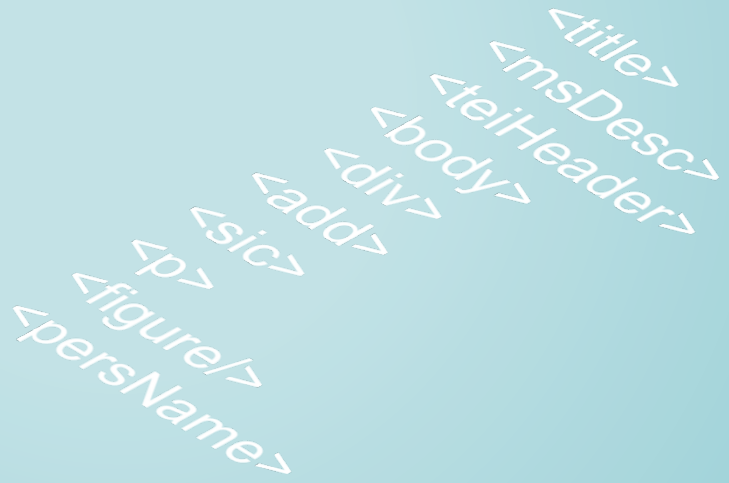
XML oder nicht XML – das ist hier die Frage

Perspektivwechsel bei der Langzeitarchivierung von digitalen Editionen (Düsseldorf; 17.9.2018)

Thomas Stäcker (thomas.staecker@ulb-tu-darmstadt.de)

TECHNISCHE UNIVERSITÄT DARMSTADT



ULB

Was ist eine digitale Edition?

TECHNISCHE UNIVERSITÄT DARMSTADT

Einige Thesen zur Frage der Nachhaltigkeit der Edition als „Dokument"

- Eine digitale Edition ist grundsätzlich anders als eine gedruckte Edition
- Eine digitale Edition ist keine digitalisierte Edition
- PDF ist nicht nachhaltig, auch nicht als PDF/A, weil es die gedruckte Edition lediglich simuliert.
- Eine digitale Edition findet ihre endgültige Form nicht in der Präsentationsoberfläche (*viewport*)
- Eine digitale Edition ist „Software"
- Eine digitale Edition ist in ihren Komponenten vollständig bestimm- und beschreibbar

