



# MAGAZIN FÜR DIGITALE EDITIONSWISSENSCHAFTEN

*Herausgegeben vom Interdisziplinären Zentrum  
für Editionswissenschaften  
der Friedrich-Alexander-Universität Erlangen-Nürnberg*

---

**Vorstand:**

BORIS DREYER  
GÜNTHER GÖRZ  
ANDREAS NEHRING  
KLAUS MEYER-WEGENER

**Board:**

FLORIAN KRAGL  
KLAUS MEYER-WEGENER  
WOLFGANG WÜST

1 / 2015



FAU University Press  
Magazin für digitale Editionswissenschaften  
ISSN 2364-0855

Herausgeber:  
Interdisziplinäres Zentrum für Editionswissenschaften  
Prof. Dr. Boris Dreyer (Sprecher)  
Universität Erlangen-Nürnberg  
Department Geschichte  
Alte Geschichte  
Kochstr. 4, Postfach 8  
D-91054 Erlangen

# EINE DIGITALE TRANSKRIPTION DES ›DEUTSCHEN PTOLEMAEUS‹

GÜNTHER GÖRZ / BETTINA LINDNER

## 1 Zweck des Vorhabens

Will man Aufschluss über das geographische Wissen erhalten, über das gebildete Humanisten kurz vor der Entdeckung Amerikas verfügten, so ist der sog. ›Deutsche Ptolemaeus‹ (GKW M36390, Hain 13542) eine unverzichtbare Quelle. Vermutlich im Umfeld des Nürnberger Humanistenkreises<sup>1</sup> entstanden und in bemerkenswert schlechter Qualität gedruckt, sind nur zwei Exemplare dieser Inkunabel erhalten. Der Text ist bis auf ein lateinisches Widmungsgedicht in einer nürnbergisch-schlesischen Varietät des Frühneuhochdeutschen verfasst, wobei der Druck eine für die Zeit nicht untypische orthographische Vielfalt aufweist. 1910 wurde von Josef Fischer in Strassburg ein kommentiertes, allerdings nicht fehlerfreies Faksimile veröffentlicht.<sup>2</sup> Das Exemplar der Bayerischen Staatsbibliothek wurde digitalisiert<sup>3</sup>.

Es gibt also gute Gründe, dieses wissenschaftshistorisch interessante Werk in Form einer digitalen Transkription im Web bereitzustellen. Da es dabei auch etliche Herausforderungen zu bewältigen gibt, eignet es sich gut als praktisches Beispiel zum Erfahrungsgewinn in einem Bereich der Digital Humanities. Neben den Examensarbeiten von Barbara Ries<sup>4</sup> und Sarah Schulz<sup>5</sup> wurde ein erheblicher Teil der Arbeiten im Rahmen studentischer Übungsprojekte geleistet. Diese resultierten in zwei Web-Präsentationen der Transkription. Die erste ist eine experimentelle – aus bildrechtlichen Gründen nicht frei zugängliche – Internet-Präsentation. Weiterhin konnte mit Unterstützung des Max-Planck-Instituts für Wissenschaftsgeschichte, Berlin, die Transkription zusammen mit

- 1 Siehe z. B. Medizin, Jurisprudenz und Humanismus in Nürnberg um 1500. Akten der gemeinsam mit dem Verein für Geschichte der Stadt Nürnberg, dem Stadtarchiv Nürnberg und dem Bildungszentrum der Stadt Nürnberg am 10./11. November 2006 und 7./8. November 2008 in Nürnberg veranstalteten Symposien, hg. von Franz Fuchs, Wiesbaden 2010 (Pirckheimer-Jahrbuch für Renaissance- und Humanismusforschung 24).
- 2 Der ›Deutsche Ptolemäus‹ aus dem Ende des XV. Jahrhunderts (um 1490), in Faksimiledruck hg. mit einer Einl. von Josef Fischer, Straßburg 1910 (Drucke und Holzschnitte des XV. und XVI. Jahrhunderts in getreuer Nachbildung). Siehe auch <https://archive.org/details/derdeutscheptole00fish>.
- 3 Die Seitenbilder sind einsehbar unter [http://daten.digitalle-sammlungen.de/bsb00001767/image\\_1](http://daten.digitalle-sammlungen.de/bsb00001767/image_1).
- 4 Barbara Ries: Der ›Deutsche Ptolemäus‹. Vorstudien zu einer digitalen Edition, Magisterarbeit, Universität Erlangen-Nürnberg, Philosophische Fakultät II, Erlangen, Februar 2006.
- 5 Sarah Schulz: Zwischen Latein und Volkssprache – Der Deutsche Ptolemäus auf dem Weg zu einer vielschichtigen digitalen Edition durch die Implementierung von XLink, Bachelorarbeit, Universität Erlangen-Nürnberg, Department für Germanistik und Komparatistik und Department Informatik, Erlangen, Juli 2010.

den gemeinfreien Seitenbildern des Fischerschen Faksimiles im Rahmen von ECHO (European Cultural Heritage Online) veröffentlicht werden<sup>6</sup>.

