

Digitale Editionen vom «kleinen Mann»

Christa Ackermann

Einleitung

Einordnung und Fragestellung

Frei zugängliche Forschungsergebnisse sind aktuell in aller Munde: Am 7. März 2018 stellte swissuniversities in einer Medienmitteilung ihren Aktionsplan zur Umsetzung der 2016 verabschiedeten nationalen Open-Access-Strategie vor: Sie erklärte als «gemeinsame Vision der Schweizer Hochschulen», bis 2024 freien Zugang zu allen durch öffentliche Mittel finanzierten Publikationen anzustreben. Der Schweizerische Nationalfonds zur Förderung der wissenschaftlichen Forschung (SNF) verlangte bereits im Oktober 2017, dass zukünftig die von ihm geförderten Forschungsprojekte nicht nur ihre Publikationen, sondern auch ihre Forschungsdaten in einer frei zugänglichen Datenbank veröffentlichen müssen. Auch der Bundesrat beschloss am 9. Mai 2018, dass das Departement für Wirtschaft, Bildung und Forschung «in den Jahren 2021 – 2024 in Zusammenarbeit mit den Forschungs- und Hochschulakteuren geeignete Massnahmen zur Zugänglichmachung von Forschungsdaten prüfen» soll.¹

Schon diese kurze Zusammenfassung zeigt einen klaren Trend, der momentan nicht nur in der Schweiz, sondern besonders durch entsprechende Vorgaben bei *Horizon 2020* europaweit verfolgt wird:² Mit öffentlichen Geldern finanzierte Forschungsergebnisse, seien es abgeschlossene Publikationen oder die ihnen zugrundeliegenden Forschungsdaten, sollen weiteren Forscherinnen,³ aber auch der Öffentlichkeit kostenlos frei zugänglich gemacht werden. Dadurch wird einem ursprünglichen Bedürfnis der Wissenschaft Rechnung getragen: Die besten Köpfe sollen Zugang zu den Ergebnissen ihrer Vordenker haben, um ihr maximales Potential auszuschöpfen und die Wissenschaft bzw. die Menschheit damit voranzutreiben.

Diese Aussage mag plakativ und ideologisch klingen, entspricht aber im Grundsatz dem Narrativ, das in der Open-Access- bzw. Open-Science-Bewegung

1 Die entsprechenden Mitteilungen finden sich auf den folgenden Websites: <https://www.swissuniversities.ch/de/publikationen/medienmitteilungen/> [29.06.2019], http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/default.aspx [29.06.2019], <https://www.bakom.admin.ch/bakom/de/home/das-bakom/medieninformationen/medienmitteilungen.msg-id-70694.html> [29.06.2019].

2 Die neusten Nachrichten dazu auf: <http://www.donneesdelarecherche.fr> [29.06.2019]. Eine historische Zusammenfassung für Deutschland: Winkler-Nees 2012.

3 Aus Gründen der besseren Lesbarkeit wird im Folgenden bei personenbezogenen Bezeichnungen das generische Femininum verwendet. Es sind stets alle Geschlechter gemeint.

immer wieder als treibendes Motiv erklärt wird.⁴ Der SNF betrachtet seine Richtlinie «als einen wesentlichen Beitrag zur Wirkung, Transparenz und Reproduzierbarkeit wissenschaftlicher Forschung.»⁵ In der knapp gehaltenen Medienmitteilung für die Open-Access-Strategie wird erklärt: «Open Access reiht sich in die grössere Strömung von Open Science ein und hat zum Prinzip, dass keine finanzielle, technische oder rechtliche Barriere den Zugang zur wissenschaftlichen Literatur verhindern sollte. Open Access trägt dazu bei, die Sichtbarkeit und die Verbreitung der Forschung zu erhöhen, wovon die Forschenden, die wissenschaftliche Gemeinschaft und die breite Öffentlichkeit profitieren.»⁶

Um dies zu erreichen, lehnen sich der SNF und auch die europäische Struktur für Forschungsunterstützung, *Horizon 2020*, in ihren Forderungen an die good practice der FAIR Guiding Principles an, wie sie von Mark D. Wilkinson et al. formuliert wurden: FAIR steht für findable, accessible, interoperable, reusable. Mit diesen Grundsätzen soll sichergestellt werden, dass die Forschungsdaten tatsächlich nutzbar sind, indem sie langfristig auffindbar und zugänglich sowie in einem Format mit den nötigen Metadaten versehen und unter einer Lizenz abgespeichert sind, die es fremden Forscherinnen ermöglicht, die Daten zu verstehen und weiter zu bearbeiten.⁷ Nur wenn Forschungsdaten in einer lesbaren und interpretierbaren Form zugänglich sind, können die Ergebnisse verifiziert oder Daten nachgenutzt werden.

Wenn diese Forderungen von der Wissenschaft ernst genommen werden, reicht es nicht, vorhandene Daten ohne Bearbeitung in einem sogenannten Repository⁸ abzulegen. Es wird sich von Disziplin zu Disziplin eine good practice herausbilden müssen, in welcher Form die Forschungsdaten abgelegt werden, um ihre Nutzbarkeit zu garantieren.⁹ Dank des Entscheids der grossen Geldgeber in der Forschung wird also in verschiedensten Fachgebieten die Diskussion intensiviert werden, wie dies geschehen soll. Auch die vorliegende Arbeit ist Teil dieses Diskurses mit Fokus auf eine spezifische Art von Forschungsdaten: Das Augenmerk liegt auf der bearbeiteten (handschriftlichen) historischen Quelle – meist Archivalien oder sonstige historische Texte –, wie sie als eine Grundlage für textbasierte historisch orientierte Forschung dient.

Wenn z. B. in der Geschichtswissenschaft mit (handschriftlichen) Quellen gearbeitet wird, werden diese meistens in einem Arbeitsablauf für die Forschung

4 Vgl. Drott 2006.

5 http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/default.aspx#Grundsatzzerkl%E4rung%20des%20SNF%20zu%20Open%20Research%20Data [29.06.2019].

6 <https://www.swissuniversities.ch/de/publikationen/medienmitteilungen/> [29.06.2019].

7 Wilkinson et al. 2016.

8 Als Repository wird bei Open-Access-Publikationen ein Server bezeichnet, auf dem langfristig Publikationen aufbewahrt und zugänglich gemacht werden.

9 Vgl. Oßwald/Scheffel/Neuroth 2012.

nützlich gemacht: Um nicht örtlich ans Archiv gebunden zu sein, werden die Quellen digitalisiert; um die Texte besser zu verstehen, werden Transkriptionen angefertigt; verschiedene Textzeugnisse des gleichen Texts werden verglichen; um bestimmte Merkmale schneller aufzufinden, werden Transkriptionen annotiert. Zumindest Teile dieser Bearbeitungsschritte werden heutzutage meistens zusammen mit den Forschungsergebnissen publiziert in Form von Quellenverzeichnissen, Transkriptionen oder Quelleneditionen.

Der Buchdruck ist jedoch nicht der einzige Weg, wie Forschungsdaten anderen Wissenschaftlerinnen zur Verfügung gestellt werden; eine digitale Open-Access-Lösung für die Veröffentlichung, die auch die Nutzbarkeit der Daten garantiert, ist die Onlineedition. Diese bietet gegenüber der gedruckten Version den Vorteil, dass mehr Arbeitsschritte öffentlich und dadurch nutzbar gemacht werden können. Das heisst aber nicht, dass alle mit historischen Quellen arbeitenden Forscherinnen ihre Pflicht gegenüber dem SNF problemlos mit einer Onlineedition nachkommen können. Die Haupthürde liegt immer noch darin, dass eine gedruckte Edition für die meisten Wissenschaftlerinnen einfacher realisierbar ist als eine Onlineedition.¹⁰

Einerseits sind Onlineeditionen klar im Vormarsch, da sie viele Vorteile bieten. Ein wichtiger Punkt ist, dass der Umfang nicht mehr durch die physischen Grenzen eines Buchs bestimmt ist: das Layout ist nicht durch die Möglichkeiten einer Buchseite begrenzt, Textvarianten von verschiedenen Überlieferungen können alle vollständig abrufbar sein und müssen nicht in einem komplexen System im Variantenapparat abgebildet werden, Hilfsmittel und Kommentare können nach Bedarf ein- und ausgeblendet werden, Transkription und Faksimile können nebeneinander oder übereinander betrachtet werden. Gleichzeitig bietet das Internet einen neuen Zugang zur Edition: die Ressourcen sind dynamisch und können jederzeit dem aktuellen Stand der Erkenntnis angepasst werden, dank der zeit- und ortsunabhängigen Zugänglichkeit können sie ein grösseres und breiteres Publikum erreichen, mit neuen Technologien können die Daten mit anderen Projekten vernetzt werden, Daten können elektronisch weiterverarbeitet werden oder Benutzerinnen können selber Inhalte hinzusteuern. Eine Onlineedition wird nicht nur gelesen, sondern benutzt.¹¹ Jörg Hörschemeyer geht in seiner Dissertation zur digitalen Edition so weit zu sagen: «Heutzutage ist es wohl unbestritten, dass die Vorteile [...] im digitalen Medium liegen.»¹²

Andererseits haben Onlineeditionen jedoch auch klare Nachteile: bereits für die Erstellung ist relativ viel technisches Knowhow nötig, für die nachhaltige Benutzung muss eine Serverstruktur über die gesamte Lebensdauer der Edition gepflegt

10 Daran hat sich seit Robinson 2005 nicht viel geändert.

11 Vgl. z. B. Pichler und Bruvik 2017, S. 207-210, van Zundert 2014, S. 335-336, Hörschemeyer 2013, bes. S. 30-35 und Sahle 2013, bes. Bd. 2, S. 52-58. Zur oft zitierten Differenzierung zwischen Lesen und Benutzen von Onlineeditionen vgl. Pierazzo 2015, Kapitel 7.