ULB

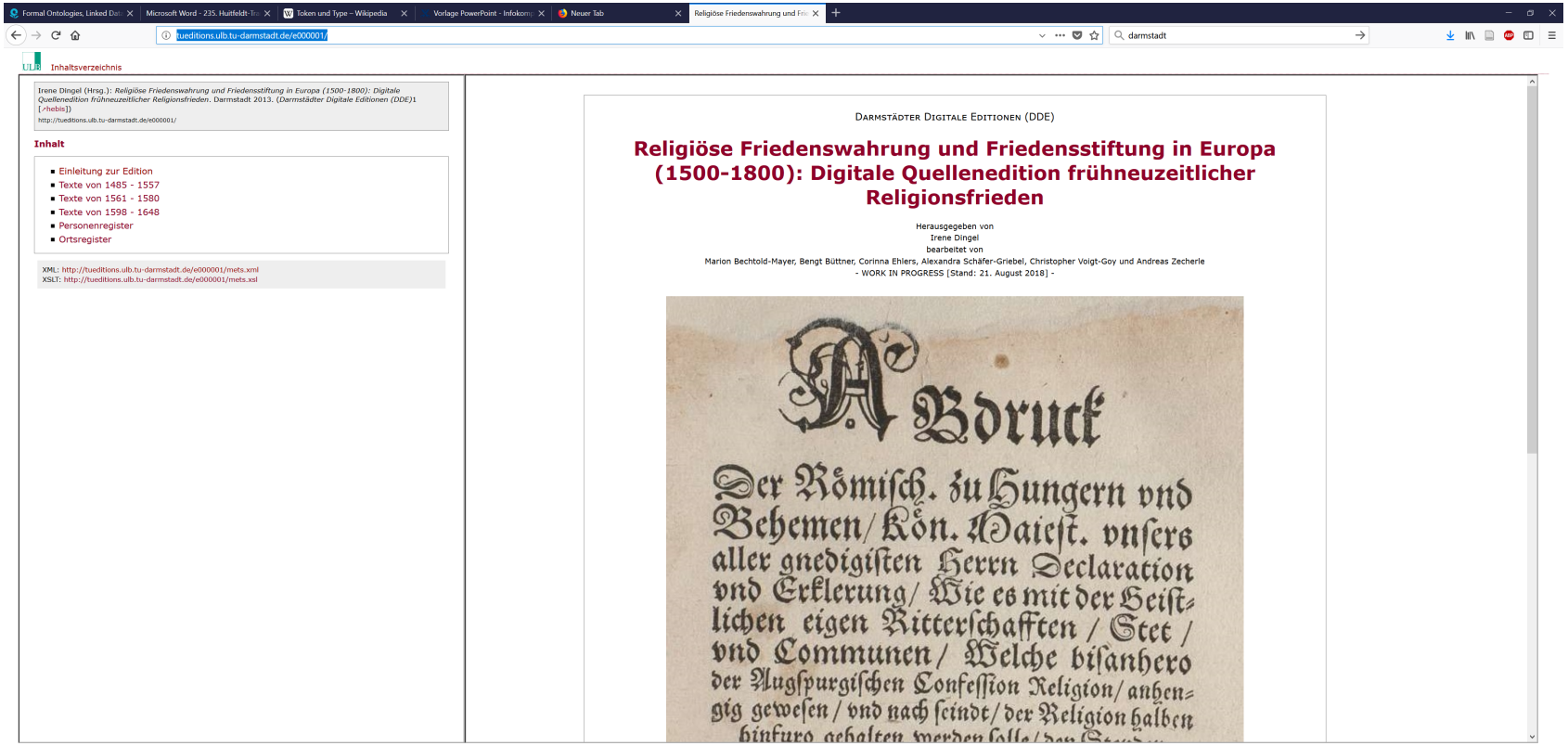
Worüber ich nicht rede

TECHNISCHE UNIVERSITÄT DARMSTADT

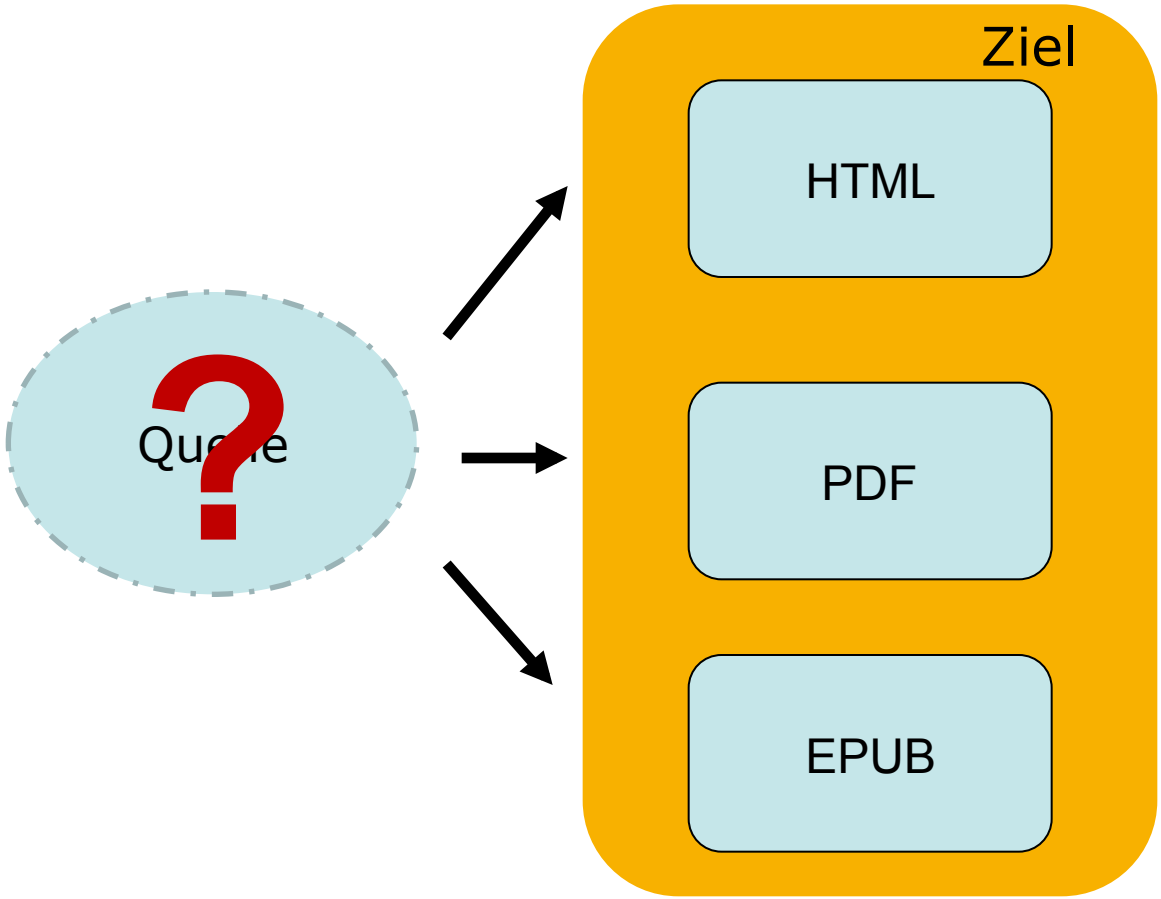
- Die Edition als „intellektuelles oder sprachliches Werk“
- Allgemeine Editionstheorie oder Textkritik
- Anwendung der TEI für Editionsphänomene

