemente der Klasse haben den gleichen Kontext

- konsens-basierte Strategie, um Text-Ressourcen zu organisieren und zu dokumentieren
- eine enzyklopädische Kodifikation aktueller editorischer Praxis
- mehrere Wege führen zum Ziel, Anpassungen möglich; lediglich wichtig: konsistente Verwendung innerhalb eines Projektes

Globale Attribute

- in allen Elementen zugelassen
- für Merkmale, die sich auf fast alles anwenden lassen
- z.B. Identität, Sprache, Aussehen
- @xml:id - "unique identifier"
- @n – Name oder Nummer
- @xml:lang – Sprache (nach ISO Standard)
- @rend und @rendition – visuelles Erscheinungsbild

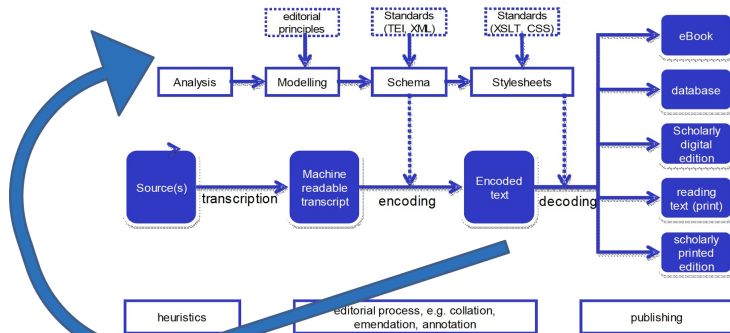
- @xml:id - "unique identifier"
- muss mindestens im gleichen TEI-Dokument, besser projektweit, einzigartig sein
- macht Elemente eindeutig identifizierbar
- Verweis innerhalb eines Dokumentes: @ref="#xml:id"
- Verweis auf ein anderes Dokument: @key="xml:id" (oder @ref, @sameAs)
- Nutzen: Verlinkungs- und Verweissystem dateiintern und -extern, projektintern oder -extern, Synchronisierung bei synoptischer Edition

Tools für TEI

- Roma – Webservice für Schemata
- TEIGarage, Dokument-Konversion nach oder aus TEI
- Publikationstools: *teiPublisher*, *Versioning Machine*, *SADE*, *TEI Boilerplate*, *TEICHI for Drupal* u.v.a.
- *TextGrid* (Arbeitsumgebung), *TEUCHOS* (*msDesc*), *eLaborate* (Transkription), *TILE* (Text/Bild), *CollateX* (Stemmata), u.v.a.
- TEI ist 100% XML: alle gegenwärtigen und zukünftigen XML-Editoren und -Tools können es verarbeiten

Kleine Übung

- Öffnen Sie die Datei `Test_TEI.xml` im Ordner `Material_TEI_Einfuehrung`
- Geben Sie eine öffnende Spitzklammer ein: `<` anstelle von `text` in `<p>` in `<body>` ein. Was passiert?
- Akzeptieren Sie den Vorschlag `<add>` von oXygen und geben sie vor der schliessenden Spitzklammer (`>`) ein Leerzeichen ein. Was passiert?
- Akzeptieren Sie den Vorschlag `@cert` und beobachten Sie was geschieht.
- Lassen Sie den öffnenden Tag mit einer Ziffer beginnen.
- Versuchen Sie ein Element um den Schlusstag zu setzen.
- Versuchen Sie ein leeres Element abzukürzen.



(Quelle: Rehbein/Fritze 2012)

Vorgehen beim praktischen Auszeichnen

- Analysieren: Textstrukturen und -Funktionen
- Modellieren: Welche Elemente werden benötigt (Grundlage für Schema)?
- Transkribieren
- Codieren: Transkription annotieren
- Visualisieren: Stylesheet überführt Code in gewünschtes Format (PDF, HTML)
- Publizieren: Buch, Website, etc.