12 Hörschemeyer 2013, S. 29.

werden, was auch langfristige finanzielle Mittel voraussetzt, die Darstellung der Inhalte muss periodisch den neusten technologischen Gepflogenheiten angepasst werden, um ein frisches Erscheinungsbild, aber auch den technischen Zugang zu den Daten zu gewährleisten, Verlinkungen müssen gepflegt werden, da sich auch die anderen Websites verändern können. Weil einzelne Forscher selten die nötigen Ressourcen haben, werden Onlineeditionen hauptsächlich von an Hochschulen angebotenen Forschungsgruppen erstellt, die durch interdisziplinäre Zusammenarbeit den Geisteswissenschaftlerinnen das technische Knowhow bieten und eine längerfristige Finanzierung garantieren können. Trotzdem sind bereits viele Onlineeditionen wieder aus dem Internet verschwunden, weil die langfristige Investition nicht garantiert war.¹³ Als einfacher Weg zur Veröffentlichung von Forschungsdaten taugt eine Onlineedition im Moment also nicht.

Ziel dieser Arbeit ist, einen Beitrag zu leisten, dass in Zukunft auch Wissenschaftlerinnen ohne grosse Ressourcen die Möglichkeit haben, ihre während der Forschung erstellten editorischen Produkte als Forschungsdaten in Form einer Onlineedition zu veröffentlichen. Damit dies möglich ist, müssen die technischen Probleme so weit gelöst werden, dass sich die Forscherin nur noch um den editorischen Teil kümmern muss und die digitale Publikation möglichst selbsterklärend nebenbei erstellt werden kann.¹⁴ Dafür braucht es eine Plattform, auf der eine Onlineedition aus technischer Sicht so einfach erstellt werden kann, wie ein Blog mit WordPress oder ein Enzyklopädieartikel mit Wikipedia: Die Onlineedition muss im Web 2.0 ankommen! Wenn ein solches Produkt frei zugänglich wäre, würde es viel mehr bieten als nur die Möglichkeit, der vom SNF formulierten Veröffentlichungspflicht nachzukommen: Es wäre ein noch viel grundlegenderer Beitrag zur Demokratisierung der Wissenschaft, da theoretisch jede Forscherin die Möglichkeit hätte, ihr Wissen in Form einer Onlineedition frei zugänglich zu machen, unabhängig von einer Hochschulanbindung.¹⁵

Meine Hauptanforderungen an eine solche Editionsplattform sind klar und im Grunde auch einfach: Die Veröffentlichung einer Onlineedition soll in einem finanziellen und technischen Rahmen stattfinden können, der nicht grösser ist als bei einer gedruckten Edition. Dabei darf das Produkt nicht schlechter sein als sein analoges Pendant: Es müssen mindestens diejenigen Auswertungsmöglichkeiten angeboten werden, die eine gedruckte Edition bietet. Wie bei einer gedruckten Edition muss die persistente und langfristige Verfügbarkeit garantiert sein. Ausserdem soll die Plattform soweit der Demokratisierung der Wissenschaft entsprechen, dass auch Forscherinnen ohne Hochschulanschluss die Möglichkeit haben, Projekte zu veröffentlichen.

13 Vgl. Sahle 2013, Bd. 2, S. 78-81.

14 Vgl. Sahle 2013, Bd. 2, S. 111.

15 Ähnlich argumentiert auch Robinson 2016a. Vgl. OCDE 2004, bes. S. 63-68 zu den sozialen Vorteilen von User Created Content. Kritisch: Pierazzo 2015, S. 129-130.

Während es bereits einige Bestrebungen in diese Richtung gibt und es der Forschungsgemeinschaft grundsätzlich klar ist, dass eine Editionsplattform einer gedruckten Edition gleichwertig sein sollte, ist ein uneingeschränkter Zugang alles andere als selbstverständlich: Manche Hochschulen bieten Lösungen zum Erstellen von online zugänglichen Editionen an, jedoch nur für ihre Angehörigen. Andere Editionsprojekte binden zwar Laien bzw. nicht institutionell gebundene (Fach-)Personen im Rahmen von Crowdsourcing ein, z. B. um Transkriptionen anzufertigen.¹⁶ Diese bekommen dadurch aber nur als Konsumentin bzw. als billige Arbeitskraft Zugang zu den Editionsprojekten. Entscheidungen über den Inhalt und Umfang trifft immer noch ein enger Kreis von Spezialistinnen.¹⁷

Bevor sich eine digital wenig affine Forscherin an die Möglichkeit einer Onlinedition heranwagt, muss trotz vielversprechenden Ansätzen noch viel Arbeit geleistet werden. Diese zu definieren ist Ziel der vorliegenden Arbeit: Wie können bei der Bearbeitung von historischen Quellen entstandene Forschungsdaten langfristig und nachnutzbar online zur Verfügung gestellt werden? Bruchstückhaft erstellte Editionsarbeiten sollen als Forschungsdaten genauso berücksichtigt werden wie wissenschaftlich-kritische Editionen: Mit dem Entscheid des SNF wird das Thema dringend, wie diese Daten langfristig und nutzbar zur Verfügung gestellt werden können.

In diesem einführenden Teil werden gleich anschliessend zuerst einige Begriffe geklärt und danach mein eigenes Editionsprojekt vorgestellt, das für den Rest der Arbeit als roter Faden dienen soll. In einem nächsten Schritt werden diverse Standards besprochen, auch dies eine Grundlage, die für die nachstehenden Erläuterungen und Überlegungen zentral ist. Im dritten Teil wird eine Auswahl an bereits bestehender Editionssoftware vorgestellt. Der Blickwinkel auf diese ist derjenige der Bedürfnisse einer Klein- oder Teiledition. Es geht darum zu verstehen, was für Lösungen bereits vorhanden sind, wie sie entstanden sind, wo ihre Limitationen liegen und inwiefern sie einen Ausgangspunkt für eine Weiterentwicklung bieten. Abschliessend steht die theoretische Auseinandersetzung mit den Rahmenbedingungen einer Editionsplattform im Vordergrund: Es werden die Hürden aufgelistet, die eine Editionsplattform für Kleinprojekte beseitigen muss und die Minimalanforderungen, die aus wissenschaftlicher Sicht zwingend umgesetzt werden müssen. Abschliessend folgt eine Auswahl an zusätzlichen Anforderungen, die in erster Linie aus den Bedürfnissen meines Projekts entspringen, aber zum grössten Teil bereits in der Forschungsliteratur formuliert wurden.

16 Robinson 2016b. Ein Beispiel einer solchen Plattform ist Velehanden von Picturae: <https://velehanden.nl> [29.06.2019].

17 Vgl. Dickel/Franzen 2015, S. 338.

Grundbegriffe

Digitale Edition und ihre Unterkategorie Onlineedition ist zwar ein relativ neues Forschungsfeld, jedoch in der Forschungslandschaft bereits deutlich angekommen: Es gibt nicht nur Zeitschriften wie das *Journal of the Text Encoding Initiative*, das sich immer wieder ausführlich mit dem Thema beschäftigt, sondern auch Sammelbände wie der 2017 erschienene französisch-norwegische Band *L'édition critique à l'ère du numérique* oder der 2014 erschienene *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*. Auch als Monographien wurde das Thema bearbeitet, z. B. von Elena Pierazzo in ihrem 2015 erschienenen und mehrfach nachgedruckten Band *Digital Scholarly Editing* oder im 2013 erschienenen dreibändigen Werk von Patrick Sahle *Digitale Editionsformen*. Für einige Forscherinnen und Forscher wie Tara Andrews oder Peter Robinson ist die digitale Edition ein Hauptthema ihrer wissenschaftlichen Publikationsliste.¹⁸ Der Diskurs ist sehr international aufgeleitet und findet hauptsächlich auf Englisch statt. Wie die aufgeführten Werke zeigen, gibt es aber durchaus auch wichtige Beiträge in anderen Sprachen, die über die Sprachgrenzen hinweg rezipiert werden.

Mit der verbreiteten Diskussion über die digitale bzw. die online verfügbar gemachte digitale Edition sind auch Begrifflichkeiten immer wieder ein Thema. Dies fängt bei der grundlegendsten Frage an, was eine Onlineedition überhaupt ausmacht. Diese Diskussion beruht hauptsächlich darauf, dass das Label für eine grosse Bandbreite benutzt wird, von der retrodigitalisierten Printedition bis zur technisch ausgeklügelten wissenschaftlich-kritischen Edition. Im Gegensatz zur gängigsten Definition in der Forschungsliteratur betrachte ich jegliche Art von Edition, die online gestellt wird, als Onlineedition.¹⁹ Der Grund dafür ist, dass in meiner Arbeit die technischen Bedingungen im Zentrum stehen, die unabhängig vom wissenschaftlichen Wert einer Edition bestehen. Die Anforderungen an eine Editionsplattform werden zwar für ein eher komplexes Projekt mit wissenschaftlichem Anspruch gestellt, die technischen Grundlagen können aber genauso für einfachere Projekte eingesetzt werden.

Als Kleineditionsprojekt bezeichne ich hier Projekte für Onlineeditionen, die nicht institutionell getragen werden. Das können neben Qualifikationsarbeiten oder Arbeiten von nicht institutionell gebundenen Wissenschaftlerinnen auch Onlineeditionen von kleineren Institutionen (Archiven, Bibliotheken, Hochschulen) sein, die die nötigen technischen und finanziellen Mitteln nicht selbstständig aufbringen können. Mit dem Begriff Teileditionsprojekt wird der Tatsache Rechnung getragen, dass beim Bearbeiten von historischen Quellen nicht unbedingt eine vollständige

18 <http://artsandscience.usask.ca/profile/PRobinson#/publications> [29.06.2019] und <http://univie.academia.edu/TaraAndrews> [29.06.2019].

