

Die Text Encoding Initiative (TEI) als Werkzeug des Urkundeneditors – Erfahrungen und Desiderate

Von Georg Vogeler¹

Einleitung

Das Verhältnis zwischen dem Kodierungsstandard der Text Encoding Initiative (TEI) und diplomatischen Editionen möchte ich im Folgenden unter zwei Perspektiven besprechen:

1. einer pragmatischen: Kann man mit der TEI eine Urkundenedition erstellen? Was hat es für Konsequenzen, wenn man mit der TEI eine Urkundenedition erstellt?

Und 2. einer grundsätzlichen: Was wird eigentlich bei einer Urkundenedition ediert? Wie verhält sich das zur Vorstellung der TEI, was mit dem Standard kodiert werden kann?

Die Rolle der Text Encoding Initiative für Urkundeneditionen

Die TEI

Die Text Encoding Initiative (TEI)² ist eine Gruppe von GeisteswissenschaftlerInnen, die seit 1987 an einer Computerkodierung von Text arbeitet, welche den Austausch zwischen verschiedenen Projekten erleichtern soll. Ihre Vorschläge basieren auf der eXtensible Markup Language (XML), einer vom World Wide Web empfohlenen Form, Kodierungen direkt in Texte einzufügen, so dass die Dokumente sowohl von Maschinen wie von Menschen gelesen werden können. Die TEI hat eine Dokumentgrammatik entwickelt, welche Definitionen für Namen von XML-Kodierungen vorschlägt. XML-Software kann sich dabei eines XML-Schemas, das heißt einer technischen Formalisierung des Regelwerks der TEI bedienen³. Die TEI ist eine Erfolgsgeschichte, denn es hat sich seit ihrer Gründung eine weltweite Community gebildet, welche Texte nach den Vorschlägen der TEI kodieren. Darunter sind auch viele HandschriftenforscherInnen und EditorInnen, welche die TEI nicht nur benutzen, sondern auch aktiv Vorschläge zu ihrer Verbesserung und Erweiterung einbringen.

Urkundeneditionen mit TEI

Es sind auch einige Urkundeneditionen entstanden, die das Regelwerk der TEI verwenden. Die elektronischen Editionen *Éditions en ligne de l'École des Chartes (ELEC)* der École des Chartes setzen besonders konsequent auf die Benutzung der TEI⁴. Urkundentexte ediert zum

¹ Zentrum für Informationsmodellierung – Austrian Centre for Digital Humanities, Universität Graz, und Institut für Dokumentologie und Editorik e.V., <georg.vogeler@uni-graz.at>. Dieses Werk ist lizenziert unter einer Creative Commons <<http://creativecommons.org/licenses/by-sa/4.0/>>, Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International Lizenz.

² <http://www.tei-c.org>.

³ Das Regelwerk der TEI liegt in den Formaten Document Type Definition (DTD), XML Schema (XSD) und – bevorzugt – Relax NG (RNG) vor.

⁴ <http://elec.enc.sorbonne.fr>.

Beispiel die Gruppe um Olivier Guyotjeannin, die sich dem Kloster Saint-Denis widmet. Explizit Urkundeneditionen bieten die *Chartes médiévales de l'abbaye de Saint-Denis*⁵ und das *Formulaire d'Odart Morchesne* aus der Handschrift BnF fr. 5024⁶. Die TEI wird auch bei Retrodigitalisierungen an der École des Chartes angewendet, wie zum Beispiel den Actes royaux du Poitou (1302–1464)⁷ oder weiteren Editionen von Chartularen der Île-de-France⁸. Auch am Department for Digital Humanities des King's College London legt man Wert auf die Verwendung der TEI. Ihre Regeln kamen also auch bei der Onlinepublikation der Anglo-Saxon Charters unter der Leitung von Janet L. Nelson zur Anwendung⁹, wobei das Forschungsinteresse nicht die Edition der Urkunden, sondern ihre Verknüpfung mit den Daten der Prosopography of Anglo Saxon England war. Schon auf der Version P4 der TEI beruht die elektronische Fassung des Corpus Altdeutscher Originalurkunden, welche Kurt Gärtner, Andreas Gniffke und Andrea Rapp organisiert haben¹⁰.

Es soll im Folgenden nicht um die individuellen Funktionalitäten gehen, die diese Editionen online bieten. Die TEI beschäftigt sich nämlich nicht vorrangig mit Webfunktionalitäten, sondern mit Datenmodellen und Kodierungsformen für programmgestützte Arbeit und den Datenaustausch. Es geht in diesem Beitrag also darum, welche Arten von Informationen Diplomatiker aus digitalen Urkundeneditionen extrahieren können, die mit der TEI realisiert werden.