Analysieren: Vorüberlegungen

- nach Definition von Korpus und Editionszielen
- anhand des Beispiels der Erzählung Niggi Ju
- Albert Bitzius/Jeremias Gotthelf: Kopfarbeiter
- in der Regel: linear verfasste Texte
- Briefe:
 - linear verfasst
 - bis ins 19. Jh. Anweisungen zur Abfassung von Korrespondenz:
Briefsteller
 - Eigenheiten eines Schreibers/einer Schreiberin
 - Zweck der Korrespondenz

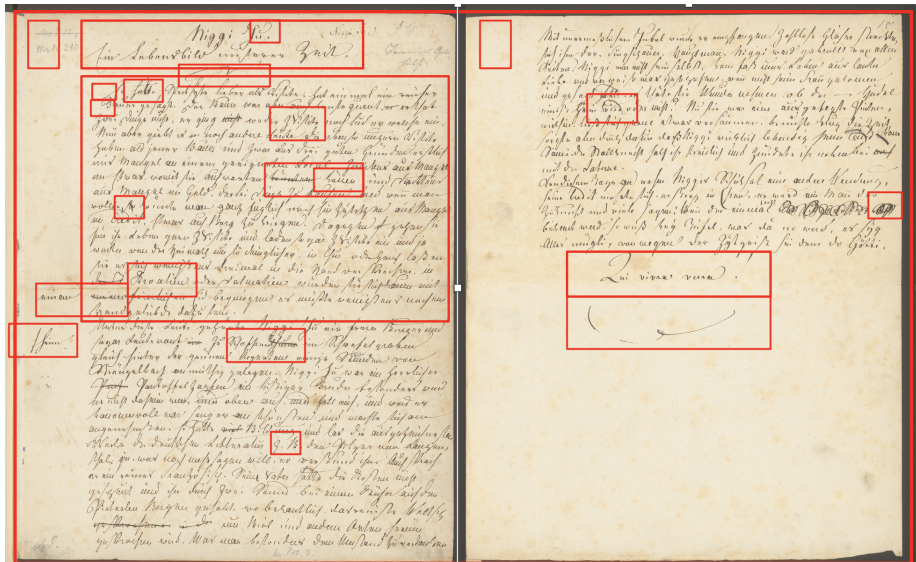
Analysieren: Textstrukturen

- je nach Korpus: einzelnes Dokument, Stichprobe oder alle Manuskripte
- physische Unterteilungen: Blätter, Seiten, Spalten
- welche Textstrukturen lassen sich identifizieren?
- von Gesamtsicht bis zu Absatzebene
- generelle Unterteilungen, Absätze, Aufzählungen
- Überschriften, Zeilen, Spalten, Seiten
- Besondere Textgattungen: Drama, Poesie, Brief

Analysieren: Wort- und Zeichenebene

- Sonderzeichen, Symbole, Abkürzungen
- Hervorhebungen: Schriftartwechsel, Unterstreichungen u.s.w.
- Papier- und/oder Textverluste
- textgenetische Phänomene
- Namen, Orte, Daten
- weiteres je nach Zielen und Richtlinien der Edition: z.B. linguistische Phänomene

Analysieren: Beispieldokument: Niggi Ju



Resultat der Analyse: Niggi Ju

- Grenzelemente (physische Strukturen): `<pb/>`, `<lb/>`
- Textstrukturen: `<div>`, darin `<head>`, `<p>` und `<lg>` mit `<l>`
- Falls Kapitel-Struktur: ein Master-`<div>` um weitere `<div>` alle mit `<head>` (für die Hauptüberschrift und die Kapitelüberschriften)
- innerhalb von `<p>`: `<hi>`, ``, `<add>`, `<subst>`, `<gap>`
- Abkürzung
- Sonderzeichen
- Namen, Orte, etc.
- Potential für Emendation und Normalisierung
- Bleistift: möglicherweise eine fremde Hand

Analyse: Grundlage für das Schema

- Identifikation von Phänomenen und Zuordnung von Elementen (TEI-Website)
- hier braucht Elemente aus den Modulen 2 `teiHeader`, 3 `core`, 4 `textstructures`, 5 `characters and glyphs`, 6 `Verse`, 10 `Manuscript Description`, 11 `Representation of primary sources`, 12 `critical apparatus`, 13 `Names and Dates`, 14 `tables and figures` und 16 `Linking`
- Verfügbare TEI-Schemata nutzen
- Eigene Anpassung: Roma – Webservice für Schemata
- Schema selber entwickeln oder anpassen
- während des Codierens: Neue Phänomene machen Anpassungen nötig

Roma

- Erlaubt die Anpassung von TEI ohne Programmierkenntnisse
 - 1 Generiert ein ODD (One Document Does It All)
 - 2 deckt nicht alle Optionen ab, aber für einfachere Zwecke geeignet
- Auf Basis der vollen TEI Spezifikationen Module oder Attribute ausschliessen (Reduce)

Modules

New Customize Language Modules Add Elements Change Classes Schema Documentation Save Customization

Module removed
drama was successfully removed.