19 Enger definiert prominent bei Apollon/Régnier/Bélisle (Hgg.) 2017, S. 395 (édition électronique) sowie Sahle 2013, Bd. 2, S. 27 (digitale Edition).

wissenschaftlich-kritische Edition das Ziel ist.²⁰ Der Einfachheit halber spreche ich generisch von Forscherin, wenn die Erstellerin einer Klein- oder Teiledition gemeint ist. Diese muss jedoch nicht zwingend an einer wissenschaftlichen Institution angegliedert oder weiblich sein.

Ein weiteres Begriffsfeld, das zuerst geklärt werden muss, ist das des Texts. Hier liegt die Schwierigkeit darin, dass Text und Dokument Begriffe sind, die in vielen unterschiedlichen Kontexten verbreitet sind und in jeder wissenschaftlichen Disziplin leicht anders benutzt werden. Auch hier bin ich in der Definition pragmatisch: Während Sahle sechs verschiedene Aspekte eines Texts ausmacht und diese als Text_s, Text_w, Text_f, Text_d, Text_i und Text_z bezeichnet,²¹ versuche ich mich auf eine differenzierte Terminologie zu stützen, die in diversen Wissenschaften gebräuchlich ist. Im Vordergrund stehen dabei die Unterscheidungen mit einem technischen Bezug.

Die erste Unterscheidung kommt aus der Geschäftsverwaltung und differenziert zwischen Dokument und Datei (File): «Datei» beschreibt ein elektronisches File, wie es in einem File-Ablage-System (z. B. Windows Explorer) zu finden ist. Die Datei ist in einem spezifischen Dateiformat gespeichert und kann Metadaten über sich selbst enthalten, muss aber nicht. Der Dokumentbegriff bezeichnet in der Geschäftsverwaltung hingegen die Metadaten zu einer elektronischen Datei oder einem physischen Dokument. Sie werden normalerweise in einem Metadatenstandard abgelegt.²²

Die zweite Unterscheidung ist in den textbasierten Wissenschaften zu finden und bezieht sich auf Dokument und Text: Während das Dokument das physische Blatt Papier, Buch oder einen sonstigen Textträger bezeichnet, ist der Text eine abstrakte Entität. Ein Text braucht immer einen Träger, sei es ein physischer oder elektronischer.²³ Unter Umständen gibt es mehrere Dokumente, die den gleichen (oder einen ähnlichen) Text enthalten. Um die übergeordnete Entität zu bezeichnen, benutze ich den Begriff Werk.²⁴

Bereits gefallen ist der Begriff der Metadaten, der auch einer Erklärung bedarf: In den Textwissenschaften werden Daten als Metadaten bezeichnet, die ein Dokument oder ein Werk beschreiben. Im Kontext der Onlineedition sind zwei Arten von Metadaten wichtig: Erstens solche, die das Dokument beschreiben. Neben bibliographischen Angaben können das bei physischen Dokumenten Angaben zur Beschaffenheit sein, z. B. Beschreibstoff, Grösse (z. B. in mm) und Umfang, bei einem elektronischen

20 Vgl. Stadler 2013, S. 32-33.

21 Sahle 2013, Bd. 3, S. 9-45. Die Abkürzungen stehen für «Text als sprachliche Äusserung», «Text als Werk», «Text als Fassung», «Text als Dokument», «Text als Inhalt/Idee/Intention» und «Text als Zeichen».

22 So z. B. in den GEVER-Anwendungen Fabasoft und CMI Axioma. Peter M. Toebak unterscheidet lediglich zwischen Primärdaten und Metadaten, dahinter steht aber im Grunde das gleiche Prinzip: Toebak 2007, S. 401-415.

23 Wie Huitfeldt 2017, S. 183.

24 Wie Grewe Rasmussen 2009, S. 121-122. Vgl. auch Pierazzo 2015, S. 40-53, die sich ausführlich mit den Begriffen «text», «document» und «work» befasst.

Dokument sind damit Angaben zu Dateiformat, Grösse (z. B. in MB) und Dateiname gemeint. Zweitens sind Metadaten im Fokus, die unabhängig vom Medium den Inhalt eines Texts beschreiben. Dazu gehören Personennamen, Stichwörter oder auch kurze Zusammenfassungen. Abweichungen im Text zweier Dokumenten, die im Druck als Variantenapparat dargestellt werden, können auch als Metadaten betrachtet werden.²⁵

In der Welt der Programmierinnen beinhalten Metadaten all dies, aber auch noch Zusätzliches. Im Besonderen wird oft auch eine Transkription genau gleich wie andere Formen von Metadaten behandelt. Das kommt daher, dass das Digitalisat – die digitalisierte Bilddatei der Originalquelle – bzw. das Original in den Mittelpunkt gestellt wird und die Transkription als dazugehörige elektronische Textfassung behandelt wird. So werden Transkription und Metadaten zusammen als Beschreibung des Bildes oder des nicht-digitalen Originals abgelegt. Da es sich bei beiden beschreibenden Aspekten um Textzeichen handelt, ist die Differenzierung für die technischen Umsetzung zweitrangig.²⁶

In Hinblick auf eine Onlineedition bedeuten die Definitionen folgendes: Das Dokument beinhaltet alles, was zu einem Originalstück gehört, unabhängig von der ursprünglichen Beschaffung des Originals. Das heisst, zu einem Dokument gehört das dazugehörige Digitalisat, der im Original enthaltene Text sowie Metadaten. Das Digitalisat besteht aus einer oder mehreren Bild-Dateien, je nach Umfang des Originals. Der Text – meistens Transkriptionen – bestehen ebenfalls aus einer oder mehreren Textdateien. Mehrere Dokumente können zu einem Werk gehören. Wenn dem Werk oder dem Dokument Metadaten zugeordnet werden, werden diese nach unten vererbt.

25 «Métadonnées de structures» und «Métadonnées descriptives du contenu». Apollon/Régnier/Béliste (Hgg) 2017, S. 364.

26 So z. B. beim «Presentation-API» von IIF (<https://iiif.io/api/presentation/2.1/#annotation-list> [29.06.2019]) und dem DFG-Viewer (<http://dfg-viewer.de/profil-der-metadaten/> [29.06.2019]).

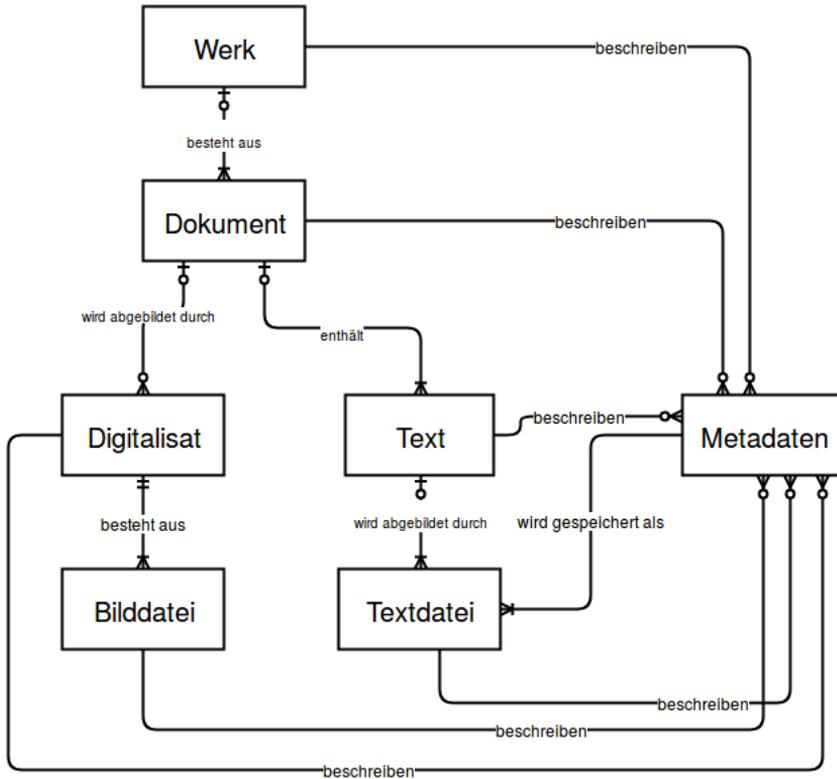


Abb. 1: Modellierung der Zusammenhänge zwischen den Grundbegriffen.

Ausgangslage - meine Edition

Als Grundlage für die Auswertung dient der ausführliche Editionsanhang meiner Doktorarbeit.²⁷ Dadurch soll sichergestellt werden, dass die Überlegungen auf einer ganz konkreten Anwendung basieren und nicht zu theoretisch werden. Im Zentrum der Edition steht ein Briefwechsel von zwei niederadligen Familien aus dem 15. Jahrhundert. Er war die Grundlage für die mikrohistorische Auswertung in meiner Doktorarbeit. Neben der intensiven Auseinandersetzungen mit diesen 60 Briefen zog ich auch eine Reihe von weiteren unedierte Quellen hinzu, die ich ebenfalls in der Edition in unterschiedlicher Tiefe beschrieb. Die einzelnen Teile sind folgende:

1. Eine Volledition von 60 Briefen.²⁸ Diese sind der Vorlage entsprechend möglichst getreu transkribiert und enthalten eine Fülle von diakritischen Zeichen. Zusätzlich zur Transkription gibt es pro Brief einen Titel, in dem Sender und Empfänger genannt werden, sowie eine Datumsangabe, ein Regest, das die wesentlichen Inhalte wiedergibt, und eine Beschreibung des physischen

²⁷ Ackermann 2014.

²⁸ Ackermann 2014, Anhang A.

Dokuments. Die eigentlichen Briefe bestehen aus drei verschiedenen graphisch eigenständigen Textteilen: die Adressierung, der Hauptteil und die Unterschrift. Da manche Briefe mehrfach überliefert sind, gibt es in der Edition einen Variantenapparat. Für inhaltliche Erklärungen gibt es einen Sachapparat.