Eine Urkunde könnte nach den Regeln der TEI wie folgt kodiert werden¹¹:

```
<TEI>
  <text>
    <front>
      <argument><p>Charte d'Hersende par laquelle elle rend en toute
      liberté à Dieu et à S. Leu, Fromond, son homme [serf] de garde de
      Cauffry moyennant la prestation annuelle d'une chandelle d'un
      denier.</p></argument>
      <docDate><date when="1104" scope="circa">Vers 1104</date></docDate>
    </front>
    <body>
      <div type="transcription" xml:lang="lat">
```

⁵ <http://saint-denis.enc.sorbonne.fr/>.

⁶ Bearb. von Olivier GUYOTJEANNIN / Serge LUSIGNAN, <http://elec.enc.sorbonne.fr/morchesne/>.

⁷ <http://corpus.enc.sorbonne.fr/actesroyauxdupoitou/>.

⁸ <http://elec.enc.sorbonne.fr/cartulaires/>.

⁹ <http://www.aschart.kcl.ac.uk/index.html>. Vgl. zum Projekt Arianna CIULA / Paul SPENCE, Threads of integration. The Anglo-Saxon Charters pilot project at the Centre for Computing in the Humanities, in: Digitale Diplomatie, Neue Technologien in der historischen Arbeit mit Urkunden, hg. von Georg VOGELER (AfD Beiheft 12), Köln u. a. 2009, S. 40–55.

¹⁰ <http://urts55.uni-trier.de/cgi-bin/iCorpus/CorpusIndex.tcl>.

¹¹ Beispiel aus dem Cartulaire du prieuré de Saint-Leu d'Esserent (1080–1538), bearb. von Eugène MÜLLER, Pontoise 1900–1901, n. V, online unter: <http://elec.enc.sorbonne.fr/cartulaires/S-Leu-d-Esserent/0007>.

```

<p>Notum sit omnibus tam futuris quam presentibus quod Ego
<persName>Hersendis</persName> hominem meum <persName>Frotmundum
de Cauferei</persName> quem possidebam sicut meum hominem
videlicet de custodia, annuentibus filiis meis, pro animabus
antecessorum meorum et pro salute mea reddo in omni libertate Deo
et <placeName><abbr>Sco</abbr> Lupo</placeName>, per unam candelam
denarii uno quoque anno reddendam in predicto loco et in predicta
Ecclesia. ...</p>
</div>
</body>
</text>
</TEI>

```

Das Regest ist als *argument* kodiert, das Ausstellungsdatum unter der Bezeichnung *docDate*, der eigentliche Text ist ein Abschnitt *div*(ision), die vom Benutzer einem Typ „transcription“ zugeordnet wird. Die Sprache des Urkundentextes ist im Attribut *xml:lang* als Lateinisch angegeben. Im Text sind Personen- und Ortsnamen markiert. Das näherungsweise angegebene Datum ist im Attribut *when* numerisch kodiert, im Attribut *scope* wird darauf hingewiesen, dass es ein erschlossener Wert ist¹².

Die TEI bietet aber mehr als die elementaren Strukturelemente, denn sie gibt auch Anregungen, wie ForscherInnen Befunde in der Handschrift beschreiben können, das heißt sie benennt explizite Kodierungsmöglichkeiten für den kritischen Apparat, Orts-, Personen- und Sachregister, die Beschreibung von Originalen und handschriftlicher Überlieferung. Diese Kodierung erweitert die Repräsentationsmöglichkeiten von editorischen Befunden gegenüber den in der Druckwelt etablierten Formen. Sie ermöglicht nämlich erstens Abkürzungen, Korrekturen und Normalisierungen direkt im Text in beiden Varianten zu speichern. Dazu werden die Alternativen als eine Auswahl (*choice*) markiert, die Abkürzungszeichen (*am*) und Auflösung (*ex*), originale Schreibung (*orig*) und Korrektur (*corr*), Vorlagenfassung (*sic*) und Normalisierung (*reg*) enthalten kann¹³. Zweitens kann die Transkription in direkter Beziehung zu einem Bild des Textes gespeichert werden, entweder über einen Verweis mit Attribut *@facs* oder indem man den Text in die Kodierung der Bilder (*surface* und *graphic*) und Bildbereiche (*zone*) einbettet¹⁴.

¹² Die Definitionen zu den einzelnen Elementen findet man unter <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-argument.html>, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-div.html>, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-docDate.html> und <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-date.html>.

¹³ Vgl. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COED>, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COEDREG>, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COEDCOR> und <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PHAB>.

¹⁴ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.global.facs.html> und <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PHFAX> und für die eingebettete Transkription <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PH-transcr>.

Wenn man von diesen Möglichkeiten nur einen Teil in den Editionen realisiert sieht, liegt das teilweise daran, dass die Editionen entstanden sind, bevor diese Modelle in die Richtlinien der TEI aufgenommen wurden. Es liegt aber auch an Entscheidungen der EditorInnen, praktische Lösungen zu verwenden.

