

Transkription und Annotation. Die Eingabeseite digitaler Editionen

Stephan Kurz

Der Artikel führt ein in den aktuellen Stand der Grundlagen und Techniken bei der Erstellung von (Quellen-) Editionen geschichtswissenschaftlicher und verwandter Art. Editionen verfahren in Kenntnis der Quellenlage stets regelhaft; für digitale Editionen – als Normalfall jeden Edierens im 21. Jahrhundert – werden die Vorschläge der *Text Encoding Initiative* (TEI) vorgestellt. Editionen folgen grundsätzlich einer Logik von Eingabe, Verarbeitung und Ausgabe, sie trennen analytisch immer schon zwischen Struktur- und Darstellungsphänomenen. Das erste Ergebnis der Editionsarbeit sind strukturierte Daten.

Es gibt für die geschichtswissenschaftliche Arbeit im analogen wie im digitalen Raum sehr unterschiedliche Quellen. Für Historiker:innen ist es nicht nur wichtig, darüber Bescheid zu wissen, welche Quellentypen und Quellen für ihre jeweilige Forschungsarbeit zur Verfügung stehen, denn mitunter gehört es zu den Erfordernissen geschichtswissenschaftlicher Forschung, Quellen selbst zu edieren und aus unstrukturierter natürlicher Sprache in strukturierte Daten umzuwandeln.

Der vorliegende Beitrag führt zunächst in Grundbegriffe des Edierens ein, um dann anhand der wichtigsten Werkzeuge zur Bereitstellung digitaler wissenschaftlicher Editionen den Stand der Technik vorzustellen. Dabei werden beispielhaft zentrale Elemente der Textwiedergabe mithilfe des von der TEI vorgeschlagenen Vokabulars vorgestellt.

1 Editionen

Zu Edierendes: Edenda

Aufgabe einer Edition ist es, Quellen verfügbar zu machen. Je nach Ausprägung der spezifischen Edition und den Erfordernissen der betreffenden Quellen gehört dazu auch, Bestände und Zustände aus unterschiedlichen Überlieferungen zusammenzuführen und dabei Chronologie und Abhängigkeiten zu etablieren, einen bestimmten textuellen Status nach transparenten Vorgaben herzustellen, etwa durch Kommentar zu kontextualisieren, um dadurch Quellen zu erschließen, in einen Zusammenhang zu stellen, zu bewahren und zu vervielfältigen. Dabei ist es zentral, die eigenen Eingriffe in einen Textzustand von dem durch die Urheber:innen Überlieferten zu trennen und dieses Vorgehen bei den Prozessen der Edition für die Leser:innen zu dokumentieren. Ziel einer wissenschaftlichen Edition ist es, das zu Edierende in „möglichst authentischer Form“ (Grubmüller & Weimar 2007) zu bieten. Für das zu Edierende steht der Begriff des *Edendums*:

„Die Grenzen der Edition liegen nicht an den Grenzen von Werken (z. B. literarische, musikalische, bildnerische), Texten (auch des weitesten Textbegriffs) oder Dokumenten (z. B. Briefe, Archivalien, Objekte, audiovisuelle Medien, born digital-data). Grundsätzlich ist jede kulturelle Äußerung unabhängig von ihrer Herkunft, Modalität, Medialität und Materialität editionsfähig“ (Institut für Dokumentologie und Editorik & DHd2022 2022, Absatz 2).

Geschichte der Edition

In diesem Sinne sind die Anfänge der Edition dort zu suchen, wo Texte überhaupt vervielfältigt wurden – *ad fontes* zu gehen hieße dann, mit der Geschichte der Schrift zu beginnen. Die wissenschaftliche oder kritische Edition, die ihre eigenen Voraussetzungen und Handlungen mittransportiert, hat ihre Ursprünge im Zeitalter des Humanismus, als zahlreiche Quellen der Antike neu herausgegeben wurden und mit dem beginnenden Druckzeitalter (Giesecke 1998 spricht vom *Typographium*) neu zirkulieren konnten. Eine besondere Bedeutung hat die Edition wissenschaftsgeschichtlich in den „positivistischen“ Unternehmungen der zweiten Hälfte des 19. Jahrhunderts erlangt, als sowohl in der Geschichtsschreibung (Geschichtswissenschaft)

als auch der Literaturgeschichtsschreibung (Literaturgeschichte als Königsdisziplin in den Philologien) große Editionsunternehmen begonnen wurden, teilweise explizit als nationale Projekte begründet. Im Spannungsfeld von Bewahrung alter Quellen und Erschließung im Horizont des jeweils aktuellen wissenschaftlichen Diskurses hat sich seit damals die Editionswissenschaft als eigenes Fach etabliert.