ULB

Was ist eine digitale Edition?



Was ist eine digitale Edition?

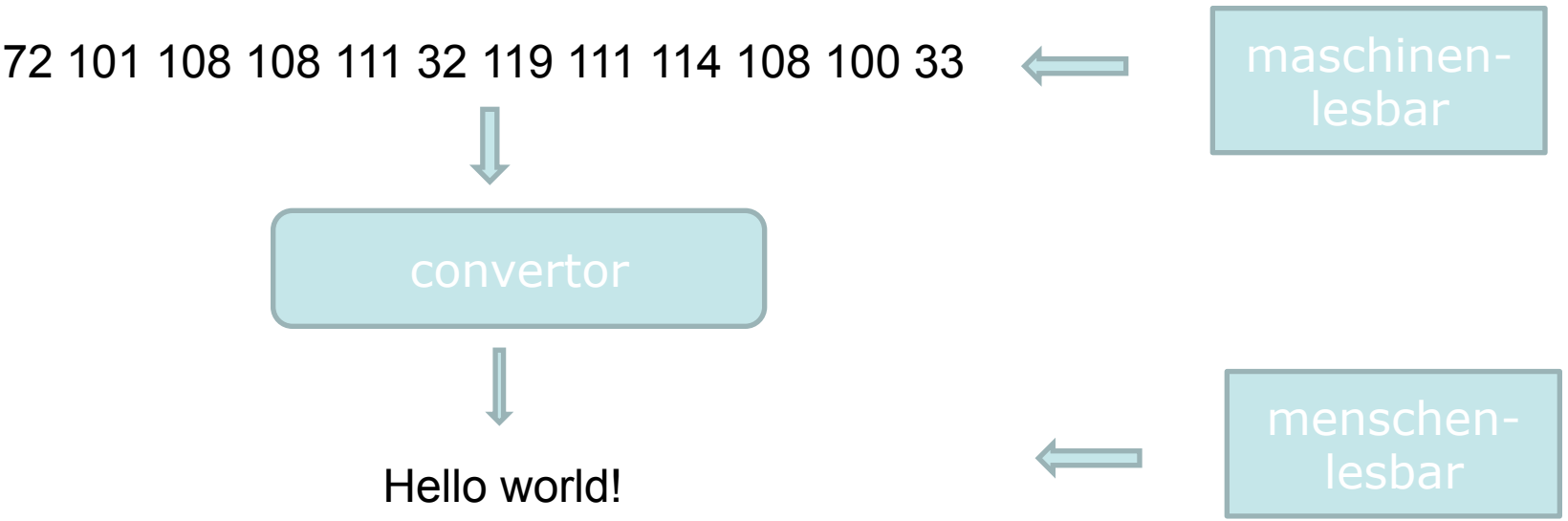


Was ist eine digitale Edition?