## 2 Was ist der ›Deutsche Ptolemaeus‹?

Beim ›Deutschen Ptolemaeus‹ handelt es sich um eine kleine Kosmographie aus dem Ende des 15. Jahrhunderts. Obwohl der Autor sie als deutsche Übersetzung der ›Geographia‹ des Ptolemaeus bezeichnet, darf sie als eigenständiges Werk gelten, auch wenn sie sich erkennbar an Ptolemaeus orientiert. Besonderen Wert erhält sie durch die beigelegte, auf der Ulmer Ptolemaeus-Ausgabe basierenden Weltkarte. Eines der zwei bekannten Exemplare befindet sich in der Public Library in New York und eines in der Bayerischen Staatsbibliothek in München. Ein dritter Druck, der in Berlin gewesen sein soll, gilt seit dem zweiten Weltkrieg als verschollen.

### 2.1 Zum Stand der Forschung

In der Forschung ist der ›Deutsche Ptolemaeus‹ bisher eher vernachlässigt worden. Die Ergebnisse von Michael Herkenhoff<sup>7</sup> und Barbara Ries sind jüngerer Datums; siehe auch Brévarts Lexikonartikel<sup>8</sup>. Daneben erschienen bereits 1910 die von Joseph Fischer herausgegebene Faksimileausgabe<sup>9</sup> und bis in die 1960er Jahre hinein einige weitere grundlegende Veröffentlichungen, die sich mit unterschiedlichen Fragen zum ›Deutschen Ptolemaeus‹ beschäftigen<sup>10, 11</sup>.

Die Frage der Autorenschaft ist nach wie vor ungeklärt. Das Buch selbst gibt keinerlei Auskunft über seinen Verfasser oder Entstehungsort. In der Forschung ist man sich seit Fischer aber einig, dass die Hervorhebung der Städte Krakau, Neisse und Nürnberg auf eine besondere Beziehung des Autors zu diesen Orten schließen lässt. So vermutet man Neisse als Geburts- und Krakau als möglichen Studienort des Verfassers.

Unsicher ist auch, in welchem Jahr der ›Deutsche Ptolemaeus‹ erschien. Da der Verfasser sowohl die 1486 erstmals gedruckte ›Peregrinatio in terram sanctam‹ des Bernhard von Breydenbach als auch die im selben Jahr von Johannes Reger besorgte Ulmer Ptole-

6 <http://echo.mpiwg-berlin.mpg.de/MPIWG:X3Y43643>.

7 Michael Herkenhoff: Die Darstellung außereuropäischer Welten in Drucken deutscher Offizinen des 15. Jahrhunderts, Berlin 1996.

8 Francis Brévart: Ptolemäus (›Cosmographia Phtolomei Dewtsch‹), in: Die deutsche Literatur des Mittelalters. Verfasserlexikon, 2. Aufl. hg. von Kurt Ruh u. a., Bd. 7, Berlin 1989, S. 899–902.

9 Fischer [Anm. 2].

10 Erwin Rosenthal: The German Ptolemy and its World Map, in: Bulletin of the New York Public Library 48/2 (1944), S. 135–147 plus Karte.

11 Walther Matthey: Wurde der ›Deutsche Ptolemäus‹ vor 1492 gedruckt?, in: Gutenberg-Jahrbuch 36 (1961), S. 77–87.

maeus-Ausgabe verwendete, geht man in der Forschung davon aus, dass das Werk frühestens 1486 geschrieben sein kann. Eine Abschrift des ›Deutschen Ptolemaeus‹ aus dem Nachlass Johann Schöners von 1509 markiert das späteste mögliche Erscheinungsjahr<sup>12</sup>.

Zweifelsfrei wurde der ›Deutsche Ptolemaeus‹ mit einer Type des Nürnberger Frühdruckers Georg Stuchs gedruckt. Allerdings ergaben paläotypische Untersuchungen, dass es sich dabei nicht um die Type 14 handelte – wie Joseph Fischer noch annahm<sup>13</sup> –, sondern dass die kleine Kosmographie sehr viel wahrscheinlicher mit der Type 13 frühestens Mitte der 90er Jahre gedruckt wurde<sup>14</sup>. Für Stuchs spricht auch die Tatsache, dass dieser Ende des 15. Jahrhunderts vor allem liturgische Bücher für ostdeutsche Diözesen herausgab und zeitweise auch für das Erzbistum Krakau tätig war. Andererseits wurde in der Forschung darauf hingewiesen, dass der schlechte und ungleichmäßige Satz und die Verwendung eines spärlichen und dazu abgenutzten Typenvorrats nicht den sonstigen qualitativ hochwertigen Arbeiten Stuchs entspricht<sup>15</sup>. Matthey hält es daher für möglich, dass ein Anfänger im Buchgewerbe den ›Deutschen Ptolemaeus‹ mit zuvor bei Stuchs erworbenen Typen druckte<sup>16</sup>.