Die Edition geht in ihrem Tun geleitet von jeweils zu entwickelnden Editionsprinzipien vor. Die Editionswissenschaft hat ein ausdifferenziertes Instrumentarium entwickelt, um die Vorgänge der Textkritik zu systematisieren (eine Reihe von Fachbegriffen siehe Infobox 1 u.).

Editionstechniken

Eine Edition befasst sich im Wesentlichen mit der kontextualisierenden Wiedergabe von materialgebundenen Inhalten. Für die Geschichtswissenschaften sind das hauptsächlich medial überlieferte Dokumente textueller Natur. Grenzwerte dazu wären etwa: die Musikedition, sie erfasst Notentexte (siehe auch den Beitrag von Andrea Lindmayr-Brandl in diesem Band); manche Altertumswissenschaften edieren Tonscherben als *text bearing objects*; die *Oral history* muss beispielsweise Tonbänder aus den 1950er-Jahren erhalten.

Wichtig ist für eine Bestimmung der Edition neben der Art ihrer Edenda aber auch, dass es in den meisten Fällen nicht um bloße Reproduktion geht, die der Erhaltung, Überlieferung und dem Zugänglichmachen des kulturellen Erbes dienen würde, sondern dass die Edition ein Meta-Wissen über die edierten Quellen hinzufügt und im Minimalfall zumindest evidente Fehler der Überlieferung tilgt und dies transparent dokumentiert. Die bloße Sicherung von Archivquellen auf Mikrofilm oder in Bilddigitalisaten allein ist daher noch keine Edition. Auch hierfür ein Beispiel: Ausgehend von der Frankfurter Hölderlin-Ausgabe (D. E. Sattler, ab 1972) hat sich in der Germanistik für die textgenetische Edition ein spezifischer Typ der Edition bewährt, der ganz wesentlich auf Faksimiles und deren diplomatischer Umschrift beruht – auch in der Präsentation im Druck. Dieses Verfahren mag oberflächlich betrachtet als Anti-Edition erscheinen, es ermöglicht aber gerade ein Nachvollziehen editorischer Entscheidungen (für eine Lesart, für ein Komma statt eines physisch im Manuskript zu liegen gekommenen Fliegenbeins), die vor dem Hintergrund einer angenommenen auktorialen oder aufzeichnenden Intention dann transparent gemacht werden können.

Quelle/Edition

Festzuhalten bleibt: Eine Edition ist stets mehr als Reproduktion. Eine edierte Quelle ist etwas anderes als die ihr zugrunde liegende Quelle erster Ordnung. Editionen „gehören, welchen Anspruch sie auch immer haben, zu wissenschaftlichen Großunternehmungen, sind zeit-, arbeits-, personal- und kostenintensiv“ (Plachta 2019, 11).

Infobox 1 – Fachbegriffe der Editionswissenschaft, in der Reihenfolge der Textkritik

Heuristik

Suche und Sammlung von Textzeugen

Kollation

Abgleich zu edierender Textzeugen

Recensio

Variantenanalyse, Auffinden von Lesarten, dabei entsteht nach Möglichkeit ein

Stemma

„Stammbaum“ der Textzeugen, aus dem die Chronologie und Abhängigkeit ablesbar ist

Emendation

Verbesserung, Fehlerkorrektur, mitunter auch

Normalisierung

von Schreibweisen, Abkürzungen, Rechtschreibung und Zeichensetzung

Konjektur

„Deutung“, „Vermutung“, inhaltliche und stilistische Korrektur

Textkritik

Editionen haben in der Textkritik ein standardisiertes Verfahren (siehe Infobox 1; vgl. auch Schwentner 2021), das abhängig vom Editionsgegenstand und von den meist fachspezifischen Erkenntnisinteressen unterschiedlich gewichtet. Auch bedingen international variierende Wissenschaftstraditionen unterschiedliche Herangehensweisen an Editionen und deren Konzepte (Plachta 2012). Die Editionen der philologischen Fächer fokussieren, da sie zumindest implizit dem Paradigma einer auktorialen Intention folgen (Martus 2007; Bosse 1981), auf die Herstellung gültiger Texte (auch sie, nicht nur die Kommentare einer Edition, sind stets Kinder ihrer Zeit und des je aktuellen Standes der Wissenschaft, was auch neue Editionen klassischer Autor:innen rechtfertigt), während andere Disziplinen die Quellenkritik im Editionsprozess eher auf den wissenschaftlichen Kommentar auslagern.