TEI in der editorischen Praxis

Grundlage für die Benutzung der TEI sind die vom TEI-Consortium gepflegten Richtlinien („Guidelines“)¹⁵, die in der Version P5-2.7 vom 16.9.2014 546 Elemente und 108 Attribute definieren. Sie decken zentrale Textphänomene¹⁶ wie Abschnitte, Überschriften oder Hervorhebungen ebenso ab wie genrespezifische Kodierungen für Dramen¹⁷, Gedichte¹⁸, Lexika¹⁹, Handschriftenbeschreibungen²⁰, die Transkription von handschriftlichen Quellen²¹ und den kritischen Apparat²².

Die Richtlinien der TEI erheben nicht den Anspruch, alle möglichen textuellen Phänomene zu beschreiben. Es können projektspezifische Anpassungen erzeugt werden, und es kann mit Hilfe der TEI dokumentiert werden, wie sich diese Anpassungen zu den vorhandenen Elementen verhalten²³. Als Werkzeug zur Erstellung einer angepassten Version der TEI hat das TEI-Consortium unter dem Namen *Roma* einen Webdienst entwickelt²⁴. Damit kann man auch ohne technische Spezialkenntnisse ein Schema erzeugen, das eine Auswahl aus dem Regelwerk der TEI trifft, Elemente hinzufügt und ihren Gebrauch modifiziert.

Für die Erfassung von Daten im Format der TEI stehen unterschiedliche Werkzeuge zur Verfügung. So können die Daten mit allgemeinen Textverarbeitungssystemen erfasst und mit Hilfe von Konvertierungssystemen in TEI umgewandelt werden²⁵. Dabei werden aber nur einfache Strukturen der Texte übernommen. Dedizierte Software zur Erstellung kritischer Editionen wie der *Classical Text Editor* von Stefan Hagen²⁶ erlaubt einen Export als TEI-XML, der mehr Kodierungsmöglichkeiten berücksichtigt. Den ganzen Beschreibungshorizont der TEI bilden aber nur XML-Editoren ab, unter denen der von der Firma SyncRo hergestellte Editor *oXygen* die TEI besonders gut unterstützt. Prinzipiell kann aber jeder XML-Editor zur

¹⁵ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>.

¹⁶ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html> und <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html>.

¹⁷ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DR.html>.

¹⁸ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/VE.html>.

¹⁹ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>.

²⁰ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>.

²¹ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>.

²² <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>.

²³ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TD.html>.

²⁴ <http://www.tei-c.org/Roma/>.

²⁵ Vgl. zum Beispiel das von den IT-Services der Universität Oxford entwickelte Werkzeug *OxGarage*: <http://www.oucs.ox.ac.uk/oxgarage/>.

²⁶ <http://cte.oeaw.ac.at/>.

Kodierung verwendet werden. Vereinzelt sind auch dedizierte TEI-Editoren in Entwicklung²⁷, die aber bislang auch nur einen Ausschnitt der Elemente der TEI verarbeiten.

Die Konzepte der TEI folgen überwiegend einem Verständnis von Text, das von DiplomatikerInnen geteilt wird. UrkundenforscherInnen können diese Kodierungen deshalb nach einer überschaubaren Einarbeitungszeit erlernen und selbst anwenden. Dagegen benötigt die Erzeugung von Druckvorlagen und Onlinepräsentation aus den TEI-Daten technische Kenntnisse aus dem Bereich der Softwareentwicklung, die üblicherweise nicht Teil der Ausbildung von DiplomatikerInnen sind. Diese Kompetenz ist inzwischen aber besonders an Zentren für digitale Geisteswissenschaften verfügbar, die an vielen Universitäten und Forschungseinrichtungen aus der Notwendigkeit zur Unterstützung von Geisteswissenschaftlern in IT-Fragen entstanden sind und ihre Dienste sowohl universitätsintern wie -extern bereitstellen. Diese Zentren sind häufig nicht nur Serviceeinrichtungen für Softwareentwicklung, sondern bieten auch Speicherungs- und Publikationsdienste an, stellen also die entsprechenden Webserver zur Verfügung und kümmern sich um die langfristige Sicherung der Daten. Im deutschsprachigen Raum sind das zum Beispiel TELOTA an der Berlin-Brandenburgischen Akademie der Wissenschaften in Berlin, das Cologne Center for e-Humanities (CCeH) als Dienstleister der Universität Köln und der Nordrhein-Westfälischen Akademie der Wissenschaften oder das Zentrum für Informationsmodellierung – Austrian Centre for Digital Humanities (ZIM – ACDH) der Universität Graz²⁸.

Modifikationen an der TEI

Wenn man nun die Urkundeneditionen unter Verwendung des Regelwerks der TEI näher untersucht, fällt auf, dass in allen Beispielen eine Modifikation der TEI verwendet wird: Im oben zitierten Beispiel aus den elektronischen Editionen der *Écoles des Chartes* war zum Beispiel die Transkription mit Hilfe einer Typisierung des allgemeinen Elements *div* markiert. Um das diplomatische Formular zu kodieren, wurde bei den *Anglo-Saxon Charters* zum Beispiel zunächst ein eigenes Schema verwendet, das erst nachträglich verallgemeinert und der TEI angepasst wurde. Dabei werden zum Beispiel diplomatische Textstrukturen wie das Urkundenformular als Spezialfälle von TEI-Textsegmenten (*seg*) verstanden²⁹. Die TEI-Fassung des *Corpus Altdeutscher Originalurkunden* bezeichnet die Rollen von Personen und Institutionen als Aussteller oder Empfänger mit dem Attribute *@type* zum allgemeinen TEI-Element *name*.