## 2.2 Inhalt

Inhaltlich ist der ›Deutsche Ptolemaeus‹, wie die meisten Kosmographien der Zeit, zweigeteilt. Einem lateinischen Widmungsgedicht, das, so Herkenhoff, den wissenschaftlichen Charakter des Werkes betonen soll<sup>17</sup>, folgen ein theoretisch-mathematischer erster Teil und eine Erdbeschreibung. Gerade in letzterer orientiert sich der Verfasser an der ptolemaeischen ›Geographia‹. Er beschreibt in diesem zweiten Teil alle damals bekannten Kontinente. Besonders charakteristisch für den ›Deutschen Ptolemaeus‹ ist das dezidierte Interesse an Sprachen und Alphabeten, denn der Beschreibung jedes Kontinents folgt eine Auflistung der dort gesprochenen Sprachen und Schriften. Hier ähnelt die Kosmographie zeitgenössischen Reiseberichten, »die gleichfalls ein deutliches Interesse für die Sprachen fremder Völker und Kulturen erkennen lassen«<sup>18</sup>.

## 2.3 Weltkarte

Wie schon erwähnt, stellt die beigelegte Weltkarte eine Besonderheit dar. Sie ist nicht zusammen mit einer der beiden bekannten Ausgaben überliefert, sondern wurde erst

12 Cod. Vindob. lat. 2992; siehe Herkenhoff [Anm. 7], S. 135; Matthey [Anm. 11], S. 79.

13 Fischer [Anm. 2], S. 23.

14 Matthey [Anm. 11], S. 83.

15 Matthey [Anm. 11], S. 80; Herkenhoff [Anm. 7], S. 136

16 Matthey [Anm. 11], S. 80.

17 Herkenhoff [Anm. 7], S. 139.

18 Herkenhoff [Anm. 7], S. 140.

Anfang des 20. Jahrhunderts von Joseph Fischer in der Kantonsbibliothek Vadana in St. Gallen<sup>19</sup> wiederentdeckt, und zwar in einer Ulmer Ptolemaeus-Ausgabe. Sie befindet sich heute in der Public Library in New York. Dass sich weder in der gebundenen Kosmographie in New York noch in München eine Karte erhalten hat, erklärt Fischer damit, dass die Werke vermutlich nicht gebunden, sondern nur geheftet waren bzw. die Karten nur lose beilagen, um eine bequemere Nutzung zu ermöglichen<sup>20</sup>. Die Karte ist die erste Weltkarte, die in Planiglobular-Projektion entstanden ist und somit einen direkten Hinweis auf die Kugelgestalt der Erde gibt. Fischer konnte anhand der Übereinstimmung von Nummern in Karte und Text eindeutig die Zusammengehörigkeit feststellen<sup>21</sup>.

### 3 Probleme der Digitalisierung von Inkunabeln

Die bei der Digitalisierung von Inkunabeln zu lösenden Probleme sind aus einer Reihe von Projekten hinreichend bekannt, siehe z. B. die Überlegungen von Rydberg-Cox<sup>22</sup> oder die Website des ›Incunable Project‹<sup>23</sup>. Prominent sind dabei neben der nicht normalisierten Orthographie die Vielfalt an Typen, Ligaturen, Diakritika, Abkürzungen, Wortgrenzen und -trennungen; Vieles ist dabei auch der Nähe zur spätmittelalterlichen Handschriftenkultur geschuldet. In vielen Fällen, so auch hier, ist trotz intensiver Forschung an automatischer optischer Zeichenerkennung (OCR)<sup>24</sup> eine automatische Transkription in maschinenlesbaren Text noch in weiter Ferne; zumindest wäre in jedem Fall der Aufwand der Nachbearbeitung im Vergleich zu einer manuellen Transkription kritisch ins Verhältnis zu setzen.

Die wesentlichen Probleme bei unserem Exemplar sind neben einem ziemlich dilettantischen Satz und den teilweise sehr abgenutzten Drucktypen die eben genannten. Interessant wäre ein Vergleich mit dem New Yorker Exemplar, denn bei Frühdrucken sind durchaus Unterschiede zwischen einzelnen Exemplaren zu beobachten; dieses konnte bisher nicht in Augenschein genommen werden und digital liegt es nicht vor. Offensichtlich stand dem Drucker ein zu geringer Typenvorrat zu Verfügung, so dass manche Typen vertauscht oder nachbearbeitet bzw. fehlende Typen aus anderen (z.B. iv für w) zusammengesetzt wurden. Die meisten der bei unserem Beispiel gewonnenen Erfahrungen sind also verallgemeinerbar – was auch eine Publikation an dieser Stelle rechtfertigen mag.

19 Siehe <http://www.sg.ch/home/kultur/kantonsbibliothek.html>.

20 Fischer [Anm. 2], S. 12.

21 Fischer [Anm. 2], S. 11.

22 J. A. Rydberg-Cox: Digitizing Latin Incunabula: Challenges, Methods, and Possibilities, in: Digital Humanities Quarterly 3/1 (2009), S. 1–6.

23 <http://daedalus.umkc.edu/incunables/>.

24 Vgl. C. Kämmerer: Vom Image zum Volltext – Möglichkeiten und Grenzen des Einsatzes von OCR beim alten Buch, in: Bibliotheksdienst 43/6 (2009), S. 626–659.