Je nach Gattung oder Textsorte der Edenda erübrigt sich oft auch die Suche nach Varianten und Textzeugen – von eigenhändigen Briefen, Tagebüchern oder von geheimen Sitzungsprotokollen existieren üblicherweise schlicht keine Varianten, sondern allenfalls Abschriften oder technisch erzeugte Kopien.

Wissenschaftliche Editionen verfolgen wissenschaftliche Ziele. Diese können sehr unterschiedlich gesteckt sein: Geht es etwa um die dringende Konservierung eines Korpus von drei neu aufgefundenen fragilen Papyrusfragmenten durch Faksimilierung? Oder darum, der Forschung möglichst rasch einen Überblick über die Entscheidungsprozesse eines Regierungsorgans der Neuzeit über einige Jahrzehnte zu ermöglichen? Im ersten Fall wird die Edition allein aufgrund des geringeren Quellenumfanges ausführlicher auf Informationen zum Textträger abheben können und jedenfalls Abbildungen einbeziehen, im zweiten Fall werden aus Dutzenden Aktenlaufmetern vielleicht „nur“ Metadaten und Verschlagwortung zur besseren Auffindbarkeit als Regestenedition erstellt werden können.

Editionsziele

Je nach Zielgruppe, Umfang des Quellenbestandes, den wissenschaftlichen Zielen der Edition und nicht zuletzt auch der personellen und finanziellen Ausstattung eines Editionsunternehmens kommen dabei unterschiedliche Editionsmodelle zur Anwendung (siehe Infobox 2).

Editionsmodelle

Infobox 2 – Editionsmodelle

Regestenedition

Sie nimmt die Quelle/n dem Inhalt nach auf, indem sie (meist kurze) Zusammenfassungen (Regesten) erstellt. Beispiel: die *Regesta Imperii* (<http://www.regesta-imperii.de>)

Historisch-Kritische Edition

Sie nimmt alle verfügbaren Lesarten zu einer Quelle auf und unterzieht sie kritischer Bewertung sowie textgenetischer Analyse, meist mit einem ausführlichen Kommentar. Ausgeführt ist sie als

- diplomatische Ausgabe, die eine Quelle anhand eines Leittextzeugen in dessen ursprünglicher Form wiedergibt, oder als
- eklektische Ausgabe, welche Bestandteile verschiedener Zeugen kombiniert, wenn etwa kein Textzeuge einen vollständigen Text bietet.

Wissenschaftliche Studienausgabe

Sie ermöglicht eine eingehende Beschäftigung mit der Quelle, der Kommentar und fallweise auch die Lesarten sind weniger ausführlich gehalten. Dieses Editionsmodell verzichtet auf die Vollständigkeit der Erfassung aller Überlieferungen zugunsten eines gesicherten Textes.

Auswahledition

Anders als die meisten historisch-kritischen Editionen bietet sie keine vollständige Überlieferung ihres Editionsgegenstandes, sondern eine – jeweils zu begründende – Auswahl.

Leseausgabe

Sie verfolgt das Ziel, eine Quelle für die Lektüre verfügbar zu machen, nicht unbedingt für die wissenschaftliche Auseinandersetzung damit, und richtet sich an ein breiteres Publikum.

1.1 Digitale Edition

Edenda heute Im 21. Jahrhundert ist die digitale Edition der Normalfall des Edierens geworden (Institut für Dokumentologie und Editorik & DHd2022 2022, Abs. 3). Diesem Umstand tragen auch die Forschungsförderungsinstitutionen bei ihrer Förderpolitik Rechnung. Digitales Editieren ist gewissermaßen die Fortsetzung der Editionspraxis mit Mitteln des digitalen Mediums. Die oft zitierte zirkelschlüssige Definition von Patrick Sahle (et al. 2014), eine digitale Edition sei eine Edition, die einem digitalen Paradigma folgt, lässt sich vor dem Hintergrund der von ihm klar ausdifferenzierten Textbegriffe nochmals herausarbeiten: Der Text (oder auch: das Edendum) kann betrachtet werden als abstrakte Idee, als Werk mit einer Struktur, als sprachlicher Ausdruck, als eine Fassung, als ein materielles Dokument und als Container visueller Zeichen (Sahle 2013, 9–64, dort auch eine Visualisierung als „Textrad“). Sahles Ausführungen beziehen sich klar auf Begriffe von Text, die der Edition zugrunde liegen; dies ließe sich aber auch auf beliebige *Edenda* ausweiten (zum Textbegriff vgl. auch Reuß 2005).