A.4 Bilgrin von Reischach an Hans Ulrich von Stoffeln

1470, August, 15.

Hans Ulrich hat von Graf Johann [von Tengen] (von Eengen) die mittlere Burg auf dem Hohenstoffeln als Lehen empfangen, obwohl [Herzog Sigmund] von Österreich und die [Landgrafschaft] Nellenburg das Recht zur Lehensvergabe haben. Bilgrin widerruft Kraft seines Amtes die Lehensvergabe.

Hs. vHA A 8 2: Konzept, eigenhändig; 23x12cm; kein Wasserzeichen; kein Siegel

Ed. bislang unediert

Adr. Dem edeln vnd vesten Hannß Vlrichen von Stoffeln, minem güten fründe

Min fruntlich dinste zuvor^J, lieber Hanns Ulrich. Jch wird bericht, wie du das mitel schlos Stoffeln von dem wolgebornen hern graff Johannsen von Tengen^{K,1} zu lechen empffangen habest vnd sigel^L dasselb schlos^M von minem gnedigen hern von Österich vnd dem hus Nellenburg² lechen, darvmb

¹Folgt gestrichen güttlich ^Jzu über der Zeile ^KEengen ^LÜber gestrichenem wie ^MÜber gestrichenem lehen

³Andreas Ranft und Sonja Zielke-Dünnebeil führen nur Burkhard von Homburg als Hauptmann im Jahre 1453 auf: Kruse/Paravicini, Ritterorden und Adelsgesellschaften, S. 213. Herbert Obenaus kennt Hans Ulrichs Hauptmannschaft, jedoch auch nur anhand des vorliegenden Briefs. Vgl. Obenaus, Recht und Verfassung, S. 18. Die Adelsgesellschaft war in zwei bis vier Teilgesellschaften gegliedert, wobei der Hegau eine der Landschaften ausmachte. Vgl. Kruse/Paravicini, Ritterorden und Adelsgesellschaften, S. 204-205.

¹Vgl. B.1 und Anhang C.

Abb. 2: Ausschnitt aus der Volledition. Ackermann 2014, S. 171.

2. Eine Regestensammlung von grösstenteils juristischen Texten.²⁹ Die ausführlichen Regesten sind in moderner Sprache geschrieben, enthalten aber ebenfalls einige Wörter, die inklusive diakritischer Zeichen der Vorlage getreu transkribiert wurden. Die juristischen Texte sind im Original stark strukturiert und verschachtelt, was versucht wurde mit Auflistungen wiederzugeben. Auch diese Dokumente sind teilweise mehrfach überliefert. Auf einen Variantenapparat wurde in der Edition jedoch verzichtet und auf Überschneidungen stattdessen mit Verweisen oder farblicher Markierung hingewiesen. Wo nötig, wurden Eigenheiten der einzelnen Dokumente vor dem eigentlichen Inhalt erläutert. Ein Sachapparat, Titel, Datum und die physische Beschreibung der Dokumente ist vorhanden.

29 Ackermann 2014, Anhang B.

worteten, dass niemand die Aussage verweigern darf, wenn sie für den Rechtsgang von Nutzen ist.

- d) Die Weiterdinger: Sie sollen wenigstens von ihren Eiden befreit werden, damit sie wahrheitsgetreuer aussagen können.
- ii. Auf [Konrad Müllers] Frage, ob dies gestattet sei, erließ [Bilgrins] Vogt [Fridolin Belzer den Weiterdingern] die Eide für die Zeit ihrer Aussage und die Weiterdinger sagten unter Eid aus *wie dann z^u Wittertingen recht ist*.

- 1) Konrad Bucher: Als er Vogt war, kam [Bilgrin] zu ihm und erzählte, dass der Stoffler mit der Sägi (*segi*) gefischt habe, dass aber nach alter Gewohnheit nur mit Reusen, Bären (*ruschen, berren*) und Angeln gefischt werde und er [den Stoffler] deswegen vor den Hauptmann [der Gesellschaft mit St. Jörgenschild] zu Verantwortung gezogen (*furgenomen*) habe. Der Hauptmann habe geurteilt, dass keiner ohne den anderen [mit der Sägi] fischen (*ziehen*) dürfe. [Bucher] fragte darauf, was er tun soll, wenn [Bilgrin] außer Haus (*nit anhaimsch*) ist. Dieser antwortete, er solle sich so weit er könne wehren. Kurze Zeit später sah [Konrad Bucher] Heinrich und seine Leute beim See und er fragte [Heinrich], wieso sie fischten, obwohl [Heinrich] doch wisse, dass [Hans Ulrich von Stoffeln] rechtlich angewiesen worden sei, keiner dürfe ohne den anderen [mit der Sägi] fischen. Da antwortete ihm Heinrich, er wolle fischen *ob eß nimer recht wurd*. Darauf sagte [Bucher], er wolle schauen, ob [Bilgrin] sich das Seine nehmen lasse. [Bucher] suchte Rat bei Heinrich von Randegg (*Randegk*), der Hauptmann war, und der sorgte, dass der Stoffler <die Sägi nicht mehr benutzte. Darauf begegnete er [Hans Ulrich]> in Weiterdingen. Der sagte ihm, er habe unrecht getan und er solle in

²³Hier steht nochmals Bolli, was vermutlich ein Abschreibefehler ist, da der Name nicht mit dem Namen des Kundenschaftgebenden übereinstimmt.

Abb. 3: Ausschnitt aus der Regestenedition. Ackermann 2014, S. 263.

3. Eine Tabelle mit Exzerpten aus verschiedenen Grundbüchern.³⁰ In den fünf Spalten sind Name des Hintersassen (Pächter), Ortschaft des Grundstücks, Grundherr (Besitzer), Beschreibung des Grundstücks und Querverweise auf andere Quellen zum Hintersassen eingetragen. Wie bei den Regesten ist der Inhalt der Tabelle in moderner Sprache erfasst, enthält jedoch auch treu transkribierte Wörter (Originalschreibweise von Namen und Flurnamen sowie Wörter ohne modernes Äquivalent).

Basül (Bawlen; Bawli; Bawly; Bawle; Pantle; Pantin)				
Belzer (Beltzer)	Weiterdingen	Bilgrin		Anhang B.26
Bollinge	Weiterdingen	Helms		Anhang B.21
Häppli (Häpflin; Huppfl; Hapfle; Huppfl)	Weiterdingen	Bilgrin	Der kleine Zehut an Helmsaus von Stoffeln (B84 fol. 25r) Seine Kinder aus seiner ersten Ehe sind alle regn (B84 fol. 26r).	Anhang B.15 und B.20

Burkhard (Burgkart; Burghart; Burkli; Bärekin; Bäreki; Bürkli; Bärekin)				
Amman (Aman)	Weiterdingen	Bilgrin (Heinrich)	/ Von Rätts Gut 1/5 V Roggen, 1 V Hafer, 2 sh. 10 dn., 8 Eier. Von einer Schuppos 2 V Roggen, 1 Mu Hafer 9 sh. 5 dn., 2 Hühner, 1 Gans, 30 Fier. <Vom andern brüchel 2 lb. dn.>. Von der Wiese an bennen riet do das klein wgerlin lit 5 Mu Dinkel, 10 Hühner und 2 Fasnachtsühner. Von einer rütty lit un bettenlowe under benenriet 9 V Dinkel Zins (B83a). Von einer Schuppos 1/5 V Roggen 1 V Hafer 2 sh. 10 dn. 8 Eier. Von einer Schuppos	Anhang B.15, B.3 und B.26

Abb. 4: Ausschnitt aus dem Hintersassenverzeichnis. Ackermann 2014, S. 281.

Als zusätzliche Hilfsmittel bzw. zur Erschliessung dient ein Glossar, eine tabellenförmige chronologische Übersicht über die in den Quellen vorkommenden juristischen

Handlungen, ein Personen- und Ortsnamenverzeichnis, ein Handschriftenverzeichnis, ein Siegelverzeichnis und ein Wasserzeichenverzeichnis.³¹ All diese Hilfsmittel verweisen mit Quellenangaben auf die in den Punkten eins bis drei genannten edierten Quellen.

Die unterschiedliche Tiefe, in der die Quellen ausgewertet wurden, ist für die vorliegende Arbeit ein Vorteil, da sie sowohl als Ausgangslage für Überlegungen zu einer Kleinedition dienen kann, aber auch Aspekte einer Teiledition hat: Ein Teil der erhaltenen Daten sind Forschungsdaten, die hauptsächlich dem Zweck der Auswertung dienen und dementsprechend undetailliert sind. Da sie aber zur Nachvollziehbarkeit der Auswertung beitragen, schlecht erschlossene Quellen ausführlicher beschreiben und anderen Forscherinnen unter Umständen sogar den Gang ins Archiv ersparen können, ist klar, dass sie eine für die Forschung nützliche Grundlage bilden und veröffentlicht werden sollen.