MitarbeiterInnen der *École des Chartes* haben ihre Modifikationen der TEI unter dem Namen „Diple“ zusammengefasst, welche nicht nur eine projektspezifische Modifikation des TEI-

²⁷ Zum Beispiel EditMOM im Rahmen des Projekts *monasterium.net* <https://github.com/icaruseu/mom-ca> oder der DHWriter <http://dhwriter.org>.

²⁸ Längere Erfahrung besitzen hier noch das Kompetenzzentrum elektronisches Publizieren der Universität Trier und das Göttingen Center for e-Humanities. Aber auch Fakultätsrechenzentren wie die Münchner IT-Gruppe Geisteswissenschaften können solche Dienstleistungen erbringen. Zusätzlich ist auch ein privater Markt an DienstleisterInnen in diesem Bereich im Entstehen.

²⁹ Vgl. die Code-Ausschnitte in CIULA / SPENCE, *Threads of integration* (wie Anm. 9), S. 43.

Schemas sind, sondern als allgemeines TEI-Schema publiziert vorliegen, zu dem es eine veröffentlichte Schemadokumentation und Transformationsszenarien gibt³⁰. Eine systematische Anpassung der TEI hat auch die Charters Encoding Initiative (CEI) versucht³¹, die auf der Vorgängerversion der TEI, der TEI P4³² aufbaut. Diese Gruppe von ArchivarInnen und DiplomatikerInnen hat so ein XML-Schema entwickelt, das für Urkunden spezifische Elemente enthält, wie zum Beispiel das diplomatische Formular, Elementnamen für die Rollen im Rechtsakt und Beglaubigungsmittel³³.

Digitale Urkundeneditionen ohne die TEI

Die TEI ist heute nicht mehr nur eine Angelegenheit von SpezialistInnen, sondern wird auch von Forschungsförderungsinstitutionen als Archiv- und Datenaustauschstandard empfohlen. Das Format sollte deshalb bei allen Projekten, welche die Wiederverwertung der digitalen Daten durch andere ForscherInnen wünschen, zumindest als Exportformat verwendet werden. Dennoch sind viele anspruchsvolle digitale Urkundeneditionen entstanden, die nicht die TEI verwendet haben.

Dazu gehört zum Beispiel die Edition der französischen Königsurkunden, welche die Académie des Inscriptions et Belles Lettres herausgibt³⁴. Auch die Texte der in Frankreich original überlieferten Urkunden aus der Zeit vor 1121 wurden nicht in TEI kodiert³⁵. Viele online gestellte Datenbanken geben nicht einmal an, welche Strukturen ihre Datenbanken haben³⁶.

Es gibt einige Gründe dafür, dass so viele Urkundeneditionen nicht die Standards der TEI verwenden. Ein gewichtiger ist, dass EditorInnen nach einer praktischen Lösung für ihre Edition suchen und deshalb lokale Datenbanksysteme beziehungsweise von ihren

³⁰ Schema, Dokumentation und Transformationsszenarien findet man unter <http://developpements.enc.sorbonne.fr/diple/schema/>. Zum Konzept dieses Schemas vgl. Camille DESENCLOS / Vincent JOLIVET, Diple, propositions pour la convergence de schémas XML/TEI dédiés à l'édition de sources diplomatiques, in: Digital Diplomatics. The Computer as a Tool for the Diplomatist?, hg. von Antonella AMBROSIO / Sébastien BARRET / Georg VOGELER (AfD Beiheft 14), Köln u. a. 2014, S. 23–30.

³¹ <http://www.cei.lmu.de>; vgl. dazu auch Georg VOGELER, Ein Standard für die Digitalisierung mittelalterlicher Urkunden mit XML. Bericht von einem internationalen Workshop in München 5./6. April 2004, in: AfD 50 (2004), S. 23–34, DERS., Towards a standard of encoding medieval charters with XML, in: Literary and Linguistic Computing 20 (2005), S. 269–280, und DERS.: Charters Encoding Initiative (CEI). Zu Möglichkeiten der Integration mit Hilfe eines Standards für Urkundendigitalisierung, in: Alte Archive – Neue Technologien, hg. von Thomas AIGNER / Karin WINTER, St. Pölten 2006, S. 181–198.

³² <http://www.tei-c.org/Vault/P4>.