List of TEI Modules			List of selected Modules	
Module name	A short description	Changes		
add analysis	Simple analytic mechanisms		remove tei	
add certainty	Certainty, Precision, and Responsibility		remove header	
add core	Elements common to all TEI documents		remove core	
add corpus	Corpus texts		remove textstructure	
add dictionaries	Dictionaries		remove gaiji	
add drama	Performance texts		remove verse	
add figures	Tables, formulæ, notated music, and figures		remove spoken	
add gaiji	Character and glyph documentation		remove analysis	
add header	The TEI header		remove dictionaries	
add iso-fs	Feature structures		remove msdescription	
add linking	Linking, segmentation, and alignment		remove transcr	
add msdescription	Manuscript Description		remove textcrit	
add namesdates	Names and dates		remove namesdates	
add nets	Graphs, networks, and trees		remove figures	
add spoken	Transcribed Speech		remove corpus	
add tagdocs	Documentation of TEI and other XML markup languages		remove linking	
add textcrit	Critical Apparatus		remove iso-fs	
add textstructure	Default text structure		remove nets	
add transcr	Transcription of primary sources		remove certainty	
add verse	Verses structures		remove tagdocs	

Eigenes Schema definieren

- 1 Metadaten ausfüllen: Schema benennen
- 2 Module und Elemente ausschliessen, ggf. Attribute anpassen
- 3 Schema generieren (Achtung: Compact Syntax generiert einen Fehler!)
- 4 Dokumentation speichern
- 5 Anpassung speichern als ODD für spätere Weiterbearbeitung

List of selected Modules

tei

remove [header](#)

remove [core](#)

remove [textstructure](#)

remove [gajji](#)

remove [verse](#)

remove [msdescription](#)

remove [transcr](#)

remove [textcrit](#)

remove [namesdates](#)

remove [figures](#)

remove [linking](#)

Aufbau eines TEI-Dokumentes

- `<teiHeader>`:
 - 1 Grundlegende Angaben zu Objekt, Edition und Publikation
 - 2 Angaben zum Kontext des Objektes
 - 3 materielle Beschreibung des Objektes
- `<text>`:
 - 1 Strukturen, Layout und Zeichensequenz
 - 2 ggf. weitere Phänomene: Textgenese, linguistische Phänomene
- zusammen: umfassende Repräsentation der Handschrift

Hauptelemente des teiHeader

- `<teiHeader>` mit `<fileDesc>`
- 3 zwingend einzuschliessende Elemente:
 - 1 `<titleStmt>`: Angaben zum Titel, mit:
`<title>`, `<author>`, `<editor>`
 - 2 `<publicationStmt>`: Angaben zur Publikation
 - 3 `<sourceDesc>`: Angaben zur Quelle, auf dem das Dokument beruht
- Optionale Elemente
 - 1 `<encodingDesc>`: Infos zur Codierung, zu den Editionsrichtlinien
 - 2 `<profileDesc>`: Beschreibung des Textprofils (Sprache, Entstehungskontext)
 - 3 `<revisionDesc>`: Bearbeitungen des TEI-Dokuments, mit `<change>`
 - 4 innerhalb von `<sourceDesc>` `<msDesc>`: Manuskript-Beschreibung oder `<p>` mit Text

Handschriftenbeschreibung (HSB)

Warum noch Handschriftenbeschreibung in Zeiten von hochauflösenden Faksimiles?