Die Doktorarbeit wurde mit LaTeX geschrieben, einem Satzprogramm, das Text für den Druck bzw. als PDF-Output optimiert. Der Editionsteil ist klassisch gehalten und würde sich einfach als Buch drucken lassen. Da ich aber von den Vorteilen einer Onlineedition überzeugt bin und vergleichsweise viel Erfahrung im Web-Bereich mitbringe, möchte ich meiner Publikationspflicht nicht nur in Form eines Open-Access-PDF/A nachkommen, sondern eine ansprechendere Lösung finden, die eine bessere Weiterverarbeitung der Forschungsdaten ermöglicht. In der Hoffnung auf längerfristigen Erhalt und grössere Visibilität meiner Edition möchte ich mein Projekt auf einer bereits bestehenden Editionsplattform anbieten. Die vorliegende Arbeit ist also nicht nur als theoretische Auseinandersetzung mit den Möglichkeiten einer Onlineedition zu verstehen, sondern zeichnet auch die eigene Suche nach einer geeigneten Editionsplattform auf. Trotz Schwerpunkt auf meiner Edition wurde Wert darauf gelegt, anhand der Forschungsliteratur einen breiteren Fokus zu richten. Besonders auch der Abgleich mit den von Michael Bender bei Befragungen eruierten Wünschen von Editionserstellerinnen zeigt, dass viele aufgrund meiner Edition formulierten Bedürfnisse weit verbreitet sind.³²

Standards

Wer schon einmal eine Edition erstellt hat, weiss, dass jede Edition gewissen Standards folgt. Wegen der begrenzten Möglichkeiten beim Buchdruck hat sich in Hinsicht auf das Seitenlayout hauptsächlich die Darstellungsform des Apparats als good practice etabliert. Mithilfe eines Apparats können auf der zweidimensionalen Textseite verschiedene Dimensionen dargestellt werden, z. B. verschiedene Varianten in der

31 Ackermann 2014, Anhang D, E, F und H.

32 Bender 2016.

Überlieferung. Dafür gibt es verschiedene graphische Konventionen, die unterschiedliche Schwerpunkte setzen: Die zwei graphischen Hauptkategorien des Fussnotenapparats und des integralen (im Text integrierten) Apparats können noch in verschiedene Unterkategorien unterteilt werden.³³ In der geschichtsorientierten Editorik ist der am weitesten verbreitete und auch in meiner Edition verwendete Einzelstellenapparat in Fussnotenform quasi der alleinstehende Standard. In dieser Form werden in unmittelbarer Nähe auf den variierenden Text folgenden Fussnoten die Variationen angegeben.³⁴

Während sich die Standards bei der Printedition hauptsächlich aus den Beschränkungen des Buches bzw. der Zweidimensionalität des Papiers ergeben und sich auf das äussere Erscheinungsbild konzentrieren, werden Onlineeditionen durch die Grenzen des digitalen Mediums beschränkt. Zwar können Anmerkungen zum Haupttext vielseitig dargestellt werden, z. B. als Pop-Up, in einem Frame, der beim Scrollen des Haupttexts mitscrollt usw., es gibt aber viel grundlegendere Probleme, die daraus resultieren, dass im Digitalen alles auf Zahlen (Bytes) basiert, deren Bedeutung erst durch die Software gegeben wird. Dies ist vergleichbar mit verschiedenen Alphabeten: Die gedruckten Zeichen sind soweit standardisiert, dass sie jede Person interpretieren kann, sofern sie das einschlägige Alphabet kennt. Wenn sie das Alphabet jedoch nicht kennt, versteht sie auch die dargestellten Laute nicht. In ähnlicher Weise gibt es im Digitalen immer noch viele Möglichkeiten, was die einzelnen Zahlen bedeuten. Wenn Daten von einem Datenträger zum nächsten gegeben werden, z. B. übers Internet von einem Server zu einem Client, muss die Bedeutung der Zahlen standardisiert sein: Nur wenn beide Computer die Zahlen auf die gleiche Weise interpretieren, ist garantiert, dass die Daten so angezeigt werden, wie sie vom Sender intendiert waren.

Nachdem hier zunächst einige der wichtigsten Standards für eine Onlineedition vorgestellt werden, wird in den darauffolgenden Abschnitten deren konkrete Anwendung genauer erläutert.

Zeichenkodierung – UTF-8

Auf der untersten Ebene muss die Zeichenkodierung geregelt sein. Es gibt eine Fülle von verschiedenen Kodierungsmustern, mit denen den Bytes Zeichen aus verschiedenen Sprachen, Sprachregionen bzw. Schriften zugeordnet werden. In den Sprachen mit lateinischem Schriftsystem sind verschiedenen Versionen der ISO 8859 Reihe sowie Unicode am weitesten verbreitet. Die gängige Unicode-Kodierungen UTF-8 sowie die für lateinische Schriften konzipierten ISO 8859 Kodierungen sind alle rückwärtskompatibel mit dem älteren ASCII (*American National Standard Code for Information Interchange*). Das heisst, alle im Englischen verwendeten Buchstaben sind identisch.

33 Hörschemeyer 2013, S. 57-65.

34 Hörschemeyer 2013, S. 59. Patrick Sahle hat der Entwicklung von Standards bei der gedruckten Edition einen ganzen Band gewidmet: Sahle 2013, Bd. 1.

Sobald jedoch Umlaute oder andere sprachspezifische Zeichen verwendet werden, gibt es Probleme beim Textaustausch, wenn die Systeme unterschiedliche Kodierungen verwenden. Das führt zu den bekannten Darstellungen wie: «GrÄſse». ³⁵

Während es für verschiedene Schriften bzw. Sprachregionen eigene Kodierungen von ISO 8895 gibt (nur schon für das lateinische Schriftsystem gibt es zehn Kodierungen für verschiedene gebräuchliche Sonderzeichen), ist es der Grundsatz von Unicode, dass alle auf der Welt benutzten Zeichen darstellbar sind. Es können also nicht nur deutsche Umlaute oder slawische Haček gespeichert werden, sondern (beinahe) alle Schriftzeichen, mit denen in irgendeiner Sprache geschrieben wird, ob Chinesisch, Aramäisch, ein nicht mehr gängiges Runen-Alphabet oder Emoticons. ³⁶ Für internationale Plattformen wie Facebook ist dies ein wichtiger Erfolgsfaktor. Deshalb hat sich Unicode bzw. das auf einer 8-Bit-Abfolge stützende UTF-8 als de facto Standard für die Textübermittlung im Internet durchgesetzt. ³⁷

Für Editionssoftware ist UTF-8-Kompatibilität besonders wichtig: Grundsätzlich sollte eine Transkription möglichst treu die Buchstaben des Originaldokuments wiedergeben können. ³⁸ Da von Hand jedes erdenkliche Zeichen auch gezeichnet werden kann, muss der Computer mit möglichst vielen davon umgehen können. Bei meiner Edition sind die seltenen Zeichen hauptsächlich diakritische Zeichen (z. B. å und û), die in den ISO 8895 Kodierungen fehlen, aber in Unicode vorhanden sind. ³⁹ Oft arbeiten Editionen jedoch nicht oder nicht nur mit lateinischen Buchstaben: z. B. bei Transkriptionen von Manuskripten auf Aramäisch oder auch bei lateinischen Texten, die griechische Zitate in griechischer Schrift enthalten. UTF-8 garantiert, dass all diese Texte mit nur einem Kodierungsstandard abgespeichert werden können. Damit auch alle Unicode-Zeichen dargestellt werden können, muss ein Unicode-kompatibler Satz oder eine Kombination von verschiedenen Satzungen verwendet werden. ⁴⁰ Mit diesen Grundlagen ist bereits ein grosser Schritt gemacht, damit die Software in vielen verschiedenen Szenarien anwendbar ist.

Textkodierung – TEI

Auf dem Quellen-Original als analoger Textträger gibt es zusätzlich zu verschiedenen Zeichen auch Möglichkeiten, einen Text zu strukturieren. Das reicht von bewusst gesetzter graphischer Strukturierung (Überschriften, Absätze, unterschiedliche Schrifttypen etc.) zu Ergänzungen oder Korrekturen einer anderen Person

35 Gillam 2003, S. 38-40 und S. 195-196.

36 <https://www.unicode.org/standard/supported.html> [29.06.2019].

37 <http://www.unicode.org/standard/WhatIsUnicode.html> [29.06.2019]. Vgl. Andrews 2014, S. 177.

38 Vgl. Pichler/Brüvik 2017, S. 220-221, sowie Aliprand 1998, S. 105-106, die das gleiche Argument für die Implementierung von Unicode in Bibliotheks-Katalogisierungs-Software anfügt.

39 https://en.wikipedia.org/wiki/Combining_character [29.06.2019].

40 Gillam 2003, S. 13-14.

(unterschiedliche Handschriften, Randbemerkungen etc.). Unicode kann nur Buchstaben als reinen Text (plain text) kodieren. Wenn die Strukturierung nicht berücksichtigt wird, geht jedoch ein Teil der Information des Dokuments verloren.⁴¹ Bei der gedruckten Edition hat sich der Usus etabliert, diese Eigenheiten graphisch wiederzugeben oder in einem Apparat festzuhalten, damit nicht mehr auf das Original zurückgegriffen werden muss. Im digitalen Bereich wird dafür mit sogenannten Markupsprachen gearbeitet, die den Text unabhängig von der graphischen Darstellung durch Annotierungen strukturieren.

Für die Strukturierung wird der Text mit sogenannten Tags annotiert. Für eine sinnvolle Verwendung der Tags braucht es eine Software, die sie interpretieren kann. Dies kann entweder in einer graphischen Darstellung geschehen (auf dem Bildschirm oder im Druck) oder durch eine Bearbeitung via Software. Graphisches Markup ist z. B. die Kennzeichnung einer Überschrift, die auf dem Bildschirm mit einer anderen, meist grösseren Schriftart angezeigt wird. Ein Beispiel für einen nur von der Maschine ausgewerteten Markup ist eine Lemmatisierung eines Texts: Indem bei jedem Wort dessen Grundform angegeben wird, kann z. B. die Häufigkeit einzelner Wörter in einem Text nachgewiesen werden. Für Menschen wäre ein in Grundformen geschriebener Text vermutlich unverständlich, weshalb diese Information nicht für den Bildschirm aufbereitet wird.

Ob ein Markup graphisch dargestellt wird oder nicht, hängt in erster Linie von der Software ab. Die Annotation hat keine inhärente Bedeutung, sondern wird in der Software definiert. Je nach Ziel der Software kann sie die gleiche Annotation des gleichen Texts auf unterschiedliche Weise interpretieren, darstellen oder auswerten.