Analog/digital In Bezug auf die Mittel der Produktion, Zirkulation und Rezeption gilt dabei: „[D]ie Essenz des gegenwärtigen Medienwandels ist nicht der Wechsel von analogen zu digitalen Medien, sondern der Übergang vom Denken in Medien zum Denken in Modellen und Daten“ (Sahle 2010). Auch die damit ins Zentrum gerückten „Daten“

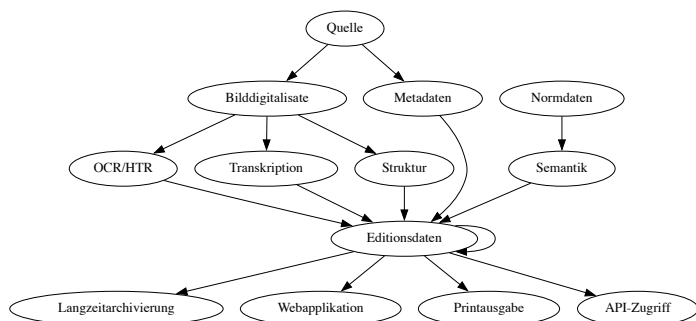


Abb. 1 Schematischer Abriss der Vorgehensweise bei der Eingabe, Verarbeitung und Ausgabe von Daten für die digitale Edition; eigenes Werk Stephan Kurz, CC-BY 4.0 <https://creativecommons.org/licenses/by/4.0/deed.de>.

sind dabei nichts epochal Neues, sie waren schon in der Arbeit mit Karteien und Zettelkästen, wie sie die großen Editionsunternehmen des ausgehenden 19. Jahrhunderts geführt haben, letztlich Grundlage einer inhärent vernetzten Arbeit (vgl. Petschar et al. 1999; Krajewski 2002; zur Methodenkritik digitaler Geschichtswissenschaft siehe die Beiträge von Mareike König sowie Aline Deicke und Stefan Schmunk in diesem Band). Im digitalen Medium kommt die technische Vernetzbarkeit von Daten hinzu, zudem orientieren sich digitale Editionen an den *FAIR-Prinzipien* (Findable, Accessible, Interoperable, Reusable) und sind üblicherweise in Form frei lizenzierter Daten und Applikationen offen verfügbar.

Die grundlegenden Abläufe bei der digitalen Edition (Eingabe – Verarbeitung – Ausgabe) zeigt Abb. 1. Dieser Beitrag konzentriert sich auf Eingabe und Verarbeitung (zur Ausgabeseite des Editionsprozesses siehe den Beitrag des Verfassers zu „Daten und Schnittstellen. Die Ausgabeseite digitaler Editionen“ in diesem Band).

Zur Vernetzung von Wissensbeständen hier ein konkretes Beispiel aus der Edition zu einem Briefindex zur Korrespondenz des habsburgischen Großbotschafters Damian Hugo von Virmont mit Karl VI. Im Rahmen des Friedens von Passarowitz/Požarevac 1718 war zwischen Sultan Ahmed III. und dem Kaiser die wechselseitige Entsendung je einer Großbotschaft vereinbart worden. Die Korrespondenzen, Reiseberichte, Zeitungsartikel, Protokollregistereinträge

Workflows

Briefedition: ein Beispiel

und anderes Material, das auf beiden Seiten zahlreich überliefert ist, werden in einem 2020 begonnenen Projekt unter Beteiligung des Verfassers am Institute for Habsburg and Balkan Studies an der Österreichischen Akademie der Wissenschaften ediert (<https://qhod.net/>). Viele Dokumente der Großbotschaft Virmonds 1719–1720 sind u. a. im Österreichischen Staatsarchiv in Wien überliefert. Darunter befindet sich auch eine zeitgenössisch angefertigte handschriftliche Liste der 57 Stücke umfassenden Hofkorrespondenz des Großbotschafters „Samt ainem indice und kurzen extract über die darinen enthaltene materias“ (ediert als <https://qhod.net/o:vipa.i.hbg.1720>). Der erste Eintrag bezieht sich auf einen Brief aus dem heutigen Rumänien auf der Hinreise der Gesandtschaft nach Konstantinopel.