- Materialität von Handschriften ein rezentes Forschungsthema
- Lange Tradition von Manuskriptbeschreibungen
- keine strukturierte, etablierte Beschreibungs- und Referenzierungskonventionen wie bei Büchern (vgl. ISBN)
- Faksimile nur 2D und nicht universell verfügbar

Modul 10: “Manuscript Description”

- sehr flexibel:
 - 1 geeignet für mittelalterliche Handschriften
 - 2 für moderne Manuskripte
 - 3 und andere text-tragende Objekte
- Vorteile von einheitlicher Erfassung:
 - 1 gegenseitig verständlich
 - 2 Datenaustausch vereinfachen

Dilemma der HSB:

- konfligierende Ziele
 - 1 Tradition der HSB: Bestehende beschreibende Prosa bewahren
 - 2 Zuverlässige Suche und Analyse der Daten ermöglichen
- Ziel des Moduls msDesc
 - wo immer möglich, beides zulassen
 - im Prosatext z.B. Namen, Orte und Organisationen taggen
 - Mischung von strukturierten und unstrukturierten Daten

<teiHeader> für Niggi Ju I

```

<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="tei_sr.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="ms01.xml">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>
          <title>Niggi Ju</title>
          <date>1851</date>
        </title>
        <author xml:id="ms01_JG">Jeremias Gotthelf</author>
        <editor xml:id="ms01_SR">Silvio Raciti</editor>
      </titleStmt>

      <editionStmt>
        <p>born digital: Summer School ZBZH, 20023</p>
      </editionStmt>
      <publicationStmt>
        <p>nicht öffentlich</p>
      </publicationStmt>
      <sourceDesc>
        <msDesc xml:id="Handschrift" n="H">
          <msIdentifier>
            <settlement>Zürich</settlement>
            <repository>Zentralbibliothek Zürich</repository>
            <collection>Handschriften aus der Stadtbibliothek Zürich, Ehemalige Secreta
              der Stadtbibliothek Zürich</collection>
            <idno
              source="https://zbcollections.ch/home/#/content/52c086bb1ca04acbb17882f87abfe1a3"
              >Ms N 210</idno>
          </msIdentifier>
        </msDesc>
      </sourceDesc>
    </fileDesc>
  </teiHeader>

```

<teiHeader> für Niggi Ju II

```

</msConts>
<physDesc>
  <objectDesc>
    <supportDesc>
      <support>
        <p>Papier gelblich, gerippt, dünn, ohne Wasserzeichen.</p>
      </support>
      <extent>
        <measure type="number" quantity="15"/>
        <measure type="size">28 cm x 22,5 cm</measure>
      </extent>
    </supportDesc>
  </objectDesc>
  <handDesc>
    <handNote scribe="eigenhändig" medium="Tinte" xml:id="ms01_Gotthelf">
      <p>Eigenhändig, schwarze Tinte (Albert Bitzius/Jeremias Gotthelf).</p>
    </handNote>
    <handNote scribe="fremdeHände" medium="Bleistift" xml:id="ms01_Fremd">
      <p>Fremde Hand, Bleistift (unbekannt).</p>
    </handNote>
  </handDesc>
</physDesc>
<history>
  <orian>

```

<teiHeader> für Niggi Ju III

```

-----
<history>
  <origin>
    <p>
      <origPlace>
        <placeName type="Gemeinde">Lützelflüh</placeName>
      </origPlace>
    </p>
    <p>
      <origDate when="1851-01-01">1851</origDate>
    </p>
    <provenance>Die Handschrift wurde der Stadtbibliothek Zürich von Jakob
      Escher-Bürkli (1864-1939) geschenkt. Die Bestände der Stadtbibliothek
      wurden 1917 von der neu gegründeten Zentralbibliothek Zürich
      übernommen.</provenance>
  </origin>
</history>
</msDesc>
</sourceDesc>
</fileDesc>
<encodingDesc>
  <charDecl>
    <glyph xml:id="ms01_uo">
      <desc>kleines u mit kleinem o darüber</desc>
      <mapping type="composed">&#x75;&#x366;</mapping>
      <graphic url="chars/umitodarueber.png"/>
    </glyph>
  </charDecl>
</encodingDesc>
<revisionDesc>
  <change n="1" when="2023-08-30" who="SR">Datei erstellt, teiHeader und Text codiert</change>
</revisionDesc>
</teiHeader>