Für die Arbeit mit historischen Texten hat sich in den letzten 20 Jahren, trotz wiederkehrender und auch berechtigter Kritik, das Format TEI-XML etabliert.⁴² TEI-XML besteht aus zwei Komponenten: der plattformunabhängigen Markup-Sprache XML (*eXtensible Markup Language*) und dem Vokabular der TEI (*Text Encoding Initiative*). Während XML die Syntax der Annotationen definiert, also den Unicode-Text (plain text) in strukturierten Text (rich text) erweitert, schreibt die TEI vor, welche Bedeutung die Strukturierung hat. Konkret sieht das so aus:

```
<rs @type="person">min gnedigen herrn von Östenrich</rs>
```

In diesem Beispiel wird die Syntax mit den sogenannten XML-Tags (<...> <\...>) markiert. Die Bedeutung der Begriffe im Tag (rs @type="person") ist im Vokabular von TEI definiert: «rs» (referencing string) ist das Tag für eine Referenzierung, während @type="person" spezifiziert, dass es sich um eine Person handelt.⁴³ Im

41 Vgl. Kurz 2017, S. 196 und Gillam 2003, S. 10-11.

42 Vgl. z. B. Pierazzo 2015, S. 117-122, Andrews 2014, S. 178 und Sahle 2013, Bd. 3, S. 341-371.

43 <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html> [29.06.2019].

Zusammenspiel von Syntax (XML) und Semantik (TEI) ist klar, dass der Text «min gnedigen herrn von Österich» im TEI-XML den Namen einer Person bezeichnet.⁴⁴

Ursprünglich wurden bei TEI im Sinne von XML bzw. dessen Vorgänger SGML graphische Aspekte eines Texts bewusst vernachlässigt, um Form und Inhalt strikt zu trennen. Besonders bei der Arbeit mit Manuskripten ist die ursprüngliche Darstellung des Texts im Dokument jedoch ein wichtiger Teil der Information für eine Edition, weshalb diese Ebene nachträglich integriert wurde.⁴⁵ Form und Inhalt bleiben aber nichtsdestotrotz getrennt: Die Information über die Darstellung des Originals fließt nicht zwingend in die Darstellung der digitalen Fassung ein. Das TEI-XML liefert nur Informationen, die in ein anderes Format umgewandelt werden müssen, bevor sie in einer für den Menschen angenehmen Form dargestellt werden. Für eine Onlineedition wird das TEI-XML in der Regel via XSLT (*eXtensible Stylesheet Language Transformation*) in HTML transformiert, das ein Webbrowser interpretieren und auf dem Bildschirm darstellen kann. Während das HTML nur die ungefähre Darstellungsform angibt, definiert ein zusätzliches CSS (*Cascading Style Sheets*) das genaue Aussehen des Texts.⁴⁶

Seit 2007 gilt TEI P5 als Standard, der zwar periodisch, aber nur geringfügig angepasst bzw. erweitert wird. TEI P5 besteht aus einem Tagset von fast 600 Tags. Diese sind in Modulen grob thematisch geordnet. Vier Module werden abgesehen von seltenen Ausnahmen immer angewendet: das *tei*-Modul, das den technischen Rahmen spezifiziert; das *header*-Modul, mit dem Metadaten zum gesamten Text bzw. Dokument abgelegt werden; das *core*-Modul, das textsortenübergreifend die wichtigsten Tags enthält; und das *textstructure*-Modul, das eine grobe Strukturierung des Texts ermöglicht. Zusätzliche Module sind für die genauere Annotierung gedacht, entweder für spezifische Textsorten oder spezifische Auswertungsmöglichkeiten, z. B. die Annotationsmöglichkeiten für Verse oder für Textkritik.⁴⁷

Ein Grund für die breite Akzeptanz des TEI-XML ist die Tatsache, dass sich das TEI-Vokabular erweitern lässt und auch fortlaufend erweitert wird. Spezifisch für Briefeditionen hat die *Special Interest Group for Correspondence* das Vokabular erweitert, um Eigenheiten dieses Genres wie Absender, Empfänger, Sendeort etc. abbilden zu können. Als Teil des *header*-Moduls können nun die korrespondenzspezifischen Metadaten gezielt angegeben werden.⁴⁸ Aber auch für individuelle Projekte kann das Vokabular eigenmächtig erweitert werden.⁴⁹

44 Vgl. Huitfeldt 2017, S. 189-200 und Gillam 2003, S. 11. Einführungen zu TEI-XML gibt es viele, z. B. Kurz 2017, S. 195-251.

45 Sahle 2013, Bd. 3, S. 370-371.

46 TEI-Consortium 2018, S. 191 sowie Pichler/Bruvik 2017, S. 221-222.

47 TEI-Consortium 2018.

48 Stadler 2016 und TEI-Consortium 2018, S. 60-62.

49 Kurz 2017, S. 243-244.

Kodierung der Semantik – LD und IIF

Ein neuer Ansatz im Bereich der Standardisierung ist die Kodierung der Semantik. Die Grundidee dahinter ist, das Vokabular in einer formalen Ontologie so aufzubereiten, dass die Zusammenhänge zwischen den einzelnen Tags nicht nur für Menschen, sondern auch für Computer verständlich ausgedrückt werden. Dafür wird mit dem Datenmodell der Linked Data (LD) gearbeitet: Einzelnen Wörtern oder Wortgruppen wird eine Bedeutung gegeben, indem sie via Prädikat in Relation zu einem anderen Objekt gesetzt werden. Dies geschieht in sogenannten Triples.⁵⁰

Um das vorherige Beispiel zu erweitern:

```
<rs type="person" key="SvOe">min gnedigen herrn von Östenrich</rs>
<relation active="SvOe" ref=https://www.wikidata.org/wiki/Property:P214
passive="http://viaf.org/viaf/45829921">
```

Mit dem neuen Tag <relation> wird in einem Triple eine Beziehung zwischen dem Textausschnitt «min gnedigen herrn von Östenrich» (active) und der Personendatenbank VIAF (passive) dargestellt. Die Beziehung wird via Vokabular von Wikidata dargestellt (ref): Das für Menschen wenig hilfreiche Prädikat P214 zeigt an, dass der Text eine Person bezeichnet, die in VIAF vorhanden ist.⁵¹ In diesem Beispiel wurde der Textausschnitt als Bezeichnung für eine Person identifiziert, und zwar als Bezeichnung für den unter vielen Namen bekannten Herzog Sigmund von Österreich (1427-1496).

Im vorhergehenden Beispiel wurde gezeigt, wie LD mit TEI-XML verbunden werden kann. Einen Schritt weiter geht das 2011 formulierte *International Image Interoperability Framework* (IIF - «triple eye eff»), das sich schnell zum neuen Standard in den Digital Humanities entwickelte. Im Gegensatz zu TEI-XML steht bei IIF nicht der Text, sondern das Bild im Zentrum. Ganz im Sinne von LD ist das Ziel nicht lediglich, den Datenaustausch zwischen Systemen zu fördern, sondern viel grundlegender, die Datensilos verschiedener Institutionen aufzulösen. Konkret heisst das Folgende: Durch eine klar definierte Schnittstelle werden Datenspeicher und Viewer (das visuelle Frontend) entkoppelt – über die Schnittstelle angebotene Daten können in jedem beliebigen kompatiblen Viewer angeschaut werden, unabhängig davon, welche Institution die Daten oder den Viewer zur Verfügung stellt.⁵²

Obwohl 2017 bereits geschätzte 335 Millionen Bilder via IIF zur Verfügung gestellt wurden, ist diese neue Technologie in der Forschungsliteratur zu elektronischen Editionen noch kaum ein Thema.⁵³ Der Ansatz ist aber für die Erstellung einer

50 Ausführlich in Sikos 2015, ein kurzer Einstieg in Toebak 2007, S. 262-264.

51 Zur Verbindung von TEI und LD und dem Tag <relation> vgl. Tupman/Jordanous 2014 sowie de la Iglesia/Moretto/Brodhun 2015, S. 95.

52 <http://iif.io> [29.06.2019]. Zur Auflösung von Datensilos via LOD vgl. de la Iglesia/Moretto/Brodhun 2015, S. 100-101.

53 Z. B. nicht erwähnt im Sammelband von Apollon/Régnier/Bélisle (Hgg) 2017, bei Kurz 2017 oder bei Pierazzo 2015.

Onlineedition aus zwei Gründen interessant. Erstens können Bilder angezeigt werden, unabhängig von deren Speicherort. Es können also auch Bilder in eine Onlineedition integriert werden, die z. B. von einer Bibliothek oder einem Archiv digitalisiert wurden, ohne dass die (meist grossen) Dateien auf einem eigenen Server gespeichert werden müssen. Zweitens kann davon profitiert werden, dass viele andere Institutionen auch Software für die Schnittstelle programmieren. Einerseits können Institutionen dadurch Entwicklungskosten senken, andererseits kann jede Nutzerin der Edition unabhängig von deren Herausgeberinnen selber entscheiden, welcher Viewer für ihre Zwecke am nützlichsten ist.