Verdatung und Entitäten

Er lautet: „Nr. 1 d. d.: Erdöd 27. Maii 1719 Berichtet den fortgang ohne beylag seiner rause, und daß ein page des grafen Bathyan einen laggey des grafen von Türheimb erstochen.“ (N. N. 2022). Die Nummerierung verweist auf einen Brief, das Datum lässt sich ohne Schwierigkeiten als maschinell verarbeitbares ISO-Datum (1719–05–27) ausdrücken; die Referenz auf Erdöd (heute bosnisch-kroatisch-serbisch Erdut) lässt sich durch Identifizierung und Disambiguierung mithilfe eines Identifikators, z. B. jenem aus GeoNames (<https://www.geonames.org/3200884/>) vereindeutigen. Dass die Diener Carl Graf von Bathyány (https://d-nb.info/gnd/116082291) und Johann Wilhelm Graf von Thürheims (kein GND-Eintrag) gemeint sind, die in den Quellen namenlos bleiben, lässt sich ebenso festmachen wie deren jeweilige Beziehung zu ihrem genannten Dienstherrn. Ein Zettelkasten hätte bis hier Abteilungen zu führen für: Briefe/Dokumente, Orte, Daten, Personen und für die möglichen Relationen zwischen diesen (z. B. Person ist Diener von Person, Ort ist Geburtsort von Person, Person ist Mitglied von Institution, Datum ist Sterbedatum von Person, Person ist erwähnt in Werk/Brief usw.). Dass sowohl hinter dem Digitalisat des Briefs Nr. 1 (ediert als von Virmont 2022, <https://qhod.net/o:vipa.i.hbg.17190527>) als auch hinter jenem des edierten Briefverzeichnisses auch ein physisches Objekt mit Eigenschaften, einem Bestandsnachweis und einer digitalen Reproduktion mit einem bestimmten Dateinamen steht, macht die Datenstruktur nur größer, ohne sie konzeptuell überkomplex zu machen: All dies ließe sich auch mithilfe eines Zettelkastens abbilden.

Derartige Datenstrukturen müssen 1. gut geplant sein, um alle möglichen Fälle abzudecken; 2. müssen Eingabeformate und -werkzeuge zur Verfügung stehen, um die Daten komfortabel erfassen und bearbeiten zu können; 3. müssen Darstellungsweisen gefunden oder geschaffen werden, um die Daten auch wieder zugänglich zu machen. Im Zeitalter der gedruckten Edition waren das Werkzeug für die drei Schritte: 1. Zettelkästen, 2. penible Datenein- und -ausgabe aus diesen Zettelkästen, 3. die Zusammenarbeit mit gut ausgebildeten Setzer:innen und Korrektor:innen. Die digitale Edition verfährt ähnlich, mit dem Unterschied, dass der Umgang mit Regeln und Werkzeugen des analogen Ordners und des Buchdrucks durch meist autodidaktisches Desktop-Publishing und durch die Vervielfältigung der Ausgabemedien ersetzt worden ist. Der grundlegende Ablauf von Eingabe – Verarbeitung – Ausgabe hat sich dabei nicht gewandelt.

Eingabe –
Verarbeitung –
Ausgabe

2 Eingabe und Verarbeitung in TEI-XML

Möglich gemacht wird die „Verdatung“ digitaler Editionen mithilfe technischer Lösungen. Die weitaus am häufigsten verwendeten Technologien bauen auf XML-Markup auf, das sich in den vergangenen Jahrzehnten gemeinsam mit den Schwestertechnologien XPath, XQuery, XSL-T und einigen weiteren als stabile technische Lösung für sämtliche oben angesprochenen Herausforderungen erwiesen hat (Becher 2009; W3C 2009; vgl. dazu auch Vogeler & Sahle 2017 sowie den Beitrag von Barbara Denicolò und Christina Antenhofer in diesem Band). Als De-facto-Standard haben sich die Empfehlungen der Text Encoding Initiative (TEI Consortium 2022) herauskristallisiert, die von einer umfassenden Gemeinde von digitalen Editor:innen zum Zweck der gleich- oder doch ähnlichförmigen Kodierung von Textphänomenen seit 1994 (mit Vorstufen) entwickelt werden. Dass damit ähnlich wie in der Bibliothekswelt Titel- und Personendaten maschinenlesbar austauschbar gemacht werden können, zeigt die Wichtigkeit und Nützlichkeit so eines Standards auf. Dazu drei Beispiele: So wird erstens eine Streichung in einem Brief von Karoline von Günderrode ebenso als `` markiert wie jene in einem Konzept zu einem Großveziersbefehl oder in einer zu edierenden Seite eines Katasters

TEI-XML

oder Grundbuches. Zweitens beinhalten TEI-codierte Texte durch den Metadatenblock `<teiHeader>` auch ihre Beschreibung und die Dokumentation der Editionsprinzipien in einer einzigen Datei. Und drittens können Texte durch strukturelles Markup in hierarchisierte Abschnitte (z. B. `<front|body|back>` und `div`, welche Absätze `<p>`, Listen `<list>` uvm. enthalten) untergliedert werden.

Alternativen zu TEI-XML

Eine gute Einführung in TEI-XML bietet Burnard (2014). Die Vorschläge der TEI werden zumindest als menschen- und maschinenlesbares Archivierungs- und Datenaustauschformat Bestand haben. Alternativen zur Modellierung in TEI wären Text-as-Graph oder andere Markup-Sprachen. Reines JSON und JSON-LD sind an der Grenze des Menschenlesbaren, gewinnen aber in der Applikationsentwicklung für die Publikation von Editionsdaten an Bedeutung.