³³ Ein Vorschlag, Urkunden nach den Strukturen der byzantinischen Herrscherdiplomatie zu kodieren, ist ohne technische Formalisierung geblieben: Sylvia POULIMENOU / Spyros ASONITIS / Marios POULOS, Metadata encoding for the documents based on the rules of diplomatics science, in: Metadata and Semantics, Proceedings of the 2nd International Conference on Metadata and Semantics Research 11–12 October 2007, hg. von A. SICILIA / M. D. LYTRAS, Berlin 2009, S. 91–100.

³⁴ Actes royaux, bearb. unter der Leitung von Hélyar XAVIER, <http://www.cn-telma.fr/actesroyaux/index/>.

³⁵ <http://www.cn-telma.fr/publication/chartes-originales-anterieures-1121-conservees-en-france>.

³⁶ So zum Beispiel die von der Niederländischen Akademie der Wissenschaften bearbeiteten Register der Grafen von Holland 1299–1345 <http://resources.huygens.knaw.nl/registershollandsegrafelijkhheid>, das Diplomatarium Danicum <http://diplomatarium.dk/dd/> oder die Cartae Cluniacenses electronicae <http://www.uni-muenster.de/Fruehmittlelalter/Projekte/Cluny/CCE/Welcome.htm>.

Institutionen zur Verfügung gestellte Infrastrukturen verwenden, deren Datenstrukturen nicht öffentlich gemacht werden. Es ist also die Frage zu stellen, wie es mit der TEI in der Editionspraxis aussehen könnte. Die TEI beruht auf XML, womit Strukturen und Annotationen direkt in den Text integriert werden können. XML ermöglicht datenbankartige Operationen, das heißt Suche nach markierten Textbestandteilen beziehungsweise Suche beschränkt auf bestimmte Textbestandteile. Deshalb gibt es auch einige Projekte, die ihre Daten mit anderen XML-basierten Strukturen speichern.

Der von Michele Ansani seit 2000 herausgegebene *Codice Diplomatico della Lombardia Medievale* (CDLM)³⁷ hat zum Beispiel eine eigene Dokumentgrammatik entwickelt, welche sich an den Editionsprinzipien gedruckter italienischer Editionen orientiert und zusätzlich Elemente für Registereinträge von Personen und Orten vorsieht, sowie das diplomatische Formular der früh- und hochmittelalterlichen Privaturkunde abbildet³⁸. Die Vorabpublikation der *Constitutiones Karls IV.* für die *Monumenta Germaniae Historica* (MGH), die an der Berlin-Brandenburgischen Akademie der Wissenschaften unter der Leitung von Michael Menzel ediert werden³⁹, verwenden eine an die TEI angelehnte XML-basierte Dokumentgrammatik, nämlich die Elementnamen der Charters Encoding Initiative (CEI)⁴⁰. Die unter der Leitung von Eliana Magnani und Marie-José Gasse-Grandjean in der UMR ARTeHIS-Arbeitsgruppe in Dijon bearbeiteten *Chartae Burgundiae Medii Aevi*⁴¹ verwenden das XML-Format des ARTFL-Projekts⁴². Sie nehmen damit Rücksicht auf die verwendete Datenbanksoftware Philologic. Die EditorInnen der Urkunden Karls IV. und der CDLM haben sich entschieden, für die Diplomatie spezifische Kodierungen zu verwenden. Warum verwenden diese Projekte nicht die TEI?

Text und Dokument: Modifikationsbedarf der TEI

Das Verständnis von Edition in der TEI und in Urkundeneditionen

Die TEI ist ein sehr allgemeines Regelwerk. Wenn ein spezielles Interesse nicht abgedeckt wird, gibt es entweder die Möglichkeit, eine frei formulierte, unstrukturierte Beschreibung zu erfassen oder ein allgemeineres Element in seiner Verwendung über ein Attribut zu spezifizieren. Freie Beschreibungen können zum Beispiel bei einigen äußeren Merkmalen angewendet werden, zu denen die TEI Kodierungen vorschlägt: Die Identifikation von Schreibern ist eine wichtige Aufgabe der Diplomatie, die bei der Edition von Urkunden eines Ausstellers Methoden gefunden hat, die Hände mit Siglen zu bezeichnen. In der TEI gibt es nur das allgemeine Element *scriptNote* zur Beschreibung von verwendeten Schriftarten, ohne die Merkmale, welche die Hand ausmachen, anders als in Fließtext beschreiben zu können. Kategorien wie Leitbuchstaben oder spezifische Gestaltungsmerkmale könnten dagegen

³⁷ <http://cdlm.unipv.it>.

³⁸ Die Dokumentation der Kodierung findet man unter <http://cdlm.unipv.it/progetto/documentazione>.

³⁹ <http://telota.bbaw.de/constitutiones/>.

⁴⁰ <http://www.cei.lmu.de>.

⁴¹ <http://www.artehis-cbma.eu/>.