Der Ausgangspunkt von IIF war zwar der interoperable Transfer von Bilddateien mit dem *Image-API*. Ein *Application Programming Interface* (API) bezeichnet eine Schnittstelle zwischen verschiedenen Systemen und basiert in diesem Fall auf der weitverbreiteten REST-Technologie. Weitere APIs unterstützen aber auch textbasierte Funktionen. Mit dem *Presentation-API* lassen sich Metadaten und Transkriptionen oder auch normalisierte Texte sowie Übersetzungen und Anmerkungen standardisiert übermitteln. Anhand der Metadaten können auch verschiedene Bilddateien als einander zugehörig beschrieben werden, z. B. zwei Seiten eines Briefs, wodurch sie im Viewer zusammen dargestellt werden. Die zu einem Werk gehörigen Metadaten werden als Manifest bezeichnet. Das weniger verbreitete *Content-Search-API* ermöglicht auch das Durchsuchen von Metadaten inklusive Transkriptionen bzw. Volltexten.⁵⁴ Diese textbasierten APIs funktionieren auf den Grundlagen von LD, das Format für den Datenaustausch ist JSON-LD (*JavaScript Object Notation – Linked Data*). Das heisst, für diese Daten muss eine den Ansprüchen von LD genügende formale Ontologie bzw. ein Schema existieren.⁵⁵

54 Vgl. Snyderman/Sanderson/Cramer 2015, S. 17-18 und Raemy 2017a.

55 Diese werden meistens in den Ontologiesprachen OWL oder RDFs geschrieben. Vgl. Toebak 2007, S. 264.

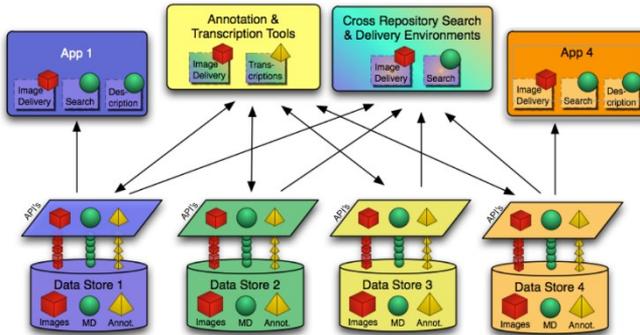


Abb. 5: Interoperabilität des IIIF, indem Datenspeicher und App via API entkoppelt werden. Snyderman/Sanderson/Cramer 2015, S. 17.

Die Transkription des Texts geschieht mit dem W3C *Open Annotation Standard*.⁵⁶ Über einen Annotations-Server⁵⁷ wird zur Bilddatei ein Manifest in JSON-LD erstellt, das Metadaten mit Verweis zur Bilddatei enthält. Das Manifest (die Metadaten) müssen nicht auf demselben Server wie die Bilddatei gespeichert sein – sie werden für die Benutzerin im Viewer zusammengefügt, so dass sie nicht wissen muss, woher die Daten stammen. Die Transkription und falls erwünscht auch inhaltliche Anmerkungen oder Übersetzungen erfolgen via Annotationsfunktion direkt über dem Bild. Diese können in Schichten (layers) abgespeichert werden, wodurch die Benutzerin jederzeit Transkription, Anmerkungen oder Übersetzung ein- bzw. ausblenden und problemlos auf das Original zurückführen kann.⁵⁸

56 Mit dem W3C-Standard kann jede Art von Web-Ressource annotiert werden: <https://www.w3.org/TR/annotation-model/> [29.06.2019].

57 Eine Liste von Software und Ressourcen für IIIF inklusive Annotationsserver findet sich unter: <https://github.com/IIIF/awesome-iiif> [29.06.2019].

58 Ein sehr schönes Beispiel: <https://falkner.freizo.org/mirador/book.cgi?catno=10002> [29.06.2019]. Technische Grundlage bei Snyderman/Sanderson/Cramer 2015, S. 19 und <http://iiif.io/api/annex/openannotation/> [29.06.2019].

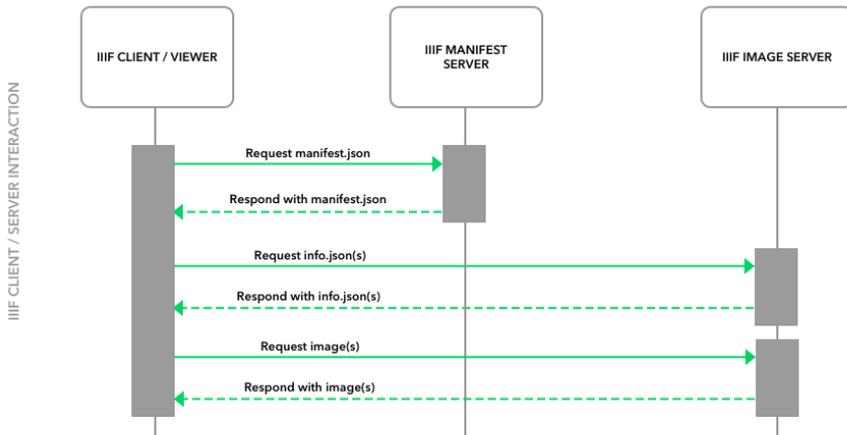


Abb. 6: Kommunikation zwischen Client und Server, um ein Bild mit dazugehörigen Metadaten zu übertragen. <https://iiif.github.io/training/intro-to-iiif/SOFTWARE.html> [29.06.2019].

Wie auch TEI-XML gibt das *Presentation-API* viel Spielraum, um Metadaten nach den eigenen Bedürfnissen zu definieren.⁵⁹ Es wird deshalb in vielen verschiedenen Bereichen eingesetzt, bei denen Bildmaterial mit Metadaten versetzt wird. In der Schweiz sind e-codices und die Digital-Humanities-Laboratorien der Universität Basel sowie die EPFL Vorreiter. Das an der HES Genf angesiedelte Projekt TICKS hat es sich zum Ziel gesetzt, IIF in der Schweiz noch zu weiterer Verbreitung zu verhelfen.⁶⁰ Wie wir anschliessend sehen werden, arbeitet das an der Uni Basel angesiedelte Projekt NIE-INE an einer Editionsplattform, die sowohl mit TEI-annotierten Texten als auch IIF-kompatiblen Bildern arbeitet und gleichzeitig auch LD-fähig ist.⁶¹ Es gibt auch einige Anstösse, um TEI-XML in IIF-gängiges JSON-LD zu integrieren bzw. umzuwandeln.⁶²

59 <http://iiif.io/api/presentation/2.1/> [29.06.2019].

60 Raemy 2018, S. 30-31.

61 NIE-INE 2018.

62 Zur Diskussion vgl. z.B. Brumfiel 2017 und Vigiante 2013. Zwei Transformationstools: <http://dmt.bodleian.ox.ac.uk/metadata-transformation-tools/> [29.06.2019] und <https://github.com/leoba/TEI-2-IIIF> [29.06.2019].

Ausgewählte bestehende Software für Onlineeditionen

Überblick

Software, die beim Erstellen von digitalen bzw. online gestellten Editionen helfen, haben eine lange Tradition. Ein frühes Beispiel ist TUSTEP (*Tübinger System von Textverarbeitungsprogrammen*) bzw. die XML-basierte Version TXSTEP, deren Entwicklung bereits 1966 anfang, die aber auch heute noch benutzt und weiterentwickelt wird.⁶³ Die meisten Programme sind aus einem spezifischen Schritt der Editionsarbeit herausgewachsen und helfen z. B. beim Transkribieren, Annotieren oder Kollationieren. Bei TUSTEP steht vor allem der Vergleich von Texten, das Sortieren und Analysieren von Textelementen sowie der Textsatz für eine gedruckte Edition im Vordergrund. Obwohl z. B. Robinson bereits 2005 als zentralen Aspekt für die Benutzerfreundlichkeit die Forderung stellte, dass eine Software alle Schritte abdecken sollte,⁶⁴ hinkt die Realität den Ansprüchen immer noch hinterher. Immerhin streben die meisten Projekte aktiv in diese Richtung, sei es über eine modulare Architektur, die eine nahtlose Integration verschiedener Softwarekomponenten ermöglicht oder durch eine komplexe, vielseitige Software.

Weil die Programme tendenziell nicht den ganzen Arbeitsprozess abbilden, sind sie entsprechend schwierig zu finden: Man muss nach den verschiedenen Stichwörtern suchen, die beim Erstellen einer Onlineedition durchlaufen werden. Auch bei Software-Listen wie der *Digital Research Tools* (DIRT) sind die Programme unter Teilschritten wie Transkription zu finden.⁶⁵ Bei DIRT und ähnlichen Auflistungen ist die nächste Schwierigkeit, dass es anhand der sehr kurzen und generischen Beschreibungen kaum auf den ersten Blick erkennbar ist, ob die Software für das eigene Projekt nützlich sein könnte. Auch die Beschreibungen auf der Homepage der Anbieter hilft nicht in jedem Fall weiter: Dann muss die Software entweder direkt ausprobiert oder ein Gespräch mit den Herstellern gesucht werden.⁶⁶

Einen konkreteren Eindruck über die Möglichkeiten der Software geben Beschreibungen in der Forschungsliteratur über online bzw. digitale Editionen. Vier Beispiele sind die Beiträge von Michael Bender (2016), von Tara Andrews (2014), von Jörg Hörnschemeyer (2013) und von Peter Robinson (2005).⁶⁷ Ein Problem bei solchen Beschreibungen ist selbstverständlich, dass sie nicht auf dem neusten Stand gehalten werden und sich Software unter Umständen schnell weiterentwickelt oder die Hersteller wegen Finanzierungsschwierigkeiten den Support nicht mehr

63 <http://www.txstep.de/programm.html> [29.06.2019].

64 Robinson 2005, § 26.

65 <http://dirtdirectory.org/> [29.06.2019].

66 Vgl. Kaden/Rieger 2015, S. 67-68.

67 Bender 2016, S. 103-108, Andrews 2014, Hörnschemeyer 2013, S. 35-47 und Robinson 2005.

gewährleisten können. Das Institut für Dokumentologie und Editorik unterhält zwar online eine Liste mit digitalen Editionen, jedoch nicht der dahinterliegenden Software.⁶⁸

Wegen all dieser Vorbehalte wird hier nicht versucht, einen abschliessenden Überblick über bereits bestehende Software für Onlineeditionen zu bieten. Vielmehr werden einige Tools mit unterschiedlichem Hintergrund exemplarisch vorgestellt und auch ähnliche Software zusammen besprochen. Der Fokus wird auf Software gelegt, die das Potential hat, für eine Klein- oder Teiledition brauchbar zu sein. Das heisst einerseits, dass sie einerseits einfach erhältlich sein muss, andererseits jedoch komplex genug, um eine Grundlage für eine kritische wissenschaftliche Edition zu bieten.⁶⁹ Es stehen diejenigen Tools im Vordergrund, die möglichst viele Arbeitsschritte abdecken und die leicht zu bedienen sind.