TECHNISCHE UNIVERSITÄT DARMSTADT

TEXT

Text = Abfolge von *code points* gemäß UNICODE

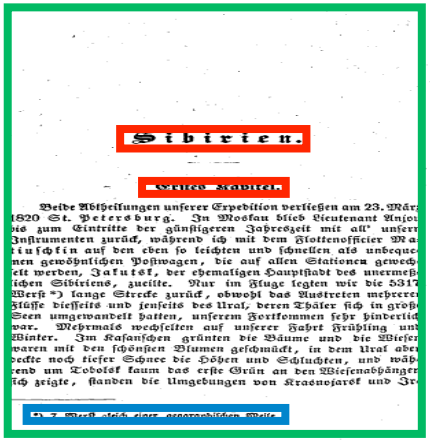


ULB

Was ist eine digitale Edition?

TECHNISCHE UNIVERSITÄT DARMSTADT

Struktur

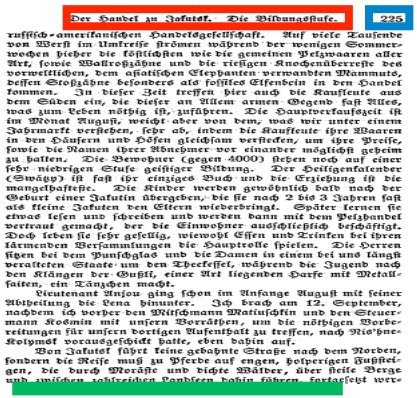


Sibirien.

Erstes Kapitel.

Beide Abtheilungen unserer Expedition verließen am 23. März 1820 St. Petersburg. In Mosfau blieb Lieutenant Anjou bis zum Eintritte der günstigeren Jahreszeit mit all' unsern Instrumenten zurück, während ich mit dem Flottenofficier Ma... in Schkin auf den eben so leichten und schnellen als unbequemen gewöhnlichen Postwagen, die auf allen Stationen gewechselt werden, Jakutsk, der ehemaligen Hauptstadt des unermeßlichen Sibiriens, zueilte. Nur im Fluge legten wir die 5317 Werst *) lange Strecke zurück, obwohl das Austreten mehrerer Flüsse diesseits und jenseits des Ural, deren Thäler sich in große Seen umgewandelt hatten, unserem Fortkommen sehr hinderlich war. Mehrmals wechselten auf unserer Fahrt Frühling und Winter. Im Kasanschen grünten die Bäume und die Wiesen prangten mit den schönsten Blumen geschmückt, in dem Ural aber deckte noch tiefer Schnee die Höhen und Schluchten, und während um Tobolsk kaum das erste Grün an den Wiesenabhängen sich zeigte, standen die Umgebungen von Krasnojarsk und Ir...

*) 1 Werst gleich einer geographischen Meile.