TEI-XML als Standard

Die in TEI-XML-Dokumenten codierten Informationen sind sowohl für den Menschen (der in den TEI-Guidelines nachlesen kann, wie das betreffende Element bestimmt ist) als auch für die Maschine auslesbar: Ging es bei der maschinenlesbaren Erfassung von bibliographischen Metadaten in der bibliothekarischen Formalerschließung darum, z. B. nur das „Titel“-Feld durchsuchbar zu machen, geht es bei der standardisierten Erfassung von Textdaten um die gleichartige Auffindbarkeit von Phänomenen des Textes (des Edendums) selbst. Mindestens ebenso wichtig ist, dass der TEI-Standard auch durch eigene Elemente erweitert werden kann:

„In brief, the TEI Guidelines define a general-purpose encoding scheme which makes it possible to encode different views of text, possibly intended for different applications, serving the majority of scholarly purposes of text studies in the humanities. Because no predefined encoding scheme can possibly serve all research purposes, the TEI scheme is designed to facilitate both selection from a wide range of predefined markup choices, and the addition of new (non-TEI) markup options.“ (TEI Consortium 2022, <https://tei-c.org/release/doc/tei-p5-doc/en/html/AB.html#ABTEI2>)

XML-Validierung

Mit TEI kodierte Texte lassen sich über ein Editionskorpus durch Festlegung des Vokabulars aus dem Gesamtumfang der bestehenden TEI-Module (etwa für Handschriften, für Dramentexte, für Wörterbücher usw.) gleichförmig machen, indem durch ein XML-

Schema bestimmte Elemente in ihrer Kardinalität und Position festgelegt oder Attribute auf eine Auswahl bestimmter Werte begrenzt werden. Die TEI-Community stellt mit ROMA für diesen Zweck ein Tool zur Verfügung (<http://roma.tei-c.org/>). Ein Korpus von einigen tausend Briefen lässt sich mithilfe eines Schemas innerhalb kurzer Zeit „validieren“, um zu prüfen, ob etwa an einer Stelle ein Titel vergessen wurde oder eine Fußnote keinen Literaturnachweis beinhaltet. Solcherart mehr oder weniger „hart“ definierte Editionsrichtlinien müssen selbstverständlich in einem vorangegangenen Schritt der Datenmodellierung erarbeitet werden (Resmini & Rosati 2011; Klingner & Lühr 2019), sie erleichtern dann aber die Arbeit mit den Quellen.

In der täglichen Arbeit mit digitalen Editionen in TEI-XML lässt sich grob unterscheiden zwischen der Aufnahme von 1. Metadaten, die die Quelle und Edition beschreiben, 2. Strukturdaten, die die Textgliederung etwa in Absätze und Abschnitte wiedergeben, sowie 3. semantischem Markup, das einzelne Zeichenketten in ihrer kommunikativen Funktion kennzeichnet. An dem obigen kurzen Zitat aus dem historischen Briefindex exemplifiziert hieße das:

TEI-XML:
Struktur, Semantik

1. Im `<teiHeader>` sind zumindest die Archivsignatur, Urheberchaften, Datierung und Titeldaten der Quelle und der Edition sowie ein Verweis auf die Editionsrichtlinien des Projekts verankert.
2. Das Element `<facsimile>` beinhaltet Verweise auf digitale Reproduktionen der Quelle und unter Umständen Koordinaten bestimmter Bereiche als `<zone>`.
3. Der gesamte zitierte Eintrag ist ein `<item>` in einer `<list>` von Briefen, er hat ein `<label>` mit einer Nummerierung (`@n`), deren Repräsentation verweist auf den an anderer Stelle edierten Brief und
4. wird semantisch erschlossen durch `<rs>`-Tags um Personen- und Ortsnamen (vgl. Infobox 3 – Named Entities) sowie die normalisierte Datierung (`<date>`).