⁴² <https://sites.google.com/site/philologic3/encoding/ate-artfl-text-encoding>.

nützlich sein, Schreiber anhand ihrer Merkmale auffinden zu können. Auch die Faltung des Pergaments kann urkundenkritische Relevanz haben⁴³. Die TEI kennt nur eine allgemeine Beschreibung des Beschreibstoffs (*support*). Für die Besiegelungsart sieht die TEI eine Unterscheidung vor, ob das Siegel zeitgenössisch ist oder nicht (*seal@contemporary*), eine nähere Beschreibung zur Befestigungsart, zum Siegelstoff, seiner Form oder seiner Legende ist aber nicht möglich. Das sind alles Fragen, die DiplomatikerInnen interessieren, die in der TEI unterspezifiziert sind. Sie bietet dafür die Möglichkeit, freien Text zu erfassen. Die Vorteile einer strukturierten Erfassung gehen verloren.

Zu den Besonderheiten von Urkundeneditionen, die in der TEI mit Hilfe von Spezialisierung allgemeiner Elemente kodiert werden können, gehören zum Beispiel Maßangaben unter Berücksichtigung der Plica. Sie müssen auf diese spezielle Faltung Rücksicht nehmen und geben deshalb gewöhnlich zwei Werte an, nämlich für das Pergament mit eingefalteter Plica und für die Plica selbst. In der TEI gibt es für die Angabe der Maße das Element *dimensions*. Seine Definition zeigt, dass dabei vorrangig an Buchseiten gedacht wurde. Die Plica müsste also zum Beispiel über die Referenz auf ihre Definition im Vocabulaire International de la Diplomatique (VID)⁴⁴ im Attribut *@scope* kodiert werden. Das diplomatische Formular kann so als Spezialfall eines Textsegmentes *seg* gelten, für das die TEI das Attribut *@function* als Mittel zu einer Spezialisierung vorschlägt. Der Datierungsstil kann im Attribut *@datingMethod* des Elements *date* angegeben werden. Der Cursus lässt sich als metrisches System *metSym* beschreiben, das in Sätzen *s*, Sprachklauseln *cl* oder Phrasen *phr* verwendet und im Attribut *@met* kodiert wird. Die Funktionen von Personen sowie von im Urkundentext oder Regest erwähnten Rechtssachverhalten können als Interpretation des linguistischen Phänomens Personennamen *persName* oder Fachterminus *term* mit Hilfe des Attributs *@ana* kodiert werden.

Generell stellt die TEI das Attribut *@type* für sehr viele Kodierungselemente zur Verfügung, das es auch ermöglichen würde, ein diplomatisches Textelement als Spezialisierung eines der abstrakten Textelemente der TEI zu kodieren. Diese Spezialisierungen sind aber nicht expliziter Bestandteil der TEI. Wenn die Kodierungen im Datenaustausch eindeutig bleiben sollen, müssten sie entweder detailliert dokumentiert oder besser noch mit einem Verweis auf eine externe Definition beschrieben werden. Dafür bietet sich insbesondere das Vocabulaire International de la Diplomatique an, das auch als formalisierte Ontologie vorliegt⁴⁵. Diesen Weg beschreitet zum Beispiel auch die TEI-Modifikation der Écoles des Chartes für das Urkundenformular⁴⁶.

⁴³ Vgl. zum Beispiel Paul BERTRAND: De l'art de plier les chartes en quatre, in: Gazette du livre médiéval: Bulletin semestriel 40 (2002), S. 25–35.

⁴⁴ <http://www.cei.lmu.de/VID/VID.php?543>.

⁴⁵ <http://www.cei.lmu.de/VID>. Auf die Bedeutung des VID für die Kodierung von diplomatischen Texten hat schon 1999 Michele Ansani hingewiesen: Michele ANSANI, Diplomatica (e diplomatisti) nell'arena digitale, in: Scrineum 1 (1999), S. 175–196, URL: <http://dohc.unipv.it/scrineum/ansani.htm>, hier S. 184. Zu seiner Aufbereitung als formalisierte Ontologie vgl. Georg VOGELER, Von der Terminologie zur Ontologie. Das „Vocabulaire international de la diplomatique“ als Ressource des Semantic Web, in: Francia 40 (2013), S. 281–297.

⁴⁶ Vgl. http://elec.enc.sorbonne.fr/cartulaires/schema#h_9.

Diplomatische Konzepte von Text und Dokument

Bei einem Vergleich des Regelwerks der TEI mit für Urkundeneditionen spezifischen Modifikationen und den davon unabhängigen XML-Kodierungen ergeben sich folgende Bereiche, in denen die TEI-Richtlinien für Editionen von Urkunden unzureichend zu sein scheinen. Erstens kennt die TEI keine spezifischen Elemente, um den Rechtsakt präzise zu beschreiben: Aussteller und Empfänger, Notar und Zeugen, Art des Rechtsgeschäfts und so weiter fehlen als Kodierungselemente. Zweitens besteht ein Interesse, auch das diplomatische Formular im Text zu kodieren. Drittens kann die Überlieferungsdokumentation auf die Beglaubigungsmittel mehr Rücksicht nehmen, als die TEI es tut: Eine paläographisch zeitgenössische Fassung einer Urkunde ohne Beglaubigungsmittel ist wahrscheinlicher ein Entwurf als ein „Original“, eine beglaubigte Kopie gibt andere Informationen über die Vorlage als eine einfache Kopie.