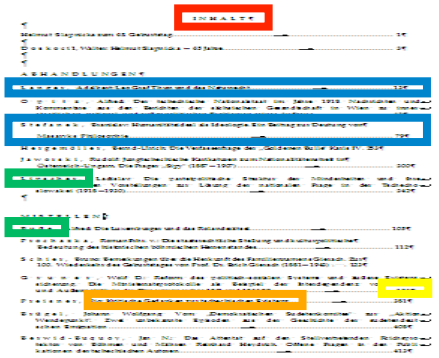


Der Handel zu Jakutsk. Die Bildungsstufe. 225

russisch-amerikanischen Handelsgesellschaft. Auf viele Tausende von Werst im Umkreise strömen während der wenigen Sommerwochen hieher die köstlichsten wie die gemeinen Pelzwaaren aller Art, sowie Wallroßzähne und die riesigen Knochenüberreste des vorweltlichen, dem asiatischen Elephanten verwandten Mammuts, dessen Stoßzähne besonders als fossiles Elfenbein in den Handel kommen. In dieser Zeit treffen hier auch die Kaufleute aus dem Süden ein, die dieser an Allem armen Gegend faßt Alles, was zum Leben nöthig ist, zuführen. Die Hauptverkaufszeit ist im Monat August, weicht aber von dem, was wir unter einem Jahrmarkte verstehen, sehr ab, indem die Kaufleute ihre Waaren in den Häusern und Höfen gleichsam verstecken, um ihre Preise, sowie die Namen ihrer Abnehmer vor einander möglichst geheim zu halten. Die Bewohner (gegen 4000) stehen noch auf einer sehr niedrigen Stufe geistiger Bildung. Der Heiligenkalender (Swätzy) ist fast ihr einziges Buch und die Erziehung ist die mangelhafteste. Die Kinder werden gewöhnlich bald nach der Geburt einer Jakutin übergeben, die sie nach 2 bis 3 Jahren fast als kleine Jakuten den Eltern wiederbringt. Später lernen sie etwas lesen und schreiben und werden dann mit dem Pelzhandel vertraut gemacht, der die Einwohner ausschließlich beschäftigt. Doch leben sie sehr gesellig, wiewohl Essen und Trinken bei ihren lärmenden Versammlungen die Hauptrolle spielen. Die Herren sitzen bei dem Punschglas und die Damen in einem bei uns längst veralteten Staate um den Theekessel, während die Jugend nach den Klängen der Gusli, einer Art liegenden Harfe mit Metallsaiten, ein Tänzchen macht.

Lieutenant Anjou ging schon im Anfange August mit seiner Abtheilung die Lena hinunter. Ich brach am 12. September, nachdem ich vorher den Wilitschnann Matiuschkin und den Steuermann Kosmin mit unsern Vorräthen, um die nöthigen Vorbereitungen für unsern dortigen Aufenthalt zu treffen, nach Nis'hne-Kolymsk vorausgeschickt hatte, eben dahin auf.

Von Jakutsk führt keine gebahnte Straße nach dem Norden, sondern die Reise muß zu Pferde auf engen, holperigen Fußsteigen, die durch Moräste und dichte Wälder, über steile Berge und zwischen felsreichen Pandiren dahin führen, fortgesetzt wer-



INHALTE

- Überschriften
- Fußnoten
- Satzspiegel
- Schrifttypen

- Kolumnentitel
- Seitenzahlen
- Kustoden

- Inhaltsverzeichnis
- Bibliographie
- Register

ULB

Was ist eine digitale Edition?

TECHNISCHE UNIVERSITÄT DARMSTADT

Struktur / *Entitäten*

- Personen
- Orte
- Körperschaften
- Zeit
- Sachen
- Werke / Zitate

ULB

Was ist eine digitale Edition?

TECHNISCHE UNIVERSITÄT DARMSTADT

Regeln („Modell“ der Edition)

- Beschreibung möglicher Elemente und Strukturen
- Beschreibung der Bedeutung von Elementen und Strukturen
- Beschreibung möglicher Inhalte
- Beschreibung von Anzeigemöglichkeiten („process model“)

ULB

Was ist eine digitale Edition?

TECHNISCHE UNIVERSITÄT DARMSTADT

Layout

- Aussehen des Textes (Grapheme, Auszeichnungen)
- Visualisierung von Strukturen und Bedeutung
- Ästhetik der Darstellung
- Bedienungsdesign

ULB

Was ist eine digitale Edition?

TECHNISCHE UNIVERSITÄT DARMSTADT

Scripte zum

- Anzeigen / Visualisieren
- Navigieren
- Suchen
- Manipulieren
- Extrahieren
- Mischen
- …