Ein wichtiges Prinzip von digitalen Editionen in diesem engeren Sinn ist die Trennung von Semantik und Darstellung: Während auf der Darstellungsebene z. B. eines Buches verschiedene textuelle Phänomene *kursiv* gedruckt erscheinen, lassen diese sich doch funktional und semantisch voneinander abtrennen: Der lebende Kolumnentitel

Semantik vs. Darstellung

einer Editionsreihe ist dann als solcher definiert (`<fw type="header" rend="#italic">Kapitelüberschrift</fw><fw type="page Num" rend="#italic">Seitenzahl</fw>`), die Emphase innerhalb des edierten Textes (`<emph>`) hat einen anderen Stellenwert als der ebenfalls kursiv gesetzte Buchtitel im Zitat (`<title><hi rend="#italic">Titel</hi></title>`). Dies ermöglicht einerseits eine zielgenaue Adressierung aller Elemente einer Klasse (etwa, um Kolumnentitel für die Analyse aus den Textdaten wegzulassen), andererseits auch die Änderung der Darstellungsweise für eine Elementklasse (sollte z. B. eine Änderung des Zielformats für Titelangaben in den Fußnoten erwünscht sein, lassen sich alle `//note//title/hi[@rend='#italic']` auch in grün pulsierenden Kapitälchen wiedergeben.

Die Trennung von Inhalt und Darstellungsweise ist erfahrungsgemäß eine konzeptionelle Hürde. Vergewagt man allerdings, dass die unterschiedliche Darstellung gleicher wie die gleiche Darstellung unterschiedlicher Sachverhalte letztlich auch nur vorher vereinbarten Konventionen folgt, so liegt nahe, mit dem Vokabular etwa der TEI-Guidelines zunächst einmal den semantischen Gehalt aufzuzeichnen (ggf. zusätzlich dessen ursprüngliche Darstellungsweise in der Quelle über ein `@rend-` oder `@rendition-`Attribut).

Erschließungstiefe

Ein oft diskutiertes Problemfeld der Edition ist jenes der Frage nach der zu erzielenden Erschließungstiefe. Sie richtet sich nach den Ansprüchen des Faches, ist aber auch Aushandlungssache zwischen dem technisch Machbaren und dem finanziell und zeitlich Umsetzbaren. Viele digitale Editionen der letzten Zeit haben z. B. die Auszeichnung von Metadaten, Schlagwörtern, Strukturdaten und Textsemantik inklusive der *Named Entities* (siehe u. Infobox 3) an einzelnen Text- (oder Objekt-)zeugen in den Vordergrund gestellt und dabei den angestammten wichtigen wissenschaftlichen Beitrag der Kommentierung auf Grundlage der Textkritik hintangestellt. Das ergibt schon ein kursorischer Durchgang durch auch nur die aus den Philologien stammenden digitalen Editionen in den gängigen Katalogen (siehe Abschnitt Kataloge digitaler Editionen im Beitrag „Daten und Schnittstellen. Die Ausgabeseite digitaler Editionen“ in diesem Band). Durch die Analyse solcher relativ oberflächlichen Strukturdaten lassen sich vor allem Fragen des *Distant Reading* (Moretti 2013) systematisch bearbeiten: Zu den Möglichkeiten (v. a. literaturwissenschaftlicher)

Analyse vgl. Schöch (2016); Korpora werden teilweise ebenfalls international vergleichbar mit TEI analysierbar gemacht (COST-Action „Distant Reading“, <https://www.distant-reading.net/>).

Infobox 3 – Named Entities

Allgemeine Bezeichnung für benannte/benennbare Einheiten, die aus Texten heraus in der digitalen Edition referenziert werden. Üblicherweise kommen Tags zur Anwendung für

- Personen – in TEI als `<person>` mit dem Bezugselement `<persName>` oder `<name type="person">` für Namensnennungen, `<rs>` (referring string) für indirekte Bezüge („ihre Schwester“)
- Orte – in TEI als `<place>` mit dem Bezugselement `<placeName>` oder `<rs type="place">` und
- Institutionen/Organisationen – in TEI als `<org>` mit dem Bezugselement `<orgName>` oder `<rs type="org">`

Weitere Entitätentypen etwa für Objekte `<object>` sind im TEI-Schema angelegt. Ähnlich verhält es sich mit bibliographischen Einheiten `<bibl>` und `<biblStruct>`, mit Nymen als verallgemeinerte Namensstandardisierung `<nym>` oder mit Taxonomien `<taxonomy>` und `<category>`. Ausbaufähig ist die in TEI vordefinierte Verwendung von Ereignisdaten als `<event>`. Diese Liste ist jedoch – wie das TEI-Schema insgesamt – durch die Benutzer:innen erweiterbar.