Dieser Modifikationsbedarf weist auf eine besondere Funktion von Urkundentexten hin, welche sie vom Textverständnis der TEI unterscheidet. Während die TEI nämlich ihren Schwerpunkt auf eine linguistische Interpretation von Text legt, sind Urkunden mehr als nur Sprachzeugnisse, denn sie sind auch Aussagen über Rechtstatsachen. Sie sind damit etwas, das ich im Folgenden als „Dokument“ bezeichnen will. Dabei meine ich die Kombination von physischer Realisierung und Text, welche den im Text formulierten Rechtssachverhalt zur Rechtstatsache werden lässt.

Der Modifikationsbedarf ist darin begründet, dass Urkundeneditionen mehr sind als die von TEI abgebildeten philologischen Editionen. Zwar hat Michele Ansani zu Recht argumentiert, dass die Rekonstruktion des Originaltextes und damit das klassische Interesse der Textkritik auch in der Urkundenedition liegt⁴⁷. Zu einer Urkundenedition gehört aber auch das *discrimen veri ac falsi*. EditorInnen von Urkunden wollen nicht nur den Originaltext rekonstruieren, sondern auch eine Aussage über den originalen Rechtsinhalt machen. Eine Urkundenedition ist deshalb mehr als nur eine Textwiedergabe, denn mit der kritischen Beurteilung der Überlieferung des Textes und mit der Rekonstruktion seiner Entstehung macht die diplomatische Edition eine Aussage über die Glaubwürdigkeit eines Textes und seines Inhalts.

Die Diplomatik kann sich dabei auf die Beglaubigungsmittel und -verfahren bei der Entstehung der Urkunde beziehen. Dazu untersucht sie die Echtheit der Siegel und Unterschriften beziehungsweise der Notarsunterfertigung. Sie verwendet auch die von den Zeitgenossen formulierten Regeln der sprachlichen Gestaltung der Texte, die ich unter dem Begriff „Solemnitäten“ zusammenfassen will, wobei Kanzleiregeln wie der *stilus curiae* ebenso gemeint sind wie die Solemnitäten der notariellen Urkunden. Diplomatik verwendet darüber hinaus retrospektive Methoden der Authentifizierung, indem sie aus den überlieferten Dokumenten den kanzlei- oder zeittypischen Gebrauch der Urkundengestaltung ableitet. Schließlich kann man aus der modernen Archivwissenschaft auch lernen, dass die Art und Weise der Überlieferung Glaubwürdigkeit für Dokumente erzeugen kann⁴⁸.

⁴⁷ Michele ANSANI, *Edizione digitale di fonti diplomatiche. Esperienze, modelli testuali, priorità*, in: *Reti Medievali* 7/2 (2006) <http://www.dssg.unifi.it/RM/rivista/forum/Ansani.htm>.

⁴⁸ Vgl. die elaborierte Taxonomie von Luciana DURANTI, *The return of diplomatics as a forensic discipline*, in: *Digital Diplomatics*, Köln u. a. 2014 (AfD Beiheft 14), S. 89–98.

Diese Überlegungen über Echtheitskritik sind nicht auf mittelalterliche Urkunden beschränkt, denn auch digitale Dokumente verwenden Beglaubigungsmittel, die zum Beispiel als Verschlüsselung Teil des Dokuments werden oder die sich wie die Notarsunterschriften aus der Glaubwürdigkeit von beglaubigenden Institutionen (Zertifikatsaussteller) ableiten. Die digitale Forensik wiederum überprüft aus der Konsistenz von Dokumenten und ihrem Überlieferungskontext retrospektiv ihre Glaubwürdigkeit. Kodierungen, mit denen DiplomatikerInnen in einer Urkundenedition ihre Expertise über die Echtheit des edierten Dokuments erfassen, könnten also auch auf moderne digitale Dokumente angewendet werden.

Auch die TEI kennt Kodierungsformen für Glaubwürdigkeit. Die Attributgruppe *att.Responsibility* enthält Attribute für die Zuverlässigkeit einer Angabe und die für die Interpretation verantwortlichen EditorInnen (*@cert*, *@resp*)⁴⁹. Die Elemente, zu denen man den Regeln der TEI folgend solche Attribute angeben kann, weisen jedoch deutlich darauf hin, dass die TEI das diplomatische Konzept von Glaubwürdigkeit damit nicht beschreiben will. Sie sind nämlich auf Interpretationen, Elemente der Textkritik und der Transkription anwendbar, nicht aber auf einen ganzen Text. Glaubwürdigkeit ist für die TEI also eine Frage des Verhältnisses zwischen Befund im originalen Dokument und seiner editorischen Interpretation, nicht eine Frage des Verhältnisses zwischen überliefertem Text und historisch belegtem Rechtssachverhalt.