## 4 Ein studentisches Projekt mit TEI/XML

Am Anfang stand eine manuelle diplomatische Transkription des Münchner Exemplars als Rohtext unter Beibehaltung des Layouts und mit Ersatzdarstellungen für alle Glyphen – z. B. Diakritika wie Nasalstriche –, die im ISO-Latin-Zeichensatz nicht enthalten sind<sup>25</sup>.

Ausgehend von dieser Transkription stellte dann Barbara Ries in ihrer Magisterarbeit systematische Überlegungen zu einer digitalen Edition an und setzte diese exemplarisch anhand des Kapitels über Europa in einer XHTML-Präsentation um. In verschiedenen Ansichten konnten in jeweils drei Spalten ein verlinkter Quellenkommentar sowie die Transkription – wahlweise diplomatisch oder bzgl. offensichtlicher Druckfehler normalisiert – und Seitenbilder (Digitalisate der BSB oder des Faksimiles von Fischer) eingesehen werden. Diese Idee auf den gesamten Textumfang anzuwenden, lag auf der Hand: Es sollte zunächst eine Web-Präsentation mit Transkription und Seitenbildern erzeugt werden, die später dann um eine neuhochdeutsche Übersetzung und um einen kritischen Kommentar erweitert werden kann.

Da grundsätzlich nur etablierte Standards und Konventionen genutzt werden sollten, gab es keine Alternative zur Codierung der Transkription in TEI/XML, womit per definitionem auch eine Entscheidung für Unicode getroffen wird. Die Umsetzung der diplomatischen Transkription nach TEI erwies sich als problemlos; Aufwand entstand durch die Auszeichnung der Druckfehler, Abkürzungen und Worttrennungen. Hinzu kamen später Auszeichnungen für "Named Entities" (Namen von Personen, Völkern, geographischen Orten) und Fachtermini. Auch wenn man ein grundsätzlich anderes Codierungssystem wählen würde, wäre der Aufwand wohl nicht geringer; letztlich hängt es von den verwendeten Werkzeugen ab, welche Belastung dem Benutzer bleibt. Der durch diese Auszeichnungen erzielte Mehrwert für Präsentation und mögliche Auswertungen ist jedenfalls enorm. Da der Text auch als Experimentierfeld für die Erprobung verschiedener Darstellungstechniken, Werkzeuge und Auswertungsverfahren dienen sollte, ist eine TEI-Codierung ein nahezu idealer Ausgangspunkt – zumal es sich ja letztlich um ein reines Textformat handelt, bei dem Nachhaltigkeit kein echtes Problem darstellt.

Auftakt des Projekts war ein Kurs bei der Sommeruniversität der Studienstiftung des Deutschen Volkes in Greifswald 2008, den der Autor zusammen mit Josef Schneeberger (Nürnberg/TH Deggendorf) mit großen Erfolg durchführte. Seitdem diente der Text immer wieder als Kern von Übungsprojekten zur mehrfach auch gemeinsam angebotenen Vorlesung ›Digitale Dokumente, Editionen und Bibliotheken‹. Sämtliche Materialien und

25 Vgl. Florian Kragl: Pseudo-Markup: Eine Eselsbrücke zwischen manueller und maschineller Textverarbeitung, in diesem Band.

Ergebnisse sind auf der Webseite ›Kulturerbe digital<sup>26</sup> und der Seite zur Vorlesung<sup>27</sup> zu finden.

## 5 Lösungsansatz

Auch wenn aufgrund der verfügbaren Ressourcen zunächst nur die digitale Publikation einer Transkription möglich war, sollten gleichwohl auch weitergehende Überlegungen zu digitalen Editionen leitend sein<sup>28</sup>. Durch die Wahl von TEI und Unicode waren wichtige editorische Entscheidungen vorgeprägt; darauf aufbauend wurden Transkriptionsrichtlinien erarbeitet.

Da es von unserer Inkunabel keine weiteren Ausgaben oder gar Auflagen gibt, sollte auf jeden Fall das Layout, d. h. die Zeilen- und Seitenstruktur in der Transkription erhalten bleiben. Da jedes TEI-Dokument aufgrund seiner XML-Basis aus nur einem Element mit hierarchischer Binnenstruktur besteht, können weder einander überkreuzende Hierarchien noch andere Arten von Graphen in TEI-Dokumenten direkt repräsentiert werden. Schon Seitenumbrüche durchbrechen aber die hierarchische Struktur des Dokuments, die durch die Kapitel- und Absatzorganisation gegeben ist. Daher wurden Letztere durch sog. Milestones dargestellt<sup>29</sup>. Für die drei hauptsächlichen Variantentypen wurde jeweils eine Darstellung mit dem choice-Element<sup>30</sup> vorgenommen:

- echte Druckfehler (vs. Schreibvarianten): Original und Korrektur,
- Abkürzungen: Abkürzungsglyph bzw. -form und Expansion,
- Worttrennungen (mit und ohne Trennzeichen): Getrennte und ganze Wortform (auf der Anfangszeile).