```


<correspDesc>: Metadaten für Briefe

- in <profileDesc> kann correspDesc erfasst werden
- Stellt Metadaten zu Korrespondenz zur Verfügung
- Absender, Empfänger, Beilage, Einlagen, Weiterleitungen, Daten, Orte, vorangehender und folgender Brief
- **CorrespSearch**: Suchmaschine, die auf diese Codierungen zugreift (Vernetzung)

TEI-Module für die Codierung von Handschriften

- Module 3 (Core) und 4 (Default Text Structure) bieten grundlegende Elemente für die Auszeichnung von Handschriften
- Modul 11 (Representation of Primary Sources): weitere Elemente für die Codierung der Textgenese
- Modul 12 (Critical Apparatus): Elemente für die Emulierung eines Apparates in print-Tradition

<text> für Niggi Ju I

```

<provenance>Die Handschrift wurde der Stadtbibliothek Zürich von Jakob
  Escher-Bürkli (1864-1939) geschenkt. Die Bestände der Stadtbibliothek
  wurden 1917 von der neu gegründeten Zentralbibliothek Zürich
  übernommen.</provenance>
</history>
</msDesc>
</sourceDesc>
</fileDesc>
<encodingDesc>
  <charDecl>
    <glyph xml:id="ms01_uo">
      <desc>kleines u mit kleinem o darüber</desc>
      <mapping type="composed">&#x75;&#x366;</mapping>
      <graphic url="chars/umitodarueber.png"/>
    </glyph>
  </charDecl>
</encodingDesc>
<revisionDesc>
  <change n="1" when="2023-08-30" who="SR">Datei erstellt, teilHeader und Text codiert</change>
</revisionDesc>
</teiHeader>
<text>
  <body>
    <div>
      <pb/>
      <head type="title">Niggi <choice><orig>J<g ref="#ms01_uo">u</g></orig><reg>Ju</reg></choice>.</head>
      <head type="subtitle">Ein Lebensbild unserer Zeit.</head>
      <lb type="strich"/>
      <p>text
        <subst n="ueberschreibung">
          <del>ha-gap/</del>
          <add>hatte</add>
        </subst> text <lb/> text <choice>
          <orig>Zvisite</orig>
          <reg resp="#ms01_SR">zur Visite</reg>
        </choice> text <supplied reason="Autorfehlen" resp="#ms01_SR">text, welchen der
          Autor vergessen hat</supplied>
      </p> text
    </div>
  </body>
</text>

```

<text> für Niggi Ju II

```

<subst n="korrektur">
  <del>könten</del>
  <add><choice>
    <orig><sic>könen</sic></orig>
    <corr resp="#ms01_SR">können</corr>
  </choice>
</add>
</subst> text <lb/> text <hi rend="antiqua"><placeName type="Land">
  >Kroatien</placeName></hi> oder <hi rend="antiqua"><placeName
  type="Region">Dalmatien</placeName></hi> text <lb/> text
<subst n="ersetzung">
  <del>mein</del>
  <add place="am linken Rand">einem</add>
</subst> text</p>
<p>text
<subst n="ersetzung">
  <del>Schoppenthurm</del>
  <add place="am linken Rand" type="Einfüguungszeichen">Schoppenheim</add>
</subst> text <lb/>
<choice>
  <orig resp="#ms01_JG"><abbr>z. B.</abbr></orig>
  <expan resp="#ms01_SR">zum Beispiel</expan>
</choice> text <lb/> text <p/> <unclear>text</unclear>
<app>
  <rdg hand="#ms01_Gotthelf">köne</rdg>
  <rdg hand="#ms01_Fremd">köffe</rdg>
</app> text</p>
<p>text
<subst n="korrektur">
  <del><gap reason="unlesbar" agent="Streichung" extent="Wort"/></del>
  <add>bekant</add>
</subst> text</p>
<lg rend="zentriert">
  <lb><hi rend="antiqua">Qui vivra verra.</hi></lb>
</lg>
<lb type="strich"/>
<p/>
</div>