ULB

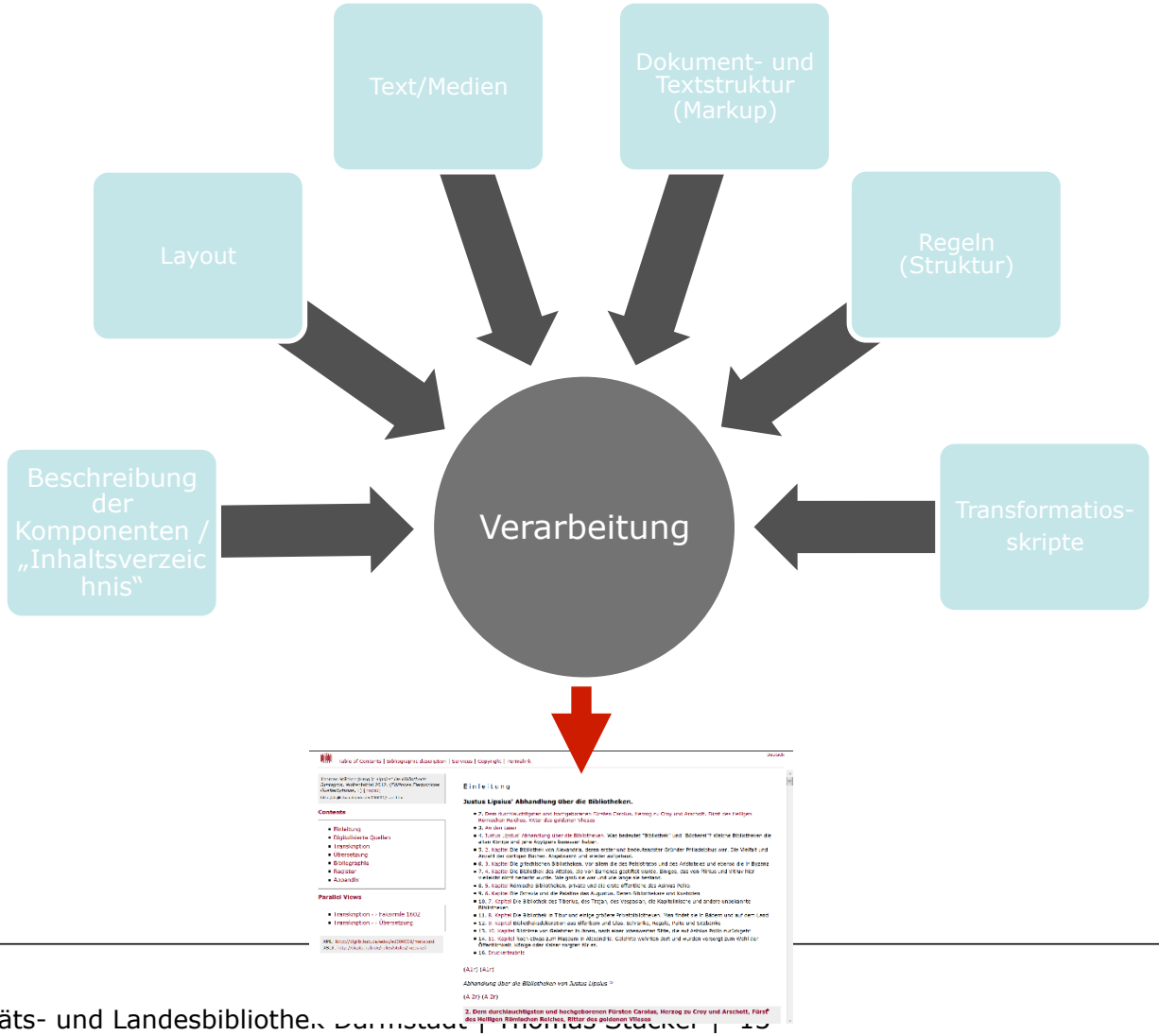
Was ist eine digitale Edition?