Wie kann man nun das von der Diplomatie verwendete Konzept von Glaubwürdigkeit in ein Modell fassen, das sich für den Computer verarbeitbar formalisieren ließe? Ein historisches Dokument wird dadurch glaubwürdig, dass erstens der *Entstehungsprozess* nachvollzogen und die Glaubwürdigkeit der am Entstehungsprozess beteiligten Personen und Institutionen beurteilt werden kann, dass es zweitens in sich und zu anderen Dokumenten *konsistent* ist, also zum Beispiel die inneren und äußeren Merkmale dem zeitgenössischen Gebrauch entsprechen beziehungsweise historisch erklärbar sind, und dass drittens die *Überlieferung* die Beglaubigungsmittel erhalten hat, so dass eine Kopie vorrangig nach ihrem Diktat als nach ihren äußeren Merkmalen beurteilt werden sollte.

Wenn man diese Kategorien in die TEI integrieren wollte, dann müsste man Kodierungen entwickeln, welche den Entstehungsprozess eines Dokuments abbilden. Dieser mehrstufige Prozess unterscheidet zum Beispiel zwischen der Sachentscheidung, einer Protokollierung des dabei besprochenen Sachverhalts, der Ausfertigung und Schritten vor der Aushändigung zum Beispiel durch Gegenzeichnung, Besiegelung oder Paraphe beteiligter Funktionäre. Die für die Beschreibung von Handschriften eingeführte Kodierung der *filiation*, das heißt des Verhältnisses eines handschriftlichen Textzeugen zu anderen des gleichen Textes als *protograph*, *antigraph* und *apograph* (Erstschrift, Vorlage, Abschrift)⁵⁰, müsste um das Konzept des Verhältnisses zwischen Rechtssachverhalt und Text (Fälschung, Ausfertigung, Konzept) erweitert werden. Die Rollen und Kategorien der in der Urkunde dokumentierten Rechtshandlung müssten kodierbar sein (Aussteller, Empfänger, Art des Rechtsakts, Ausstellungs-/Handlungsort et cetera). Schließlich bräuchte es eine Kodierung für die

⁴⁹ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.responsibility.html>.

⁵⁰ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html#msfil> und <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-filiation.html>.

Beglaubigungsmittel in ihrer rechtlichen Funktion, also nicht nur als Teil der Beschreibung der äußeren Merkmale (wie die TEI die Kodierung *seal* verwendet)⁵¹, als bibliographische Angabe auf einem Titelblatt (*byline*)⁵² oder als abschließender Bestandteil eines Briefes (*signed*)⁵³. Wie oben angedeutet sind solche Kodierungen nicht auf mittelalterliche Urkunden beschränkt. Eine Kodierung von Beglaubigungsmitteln könnte zum Beispiel auch die technische Integrität von Dateien durch Prüfsummen oder Verfahren der digitalen Signatur einschließen.

Fazit

Die TEI ermöglicht die Beschreibung des einzelnen Textzeugen beziehungsweise rekonstruierten Textes und die Identifikation historischer Kerninformationen (Orte, Personen, Sachbegriffe). Gleichzeitig ist die TEI ein etabliertes und gutes Werkzeug für Editionen, das von einer breiten NutzerInnen- und ExpertInnengruppe unterstützt wird und als Langzeitarchivierungsformat etabliert ist. Eine neu zu planende digitale Urkundenedition muss sie also in die Entscheidungen über die Methode mit einbeziehen.

Die TEI ist von einem philologischen Editionsmodell geprägt. Sie legt deshalb ein besonderes Schwergewicht auf die linguistischen und die physischen Eigenschaften von Text. Darauf können diplomatische Editionen jedoch nicht reduziert werden. Die TEI bietet keine Kodierungen an, die Sachaussage eines Textes und das Verhältnis zwischen Inhalt und Form der Dokumentation zu beschreiben. Die Regeln der TEI reichen nicht aus, um die diplomatische Beurteilung von Dokumenten zu kodieren. Die TEI müsste also für Urkundeneditionen erweitert werden. Die Konzepte, die dabei in die TEI eingeführt würden, wären aber nicht nur Spezialisierungen der allgemeineren Elemente der TEI, sondern eine Erweiterung des Textverständnisses. DiplomaterInnen kodieren nicht nur linguistisch und philologisch verstandene Texte, sondern Dokumente, das heißt die glaubwürdige Verbindung von physischem Dokument und in Sprache gefasster Aussage über Rechtstatsachen. Eine diplomatische Erweiterung der TEI würde also nicht nur die Grundlage für diplomatische Editionen legen, sondern sie würde Konzepte wie Glaubwürdigkeit und Rechtsaussagen zu den Kodierungsmöglichkeiten der TEI beitragen. Diplomatie kann nämlich eine substantiell neue Sicht in die Diskussion über das Verständnis von Text im Bereich der digitalen Geisteswissenschaften einbringen.

⁵¹ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-seal.html>.

⁵² <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-byline.html>.

⁵³ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-signed.html>.