```

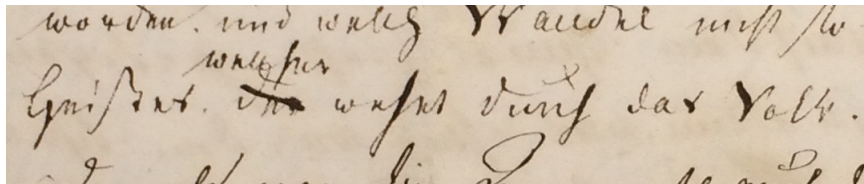
Nur teilweise lesbarer Text: <unclear>

- <damage>: Beschädigung des Textes oder Textträgers
- Für teilweise oder schlecht lesbaren Text: <unclear>
- Unleserlicher Text: <gap>
- <supplied>: hinzugefügter Text
- Attribute:
 - @reason: Grund, bspw. fehlender Tintenauftrag oder Papierverlust
 - @extent: Ausmaß
 - @unit: Einheit (mm, cm, char, lines, etc.)
 - @cert (certitude): Sicherheit, bspw. in Prozent, oder Grad (Typologie, Schema)
 - @resp (responsibility): Verantwortlicher für die Lesart und die Angabe des Grundes und der Sicherheit (Header: Liste der Editoren)

Korrekturen

- Typische Änderungen am Text aus core-Modul
 - 1 ``: Streichung oft mit Attribut `@prev` oder `@next`
 - 2 `<add>`: Hinzufügung oft mit Attributen `@place` und `@prev/next`
- Was tun, wenn ein Zusammenhang zwischen Tilgung und Ergänzung besteht?

Beispiel: Ersetzung



- "der" wurde durchgestrichen
- "welcher" wurde hinzugefügt
- vielleicht so: `der<add>welcher</add>` ?

Ersetzung + Räumlichkeit

- "der" wurde durch "welcher" ersetzt
- "welcher" wurde interlinear darüber geschrieben
- Code für Ersetzung:
- `<subst n="Ersetzung">`
`der`
`<add place="über der Zeile">welcher</add>`
`</subst>`

Sonderzeichen

- Sonderzeichen in alten Drucken oder Handschriften können Probleme bereiten
- vieles kann direkt über die Tastatur eingegeben werden
- sofern ein UTF-8-Zeichen (vgl. **Unicode-Tabellen**)
 - 1 Menü: >Bearbeiten > Einfügen aus Zeichentabelle
 - 2 &# mit dem Hexadezimalcode des gewünschten Zeichens und einem ; am Ende
 - 3 ℔ wird als Pfundzeichen ausgegeben: ℥

Sonderzeichen mit `<g>` “gaiji”

- Wenn kein UTF-8-Zeichen
- die Darstellung des Zeichens als zu wenig akkurat angesehen wird
- oder wenn gewisse nachfolgende Verwendungen, z. B. indexing, nicht erschwert werden sollen
 - 1 Zeichen definieren in `<teiHeader>` unter `<encodingDesc>` und `<charDecl>`: `<glyph xml:id="x123">`
 - 2 im Text mit `<g ref="#x123">x</g>` verwenden

Sonderzeichen

```
<encodingDesc>
  <charDecl>
    <glyph xml:id="ms01_uo">
      <desc>kleines u mit kleinem o darüber</desc>
      <mapping type="composed">&#x75;&#x366;</mapping>
      <graphic url="chars/umitodarueber.png"/>
    </glyph>
  </charDecl>
```

```
</encodingDesc>
```

im Text:

```
<g ref="#ms1_uo"/> oder <g ref="#ms1_uo">u</g>
```

```
<choice>
```

```
  <orig>J<g ref="#ms01_uo">u</g></orig><
```

```
reg>Ju</reg>
```

```
</choice>
```

<choice>: Abkürzungen, Emendationen, Normalisierungen

```
<choice>  
  <orig resp="#ms01_JG"><abbr>z. B.</abbr></orig>  
  <expan resp="#ms01_SR">zum Beispiel</expan>  
</choice>
```

- <choice>
- für Abkürzungen:
- <abbr>: abgekürzte Form (im Manuskript)
- <expan>: ausgeschriebene Form (zusätzlich)
- für Auslassungen bzw. Verschleifungen
- <am> abgekürzte Form (im Manuskript)
- <ex>: ausgeschriebene Form (zusätzlich)