Später kamen noch Auszeichnungen von “Named Entities” hinzu, d. h. der Namen von Personen, Völkern und geographischen Orten – nach TEI P5<sup>31</sup> – sowie von astronomischen und geographischen Fachtermini mit dem term-Element<sup>32</sup>.

Einige Besonderheiten, die unsere Inkunabel aufweist, sind

- eine Tabelle (der Winde)
- Marginalien (Zahlen)
- Text-Bild-Referenzen in der Form von Zahlen in Abschnitts-Überschriften, die sich in der zugehörigen Weltkarte wiederfinden.

26 <http://www.dh.cs.fau.de/IMMD8/Services/textfarm/>.

27 <http://www.dh.cs.fau.de/IMMD8/Lectures/DIGIDOK/>.

28 Siehe Patrick Sahle: Digitales Archiv und Digitale Edition. Anmerkungen zur Begriffsklärung, in Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung, Zürich 2007, S. 64–84.

29 Lou Burnard, Syd Bauman: TEI P5: Guidelines for Electronic Text Encoding and Interchange, Handbook, TEI Consortium, Oxford u. a. 2008, Kap. 3.10.

30 Burnard [Anm. 29], Kap. 3.4.

31 Burnard [Anm. 29], Kap.13; siehe auch <http://www.teic.org/release/doc/tei-p5-doc/en/html/ND.html>.

32 Burnard [Anm. 29], Kap.3.

Einige handschriftliche Marginalien wurden bisher nicht beachtet, sind aber sicher für die Erforschung der Benutzung des Exemplars wichtig. Allerdings wurde es beschnitten, so dass einige Marginalien nur fragmentarisch erhalten sind. Interessant wäre auch ein Handschriftenvergleich mit den beiden Abschriften – mit der schon erwähnten von Johann Schöner 1509 (heute in Wien) sowie einer aus dem Besitz von Siegmund Scheufler (Anf. 16. Jh., BSB München, Clm 388).

Für die TEI-Codierung des ›Deutschen Ptolemaeus‹ mit allen genannten Merkmalen wurden verschiedene Editoren eingesetzt, die alle zumindest über Module für die Syntax von XML verfügen. Von großem Vorteil ist, wenn beim Anlegen eines Elements automatisch die schließende Klammer erzeugt wird und wenn Schablonen für komplexe Konstrukte wie das choice-Element definiert werden können oder gar automatisch angegeben werden. Für die vereinfachte Texteingabe bietet sich auch an, wie Kragl<sup>33</sup> beschreibt, kurze Ersatzdarstellungen zu verwenden, die dann mittels eines Texteditors, z. B. sed (Unix) im Batch-Modus, in die entsprechend instantiierten Schablonen expandiert werden.

Zunächst wurde – und wird – Emacs<sup>34</sup> benutzt, dessen Benutzerfreundlichkeit allerdings nicht alle so positiv einschätzen wie der Autor. Als weitere kostenfreie Alternative wurde das XML-Modul für die Java-orientierte Programmierungsumgebung Eclipse<sup>35</sup> eingesetzt. Sofern man sich den Rahmenvorgaben beugt, nämlich ein Editionsprojekt im Stil eines Software-Entwicklungsprojekts zu organisieren, ist das XML-Plugin zweckdienlich. Zusätzlich hat man den Vorteil, dass mit Einrichtung einer geeigneten Speicherorganisation die Daten in einem Mehrbenutzermodus in verteilten Rollen bearbeitet werden können. Dies erspart den Einsatz einer sog. Groupware, wie wir sie am Anfang des Projekts eingesetzt hatten<sup>36</sup>. Seine Benutzung ist inzwischen einfacher geworden im Zusammenhang mit einem anderen XML-Editorprojekt, ArborealMWN<sup>37</sup>. Arboreal war ein Projekt von Harvard mit dem MPI für Wissenschaftsgeschichte, das vor allem die Integration linguistischer Dienste – Lexika, Lemmatisierung und Morphologie, Terminologieverwaltung – bei der Bearbeitung digitaler Editionen unterstützt. Die aktuelle Implementierung auf der Basis von Eclipse, die die Handhabung wesentlich vereinfacht und die weitgehend fertiggestellt ist, hat das o. g. XML-Modul integriert. Als dritter Editor wurde Oxygen<sup>38</sup> eingesetzt, der kostenpflichtig, aber äußerst komfortabel ist, jedoch eine solide Kenntnis von XML voraussetzt – und deshalb für Bearbeiter mit einem philologischen Hintergrund eine gewisse Hürde darstellen dürfte. Angeblich können spezielle

33 Kragl [Anm. 25].

34 <http://www.gnu.org/software/emacs/>.

35 <https://www.eclipse.org/>.

36 Konkret <http://bscw.de> von Fraunhofer, freie ›educational license‹.

37 <http://sourceforge.net/projects/arboreal/>.

38 <http://www.oxygenxml.com/>.

Konfigurationen für bestimmte Aufgaben generiert werden; hierzu liegen bisher keine Erfahrungen vor.