TECHNISCHE UNIVERSITÄT DARMSTADT

Metadaten/Verzeichnis der Komponenten

- Metadaten (deskriptiv, strukturell, administrativ, technisch)
- Beschreibung der vorhandenen Dateien
- Beziehungen der Dateien untereinander und zu externen Ressourcen
- Bearbeitungsvorschriften für die Dateien

Komponenten der digitalen Edition



Wie schreibt man eine digitale Edition?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Text - > UNICODE
- Struktur -> XML, TAG? (keine eigene Serialisierung), JSON?
- Regeln -> XML Schemasprachen, Shape Expressions (RDF?)
- Layout -> CSS/Javascript
- Skripte -> XSLT, XQuery, Javascript, Python, PHP, PERL …
- Metadaten/Verzeichnis -> teiCorpus/teiHeader, METS, PREMIS (XML)

ULB

XML oder Nicht-XML

TECHNISCHE UNIVERSITÄT DARMSTADT

Kritikpunkt OHCO

- **Text content ist nicht XML-content**

```
<div>
        <p>Der erste Satz ist:</p>
        <p>Hello World!</p>
</div>
```

≙

```
<div>
    <p sort="2">Hello World!</p>
    <p sort="1">Der erste Satz ist:</p>
</div>
```

- (CONCUR)
- stand-off markup
- start/end-milestones
- …

ULB

XML oder Nicht-XML

TECHNISCHE UNIVERSITÄT DARMSTADT

```
<choice>
        <sic>Hello Wörlt!</sic>
        <corr>Hello World!</corr>
</choice>
```

≙

```
<choice>
        <corr>Hello World!</corr>
        <sic>Hello Wörlt!</sic>
</choice>
```


ULB

XML oder Nicht-XML

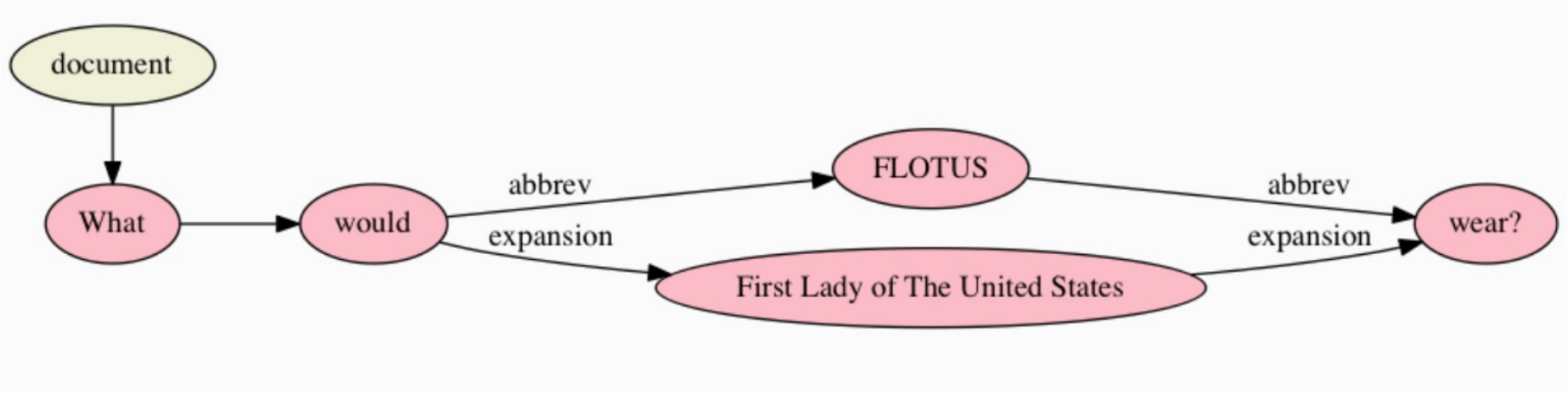
TECHNISCHE UNIVERSITÄT DARMSTADT

```
<choice>
        <sic>Hello Wörlt!</sic>
        <corr>Hello World!</corr>
</choice>
```

```
<content>
 <alternate minOccurs="2"
  maxOccurs="unbounded">
  <classRef key="model.choicePart"/>
  <elementRef key="choice"/>
 </alternate>
</content>
```