Emendationen und Normalisierungen

- offensichtliche Fehler
- markieren mit `<sic>`
- emendieren? Wenn ja, dann sowohl Zeichenfolge im Original als auch korrigierte Version festhalten
- hierfür kann man `<choice>` mit `<orig>` und `<corr>` benutzen

- so:

```
<choice>
```

```
    <orig><sic>könen</sic></orig>
```

```
    <corr resp="#ms01_SR">können</corr>
```

```
</choice>
```

- Analog für Normalisierungen:
 - `<orig>`: originale Form
 - `<reg>`: normalisierte Form
- Visualisierung als Apparateintrag

Modul 12: Elemente zur Codierung von fremden Händen/Varianten

```
</choice> text </or> text </pr> text  
<app>  
  <rdg hand="#ms01_Gotthelf">köne</rdg>  
  <rdg hand="#ms01_Fremd">köme</rdg>  
</app> text</p>
```

- jede Stelle mit Texteingriffen (egal welcher Hand)/mit abweichenden Varianten muss so codiert werden
- Präferierte Variante kann angegeben werden: <lem> statt <rdg>
- Parallelstellen Edition (parallele Segmentation):
 - klassischer Apparat
 - Print-orientiert
 - aber auch die Grundlage für neue Darstellungsmöglichkeiten
 - z.B. **Versioning Machine 5.0**

▷▷Niggi Ju◁▷1851◁◁
Jeremias Gotthelf

Silvio Raciti

bom digital: Summer School ZBZH, 20023

nicht öffentlich

▷Zürich◁▷Zentralbibliothek Zürich◁▷Handschriften aus der Stadtbibliothek Zürich, Ehemalige Secreta der Stadtbibliothek Zürich◁
Ms N 210

▷▷▷▷Albert Bitzius◁▷Jeremias Gotthelf◁◁▷Niggi Ju. Ein Lebensbild unserer Zeit.◁◁◁

Papier gelblich, gerippt, dünn, ohne Wasserzeichen.

▷measure◁▷28 cm x 22,5 cm◁

Eigenhändig, schwarze Tinte (Albert Bitzius/Jeremias Gotthelf).

Fremde Hand, Bleistift (unbekannt).

▷▷Lützelflüh◁◁

▷1851◁

Die Handschrift wurde der Stadtbibliothek Zürich von Jakob Escher-Bürkli (1864-1939) geschenkt. Die Bestände der Stadtbibliothek wurden 1917 von der neu gegründeten Zentralbibliothek Zürich übernommen.



kleines u mit kleinem o darüber

Datei erstellt, teilHeader und Text codiert

1. Niggi Ju

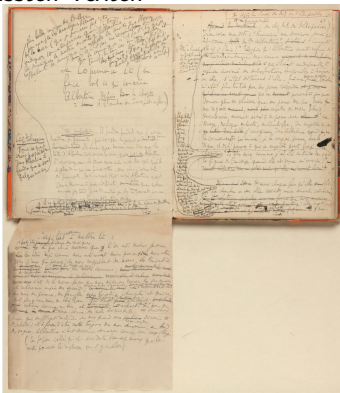
1. Ein Lebensbild unserer Zeit.

hatte text
Zur Visite zur Visite text, welchen der Autor vergessen hat
könnten können
Kroatien oder Dalmatien text
mein einem text
Schoppenturm Schoppenheim text
z. B. zum Beispiel text
text könie könie text
bekam text
Qui vivra verria.

- siehe den morgigen Vortrag zum **teiPublisher!**
- Gotthelf-Edition: im Moment nur print
- dHKG soll demnächst online gehen

Schluss

- TEI: einfache aber mächtige Werkzeuge zur Erstellung einer digitalen Repräsentation
- Probleme treten auf
 - 1 wenn die strikt hierarchische Struktur nicht eingehalten werden kann
 - 2 bei nicht-linearen Textstrukturen (morgiger Workshop)
 - 3 bei nicht-linear verfassten Texten



- Für solche Fälle TEI für Handschriften bereits weiterentwickelt:
- Prototyp eines Modell für genetische Edition (Elena Pierazzo u.a.):
Autour d'une séquence et des notes du Cahier 46: enjeu du codage dans les brouillons de Proust
- Evidenz für funktionierendes Zusammenspiel von Konsortium und Community

- Fragen
- Diskussion