Ist die Codierung des Textes in TEI einschließlich der Metadaten im TEI-Header abgeschlossen, steht die Frage der Präsentation an. Hierzu gibt es ein breites Spektrum an Möglichkeiten, wobei die einfachste darin besteht, die TEI-Datei nach XHTML zu konvertieren, so dass sie mit jedem Browser gelesen werden kann. Allerdings ist bei der Verwendung von Elementen, die, wie z. B. choice, Redundanzen einführen, ein Vorverarbeitungsschritt notwendig, um zu verhindern, dass z. B. eine Abkürzung als solche und daneben auch noch in expandierter Form wiedergegeben wird. Beispielhafte XSLT-Skripte sind auf der Webseite zur o.g. Vorlesung zu finden. Ihre Anwendung kann entweder im Kommandozeilenmodus durch expliziten Aufruf eines Prozessors wie Saxon erfolgen oder komfortabler eingebunden in einen XML-Editor – optimal mit Oxygen. Eine einfache Möglichkeit zur Konvertierung (und viele andere!) ist über die Website <http://www.tei-c.org/oxgarage/> verfügbar. Möchte man die XHTML-Transkription zusammen mit den entsprechenden Seitenbildern anzeigen, kann dies – wie bei Ries – in zwei Frames erfolgen, wobei die Verknüpfung in der TEI-Datei z. B. durch das facsimile-Element<sup>39</sup> angegeben werden kann.

Anstelle einer Transformation nach XHTML bietet ›TEI Boilerplate<sup>40</sup> die Alternative der direkten Anwendung eines Stylesheets auf die TEI-Datei. Im einfachsten Fall muss nur eine Operationsanweisung in die TEI-Datei eingefügt werden, die auf ein geeignetes Stylesheet verweist, wobei eine Reihe von Stylesheets mitgeliefert wird. Voraussetzung ist allerdings, dass ein Apache Webserver installiert ist<sup>41</sup> und die teibp-Dateien dort auch vorgehalten werden. Auch hier ist bei Redundanzen eine vorherige Transformation erforderlich; in einfachen Fällen kann die Auswahl aber auch über ein im Stylesheet implementiertes Toolbox-Element getroffen werden.

## 6 Aufbereitung für eine Web-Präsentation

Zur Web-Präsentation gibt es, wie oben erwähnt, zwei Varianten. Beide erlauben, durch Schaltflächen die Art der Normalisierung einzustellen. In der ersten, experimentellen Variante wurde auch Javascript-Code implementiert, durch den ermöglicht wird, bei Positionierung des Mauszeigers im Bild die entsprechende Zeile in der Transkription durch Einrahmung hervorzuheben. Sie benutzt allerdings lokal bearbeitete und gespeicherte Seitenbilder des BSB-Digitalisats. Die BSB lässt aber keine Gesamtpräsentation ihres Digitali-

39 Burnard [Anm. 29], Kap. 11.1.

40 <http://dc1.slis.indiana.edu/teibp/index.html>.

41 Am einfachsten mit dem PHP-Entwicklungspaket XAMPP, <https://www.apachefriends.org/de/index.html>.

satz zu, sondern nur die Verlinkung von einzelnen Bildern. Die ECHO-Version benutzt aus diesem Grund nur die Bilder des Fischer-Faksimiles.

In ECHO sind verschiedene Anzeigevarianten vorgesehen, u. a. Mini-Seitenbilder (Thumbnails) als Gesamtübersicht oder zusammen mit der Transkription – in verschiedenen einstellbaren Normalisierungen – oder XML-Code oder Vollbild. Der XML-Code kann auch im Ganzen heruntergeladen werden. Intern steht eine Reihe spezieller Werkzeuge zur Verfügung, u. a. zur Einfügung von Satzgrenzen-Auszeichnungen. Dies wurde von den Administratoren auch angewandt, so dass die angebotene Volltextsuche jeweils die Treffer im Kontext des jeweiligen Satzes anzeigt. Nebenbei bemerkt waren auch hier wegen der nicht normalisierten Zeichensetzung manuelle Eingriffe notwendig. Wünschenswert wäre noch eine Anzeige von Paralleltextrn, etwa des Textes in Ausrichtung (alignment) mit einer Übersetzung.

Weiterhin gestattet ECHO auch, jedes Wort mit einem Lexikoneintrag zu verknüpfen – sofern entsprechende Lexika zur Verfügung stehen. Dies trifft zum einen auf das lateinische Widmungsgedicht zu, wo u. a. eine Verlinkung zum (gemeinfreien, da 19. Jh.!) Lexikon von Lewis und Short und darüber dann zur Perseus Digital Library<sup>42</sup> besteht. Für unsere Varietät des Frühneuhochdeutschen gibt es kein geeignetes Lexikon, so dass nur auch im modernen Deutsch gültige Wortformen verlinkt werden. Das führt wegen der durch die nicht normalisierte Orthographie bedingten Ambiguitäten bisweilen zu bizarren Ergebnissen, etwa im Beispiel »schonen figuren«, wo zwar »schönen« gemeint ist, aber auf der (vom Benutzer konfigurierbaren) Lexikoseite dann Zugriffe für »schonen« in digitalen Lexika des Neuhochdeutschen (u. a. DWDS mit Korpusbelegen, Wortschatz Leipzig inklusive Dornseiff-Thesaurus, auch CELEX-Morphologie) angezeigt werden.