In der digitalen Edition werden *Named Entities* üblicherweise in externen Listen bearbeitet, auf die über Identifikatoren in den edierten Dokumenten verwiesen wird. Zentrales Feature dabei ist das Kenntlichmachen (Identifikation) und die Vereindeutigung (Disambiguierung) von Namen (Bezeichnen) über Identifikatoren aus sogenannten Normdateien (engl. *authority files*), also Konkordanzen von Bezeichnern und Bezeichnetem. Im deutschsprachigen Raum wird meist der Identifikator der aus dem Bibliotheksbereich stammenden *Gemeinsamen Normdatei* (GND-Id) für Personen verwendet (die GND listet beispielsweise 25 distinkte Personen, auf die das Suchmuster „Mayer, Maria“ passt). Die GND umfasst auch Daten zu Institutionen und Organisationen. Für Ortsbezeichnungen bietet sich neben den GND-Daten die *GeoNames*-Id an, welche multilinguale Toponyme mit entsprechenden geographischen Koordinaten verknüpft (darunter drei in Deutschland und zwei in den USA gelegene Orte namens Salzburg).*

Named Entity Recognition (siehe die Beiträge von Melanie Althage und Charlotte Schubert in diesem Band) ist der Prozess regel- oder modellgetriebener automatischer Erkennung von Eigennamen und Bezeichnern im Zuge der automatisierten Textverarbeitung.

* GND: Gemeinsame Normdatei, recherchierbar über <https://portal.dnb.de/>; GeoNames ist eine offene Datenbank von Ortsbezeichnungen und zugehörigen Geodaten, siehe <https://www.geonames.org/>.

Standards sunt servanda

Im Gesamtgefüge der mit „digitalen“ Mitteln betriebenen Geschichtswissenschaft kommt der wissenschaftlichen digitalen Edition ein zentraler Stellenwert zu. Sie ermöglicht einen standardisierten Zugriff auf edierte Quellen, der nicht nur auf die engen Grenzen des einzelnen Fachs beschränkt ist. Mit der an Standards wie dem TEI-Vokabular orientierten Wiedergabe der Edenda lässt sich neben dem „Inhalt“ einer Quelle auch deren Form in menschen- wie maschinenlesbarer Form abbilden.

Grundsätze der Textwiedergabe wie individuelle Verantwortlichkeiten für Lesarten, Kommentare, Änderungen werden in der digitalen Edition mit dokumentiert.

Erstes Ergebnis der Tätigkeit wissenschaftlicher Editor:innen sind damit zunächst standardisierte (XML-)Dateien. Diese beinhalten die nach den Editionsgrundsätzen aufgenommenen Informationen zu Struktur und Semantik der Quelle/n und alle erforderlichen Metadaten dazu. Der Wiedergabe dieser Informationen in verschiedenen Ausgabeformaten (Webapplikationen, aber auch Druckerzeugnisse) ist im Abschnitt „Publizieren – Präsentieren“ ein eigener Beitrag gewidmet.

3 Literaturhinweise

- Bohnenkamp-Renken, A. (Hrsg.). (2012). *Medienwandel/Medienwechsel in der Editionswissenschaft*. De Gruyter. <https://doi.org/10.1515/9783110300437>
- Cummings, J. (2018). A World of Difference: Myths and Misconceptions about the TEI [Supplement]. *Digital Scholarship in the Humanities*, 34, 158–179. <https://doi.org/10.1093/llc/fqy071>
- Flanders, J., & Jannidis, F. (2018). *Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources*. Routledge.
- Giesecke, M. (1998). *Der Buchdruck in der frühen Neuzeit: Eine historische Fallstudie über die Durchsetzung neuer Informations- und Kommunikationstechnologien*. Suhrkamp.
- Jannidis, F., Kohle, H., & Rehbein, M. (Hrsg.). (2017). *Digital Humanities: Eine Einführung*. Metzler. <https://doi.org/10.1007/978-3-476-05446-3>
- Krajewski, M. (2002). *Zettelwirtschaft: die Geburt der Kartei aus dem Geiste der Bibliothek*. Kulturverlag Kadmos.
- Plachta, B. (2019). *Editionswissenschaft*. Reclam.

4 Links

- Dumont, S., Haaf, S., & Seifert, S. (Hrsg.). (2019–2020). *Encoding Correspondence: A Manual for Encoding Letters and Postcards in TEI-XML and DTABf*. <https://encoding-correspondence.bbaw.de/> (30. 7. 2023)
- Institut für Dokumentologie und Editorik, & DHd2022. (2022, 11. März). Manifest für digitale Editionen. *DHdBlog*. <https://dhd-blog.org/?p=17563> (30. 7. 2023)
- Kamzelak, R. S. (Hrsg.). (2016). *Edlex*. <https://edlex.de/> (30. 7. 2023)
- TEI Consortium. (2022). *TEI Guidelines*. <https://tei-c.org/guidelines/> (30. 7. 2023)
- W3C. (Hrsg.). (2009, 28. Jänner). *XML Base (Second Edition): W3C Recommendations 28 Jan. 2009*. <https://www.w3.org/TR/2009/REC-xmlbase-20090128/> (30. 7. 2023)