ULB

XML oder Nicht-XML



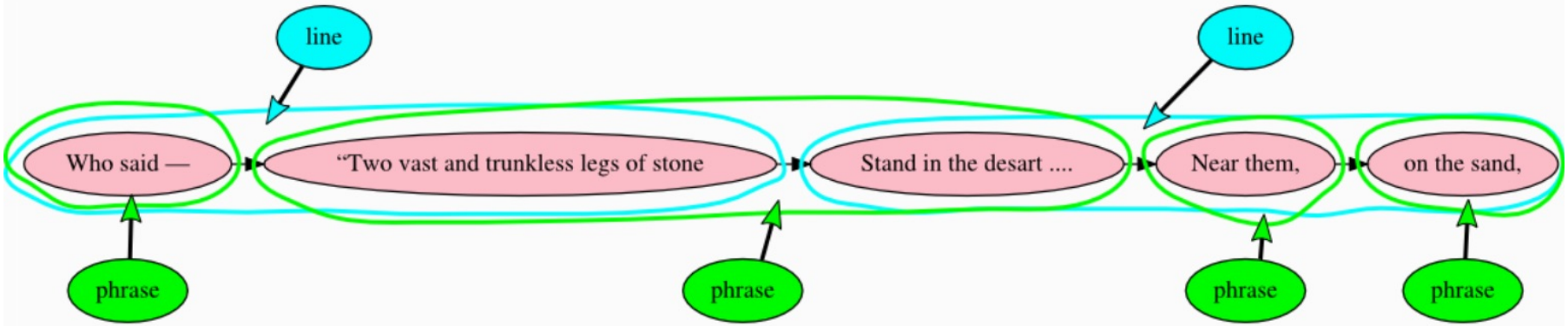
XML oder Nicht-XML

TECHNISCHE UNIVERSITÄT DARMSTADT

```
<elementSpec ident="choice" mode="change">
 <model predicate="sic and corr"
  behaviour="alternate">
  <param name="default" value="corr"/>
  <param name="alternate" value="sic"/>
 </model>
</elementSpec>
```


ULB

XML oder Nicht-XML



<seg type=„phrase “xml:id=„line1a“>Who said</seg><seg type=„phrase“xml:id=„line1b“>„Two vast and trunken legs of stone <seg type=„phrase“ xml:id=„line2a“>Stand in the desart … </seg><seg type=„phrase“ xml:id=„line2b“>Near them<seg><seg type=„phrase“xml:id=„line2c“> on the sand<seg>

<join target=„#lnie1a #line1b“ scope=„branches“ result=„l“>
<join target=„#lnie2a #line2b“ #line2c“ scope=„branches“ result=„l“>

Nachhaltigkeit

TECHNISCHE UNIVERSITÄT DARMSTADT

Eine nachhaltige digitale Edition

- besteht aus einer Anzahl von „Dateien“ (=digitale Dokumente), die mit Metadatenstandards (XML) beschreiben werden,
- stellt alle Dateien frei unter offenen Lizenzen zur Verfügung,
- existiert als ein Prozess von Eingabe, Verarbeitung und Ausgabe (EVA), nicht als resultierende Datei,
- nutzt (pragmatisch) XML/TEI für die strukturelle Textbeschreibung (Markup; nach wie vor das Mittel der Wahl, wenn eine das Textmodell abbildende Schemadatei vorliegt),
- beschreibt das *process model* möglicher Verarbeitungen,
- beschreibt das Layout mittels geeigneter Standards (insb. CSS),
- bietet Schnittstellen (APIs) an für verschiedene Nutzungen der Edition, die dem REST Modell folgen (als selbsterkärende HTTP-basierte Schnittstelle)
- Ist per PI (DOI, URN, PURL…) zitierbar

ULB

Nachhaltigkeit

TECHNISCHE UNIVERSITÄT DARMSTADT

Kernthesen

- *Nachhaltigkeit von Editionen kann durch Bewahrung von zweidimensionalen Repräsentationen, die Resultat von Prozessschritten sind, nicht erreicht werden. Eine digitale Edition besteht vielmehr in der Summe der Verarbeitungsmöglichkeiten, die in ihrem Modell vollständig beschrieben werden können.*
- *Jeder Versuch, eine konkrete technische Realisierung einer Edition zu bewahren, ist eo ipso zum Scheitern verurteilt.*
- *Die vollständige Beschreibung aller Komponenten der Edition in maschinenlesbarer und standardisierter Form ist wichtige Voraussetzung für ihre Nachhaltigkeit.*