Für die interne Webpräsentation wurde auch die native XML-Datenbank eXist-db<sup>43</sup> zur Speicherung der TEI-Datei erprobt. Die Handhabung hat sich als recht einfach erwiesen, bringt aber in unserem Fall einer einzigen TEI-Datei keine Vorteile; solche sind aber bei größeren Korpora zu erwarten.

## 7 Verarbeitung

Für die Untersuchung der sprachlichen Form des Textes bieten sich aufgrund der sehr eigenen Orthographie im Prinzip nur Werkzeuge an, die Wortformen als bloße Zeichenketten ohne Rückgriff auf linguistische Informationen verarbeiten. Dies ist mitnichten ein Sonderfall, denn dasselbe gilt für viele seltene Sprachen; es gibt schlechterdings kaum Lexika und Sprachmodelle für Varietäten mit nichtnormierter Orthographie. So sind auch für

42 <http://www.perseus.tufts.edu/>.

43 <http://exist-db.org/>.

moderne Sprachen erzeugte Tagger, z. B. TreeTagger<sup>44</sup>, bestenfalls von heuristischem Wert. Selbst für alte Sprachen wie Griechisch und Latein gibt es bisher kaum linguistische Software, die über die morphosyntaktische Verarbeitung hinaus auf die Ebene der syntaktisch-semantischen Strukturen reicht.

Wir müssen uns also auf Wort(formen)listen – auch rückläufige für die Suffixanalyse – mit Häufigkeiten, Bigramme, Trigramme, Konkordanzen und eine Vollformenindexierung beschränken. Am einfachsten geht dies mit Rohdaten ohne Auszeichnungen; hierfür stehen in der Unix-Welt diverse über die Kommandozeile aufrufbare Programme zur Verfügung. Unter den einschlägigen integrierten Programmpaketen und Webdiensten gibt es inzwischen etliche, die auch mit XML-Tags – abgesehen von Redundanzen – umgehen können, u. a. haben wir mit dem Korpus-Analysewerkzeug antconc<sup>45</sup> sehr gute Erfahrungen gemacht. Unter den webbasierten Korpus-Tools ist Voyant<sup>46</sup> besonders zu empfehlen.

Auf der Basis erzeugter Wortformenlisten wurden Untersuchungen durchgeführt, inwieweit unscharfe Suche in Texten mit nichtnormierter Orthographie sinnvoll eingesetzt werden kann. Zugrunde gelegt wurde die sog. Edit-Distanz (Levenshtein-Algorithmus<sup>47</sup>) zwischen zwei Zeichenketten. Sie wird auch in dem Unix-Programm ›agrep‹ verwendet, das zwar prinzipiell geeignet erscheint, aber noch genauer evaluiert werden muss.

Für weitergehende linguistische Analysen bietet sich als Notbehelf an, mit textnahen, d. h. eher wörtlichen Übersetzungen als Paralleltext zu arbeiten. Aus einem anderen Projekt liegen gute Erfahrungen mit dem Stanford-Parser<sup>48</sup> vor. Während dessen deutsches Sprachmodell nur Konstituentenstrukturen erzeugen kann, liefert das englische auch Dependenzstrukturen, die für die semantische Ebene besser anschlussfähig sind. Zur Erzeugung von Prädikat-Argumentstrukturen hat sich (für das Englische) u. a. der Semantic Role Labeler<sup>49</sup> der Universität Lund bewährt, der auch den Zugriff auf ein framebasiertes Lexikon ermöglicht.

## 8 Erstellung von Druckfassungen

Zusätzlich zur Online-Präsentation wird oft auch eine Ausgabe in einem Druckformat, normalerweise PDF, oder in einem anderen eBook-Format gewünscht. In der XML-Welt wurde hierzu die Transformation mittels XML-FO (Formatting Objects) entwickelt. Da zwischen der inhaltlich-deklarativen Auszeichnung in TEI/XML und einer Auszeichnung

44 <http://www.cis.uni-muenchen.de/schmid/tools/TreeTagger/>.

45 <http://www.antlab.sci.waseda.ac.jp/software.html>.

46 <http://voyant-tools.org/>.

47 Siehe <http://www.levenshtein.de/>.

48 <http://nlp.stanford.edu/software/lex-parser.shtml>.

49 <http://nlp.cs.lth.se/>.

mit LaTeX<sup>50</sup> eine Analogie besteht, liegt die Vermutung nahe, dass zur professionellen Layout-Produktion mittels XML-FO ein ähnliches Verfahren einzusetzen wäre. Tatsächlich gab es auch Versuche, die Satzmaschine von TeX für diesen Zweck einzusetzen, die allerdings bisher aus nicht klar nachvollziehbaren Gründen nicht besonders erfolgreich verliefen. Für diesen Zweck eigens entwickelte Open-Source-Tools hatten bisher nur einen begrenzten Leistungsumfang. Seit kurzem steht mit XML-Print<sup>51</sup>, das im Rahmen des Verbundprojekts TextGrid entwickelt wird, ein neues, ebenfalls mit Eclipse entwickeltes Rahmensystem mit einer flexiblen graphischen Benutzungsschnittstelle bereit. Erste Versuche mit unserem Text verliefen erfolgreich, jedoch ist noch einiger Feinschliff bei den Ausgabeformaten notwendig. In der Handhabung besteht gegenüber dem reinen XML-FO-Konzept eine signifikante Vereinfachung.

## 9 Ausblick

Mit der TEI-Codierung ist eine Basis für eine Reihe künftiger Erweiterungen gelegt. In einer Bachelorarbeit<sup>52</sup> wurde u. a. untersucht, inwieweit sich die XML Linking Language (XLink) zur Definition und Nutzung von Links in TEI-Dokumenten einsetzen lässt. Zwar verfügt die TEI auch über eine eigene Link-Komponente, aber die Nutzung der damit formulierten Links wird bisher nicht von Werkzeugen unterstützt. Auch XLink wird eher stiefmütterlich behandelt; doch es konnte praktisch nachgewiesen werden, wie man durch eine Auswertung von XLink-Konstrukten zu einer erweiterten HTML-Edition gelangt.

Ein großes Potential besteht in der semantischen Auswertung der Auszeichnungen von Personennamen, Toponymen, Zeitangaben, Ereignissen etc. durch deren Anschluss an eine semantische Modellierung durch formale Ontologien wie CIDOC CRM<sup>53</sup>. Das Conceptual Reference Model (CRM) wurde als Referenzontologie für die Dokumentation des Kulturerbes definiert und ist ein internationaler Standard (ISO 21127). Eines seiner wesentlichen Ziele ist die Herstellung von Interoperabilität und umfassender Vernetzung unterschiedlicher Ressourcen. Die Überlegung »that detailed information about persons (physical and legal), dates, events, places, objects etc. and their interpretation could be marked up outside the text, and that this could be connected to on-going ontology work« führte zur Gründung einer Ontologie-Arbeitsgruppe innerhalb der TEI<sup>54</sup>. Durch die Einbindung von Normdateien, z. B. für geographische Bezeichnungen, die wie

50 <http://www.latex-project.org/>.

51 <http://kompetenzzentrum.uni-trier.de/de/projekte/projekte/xml-print/>.

52 Schulz [Anm. 5].

53 <http://www.cidoc-crm.org/>, <http://erlangen-crm.org>.

54 Siehe <http://www.tei-c.org/Activities/SIG/Ontologies/>.

Pleiades eindeutige URIs zuordnen<sup>55</sup>, wird eine Vielfalt weiterer Ressourcen zugänglich. Eine interessante Möglichkeit besteht in der Geovisualisierung mit Web-GIS-Systemen wie GeoTemCo<sup>56</sup>.

Zugleich können die aus den Auszeichnungen gewonnenen Propositionen in der Subjekt-Prädikat-Objekt-Form, sog. RDF-Tripel<sup>57</sup>, in Graphen organisiert als »Linked Open Data« über eine standardisierte Anfrageschnittstelle (SPARQL<sup>58</sup>) allgemein zugänglich gemacht werden. Beispielhaft sei hier die von Thomas Kirchner et al. besorgte digitale Edition von Joachim von Sandrarts ›Teutscher Academie der Bau-, Bild- und Mahlerey-Künste«, Nürnberg 1675/1679/1680, genannt<sup>59</sup>. In diesem Editionsprojekt wird besonders darauf hingewiesen, »dass es sowohl machbar als auch wünschenswert erscheint, zumindest einen Teil unserer Daten als Linked Open Data zu veröffentlichen: Personen, Orte, Kunstwerke, Bibliographie-Objekte und die Bezüge zwischen diesen.« Durch diese neue Dimension der Interoperabilität gewinnt nicht nur die eigene Edition einen erheblichen Mehrwert, sondern dieser kommt in gleicher Weise auch anderen zu.

## Danksagung

Besonderer Dank gebührt Josef Schneeberger und Martin Scholz für die umfangreiche und kontinuierliche technische Unterstützung des Projekts. Ohne die vielen Beiträge von Studierenden zu Transkription, Korrekturlesen und Implementation – neben den bereits Genannten die Teilnehmer unseres Kurses bei der Sommeruniversität der Studienstiftung des Deutschen Volkes in Greifswald 2008 sowie in Erlangen Studierende aus Buchwissenschaft, Germanistik und Informatik – wäre das Projekt immer noch in den Kinderschuhen. Für die Integration in ECHO am MPIWG Berlin sei Josef Willenborg und Klaus Thoden herzlich gedankt.

55 Siehe Pelagios, <http://pelagios-project.blogspot.de/>, dort speziell auch IN USE.

56 <http://www.informatik.uni-leipzig.de:8080/geotemco/>.

57 Siehe <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>.

58 <http://www.w3.org/TR/rdf-sparql-query/>.

59 <http://ta.sandrart.net/de/>, insbesondere <http://ta.sandrart.net/de/info/services/lod-rdf/>.