FuD, TextGrid und ediarum

Im deutschen Raum wurden in den letzten zehn Jahren mit Unterstützung der Deutschen Forschungsgemeinschaft (DFG) einige Forschungsumgebungen entwickelt, deren Fokus auf die kollaborative Arbeit an Manuskripten gerichtet war.⁷⁰ Bedingt durch die Richtlinien von *Horizon 2020* wurde die Möglichkeit geschaffen, die erstellten Forschungsdaten auch in einem Repositorium abzulegen und so zu archivieren. Drei dieser Forschungsumgebungen sind FuD⁷¹, TextGrid⁷² und ediarum⁷³. Da FuD bereits häufig für Briefeditionen eingesetzt wurde und sich dementsprechend für diese Textgattung besonders eignet, werde ich den Fokus auf dieses Tool legen.

FuD nennt sich «eine virtuelle Forschungsumgebung für die Geisteswissenschaften» und verfolgt den Anspruch, geisteswissenschaftliche Projekte von der Transkription über die Indizierung und Annotation bis zur Publikation zu begleiten.⁷⁴ Die modular aufgebaute Open-Source-Software wird auf einem Server installiert, wodurch von verschiedenen Geräten aus übers Internet auf die dort gespeicherten Daten zugegriffen werden kann. Über die Benutzerverwaltung können verschiedene Forscherinnen einem Projekt zugewiesen werden und daran arbeiten. FuD wurde und wird hauptsächlich anhand der Bedürfnisse von grösseren Forschungsprojekten entwickelt, weshalb die kollaborative Arbeit einen zentralen Stellenwert hat.⁷⁵ Allerdings ist es keine echte plattformunabhängige Cloud-Lösung, da die Benutzerinnen auch

68 <http://www.digitale-edition.de/> [29.06.2019].

69 Zu Software, die diesen Ansprüchen nicht gerecht wird, vgl. Robinson 2016b, S. 4-5.

70 Zum Begriff der Forschungsumgebung und welche Anforderungen sie erfüllen sollten vgl. Franke 2018.

71 <http://fud.uni-trier.de/> [29.06.2019].

72 <https://textgrid.de/> [29.06.2019].

73 <http://www.bbaw.de/telota/software/ediarum> [29.06.2019].

74 <https://fud.uni-trier.de/software/> [29.06.2019].

75 Telefongespräch mit Yvonne Rommelfanger, Mitarbeiterin bei FuD, vom 09.07.2018.

lokal ein Programm installieren müssen und dafür ein Windows oder MacOS Betriebssystem brauchen. Obwohl FuD für diverse Forschungsvorhaben verwendet werden kann, ist der am weitesten verbreitete Einsatz eindeutig bei der Erstellung einer Edition, ob in Volltext oder nur in Regestenform, digital oder gedruckt.⁷⁶

In der Basisinstallation können Texte mit einer hauseigenen Auswahl an TEI-Tags annotiert, Digitalisate und Anhänge abgespeichert und Verknüpfungen zwischen den Dokumenten erstellt werden. Für die Erstellung von Transkriptionen digitalisierter Unterlagen kann FuD zusammen mit dem Transkriptionstool Transcribo⁷⁷ benutzt werden. Auch weitere Hilfsmittel wie das Bibliographietool Zotero lassen sich integrieren. Die Forschungsumgebung bietet Suchfunktionen sowie die Möglichkeit, die Texte automatisch analysieren zu lassen. In diesem Zusammenhang ist FuD hauptsächlich ein Instrument, um Forschungsdaten zu ordnen und auszuwerten. Für die Archivierung von Forschungsdaten können Angehörige der Universität Trier die Daten von abgeschlossenen Projekten mit nur wenigen Klicks ins Trier Repository laden, wo sie je nach Einstellung auch für andere eingeloggte Benutzerinnen herunterladbar und weiterverwendbar sind.

Zusätzlich bietet das Team von FuD die Möglichkeit, weitere kostenpflichtige Dienstleistungen in Anspruch zu nehmen. Die Dienstleistungen umfassen im Wesentlichen drei Bereiche: Beratung zum Aufbau des Projekts inklusive Parametrisierung der Software, damit sie genau den Bedürfnissen des Projekts entsprechen; Weiterentwicklung der Software für spezielle Bedürfnisse; Hilfe beim Import und Onlinestellen von Daten. Für das Erstellen einer Onlinedition bietet das Team von FuD zwei Möglichkeiten, um eine Website direkt auf der bereits vorhandenen Datenbank aufzubauen. Im cake-PHP kümmert es sich um die graphische Darstellung und zusätzliche Informationsseiten, die nicht in der Datenbank sind, wie z. B. ein Impressum. Im ViDa-System werden die Daten mit Hilfe von WordPress dargestellt, was der Editionerstellerin ermöglicht, selber eine Website um die Daten herum aufzubauen. Bei beiden Versionen wird eine Suchfunktion implementiert und selbstverständlich werden sowohl Bild- wie auch Text-Daten aus der Datenbank angezeigt.⁷⁸

In seinem Grundaufbau ist FuD sehr ähnlich zu ediarum der Berlin-Brandenburgischen Akademie der Wissenschaft oder zu TextGrid-Laboratory. Die Dokumentation und Zugänglichkeit von ediarum sind aber nicht vergleichbar mit FuD – man kann die Software zwar von Github herunterladen, muss aber selber das technische Wissen haben, um sie auf einem eigenen Server zu installieren.⁷⁹ Vermutlich sind die

76 <https://fud.uni-trier.de/community/referenzen/> [29.06.2019].

77 <http://transcribo.org/de/> [29.06.2019].

78 FuD 2018. Vgl. auch den Fragenkatalog für das Erstellen einer neuen Forschungsumgebung: <https://fud.uni-trier.de/service/beratungs-service/projektplanung-fragenkatalog/> [29.06.2019].

79 <http://www.bbaw.de/telota/software/ediarum> [29.06.2019].

Funktionen auch weniger umfangreich. Für technisch unversierte Benutzerinnen ist ediarum sicherlich ungeeignet.

Hinter TextGrid hingegen steht eine breite Community, die in regelmässigen Treffen und via Publikationen die wissenschaftliche Reflektion im Bereich der digital unterstützten Editionsarbeit stark vorantreibt.⁸⁰ Vor diesem Hintergrund ist es passend, dass TextGrid im Gegensatz zu FuD und ediarum bereits früh mit LOD arbeitete, auf den Zug des IIF aufgesprungen ist und Daten für das IIF *Presentation-API* und *Image-API* unterstützt.⁸¹ Die Software selber ist quellfrei und kann auf verschiedenen Betriebssystemen einfach installiert werden, sie ist jedoch relativ komplex in der Benutzung. Eine nur vorläufig geklärte Frage bei TextGrid ist allerdings die Finanzierung: Obwohl die Einschätzung von Fotis Jannidis aus dem Jahr 2015 diesbezüglich sicherlich zu pessimistisch war und vorerst eine weitere Finanzierung im Rahmen von DARIAH gefunden wurde, wird momentan zwar der Betrieb sichergestellt, Weiterentwicklungen aber nur im Rahmen von Anwendungsprojekten finanziert.⁸²

Im Vergleich zu den anderen Lösungen ist die Benutzerfreundlichkeit ein Alleinstellungsmerkmal von FuD: Es wird nicht mit einem XML-Editor gearbeitet, sondern für jeden Dokumenttyp wird eine Maske erstellt, in die die nötigen bibliographischen Metadaten eingegeben und mit der Tags für die inhaltliche Annotation aus einer Dropdownliste ausgewählt werden können. Somit ist zur Benutzung kein vertieftes Verständnis von XML notwendig. Auch bemerkenswert ist der sorgfältig ausgearbeitete Finanzierungsplan für einen dauerhaften Regelbetrieb, der die Grundlage für eine langfristige Finanzierung bietet. Für ein Klein- oder Teileditionsprojekt ist jedoch schnell klar, dass der benutzerfreundliche Service kaum zahlbar ist: Bereits für die Basis-Installation werden schnell mehrere tausend Euro fällig und für eine Webpräsenz fallen zusätzlich jährliche Serverkosten an.⁸³

Corpus Corporum und Digital Editions

Ein gänzlich anderes Konzept verfolgt Corpus Corporum, eine Onlineplattform für lateinische Texte.⁸⁴ Sie wurde von Philipp Roelli mit einem dreijährigen Stipendium des SNF als Alternative zur kostspieligen Brepols-Datenbank⁸⁵ aufgebaut, um die Texte Angehörigen von finanziell schlechter gestellten Hochschulen frei zur Verfügung zu stellen. Die dahinterstehende Datenbank enthält hauptsächlich von Roelli

80 <https://de.dariah.eu/textgridlab> [29.06.2019]. Vgl. Wegstein/Rapp/Jannidis 2015, S. 23-35.

81 https://dev.textgridlab.org/doc/services/submodules/tg-iiif-metadata/docs_tgrep/index.html [29.06.2019] und de la Iglesia/Moretto/Brodhun 2015, S. 95-98.

82 Wegstein/Rapp/Jannidis 2015, S. 33-35 und Telefongespräch mit Mirjam Blümm von DARIAH vom 23.07.2018.

83 Telefongespräch mit Yvonne Rommelfanger vom 09.07.2018. Der Finanzierungsplan wird in Minn et al. 2016 vorgestellt.

84 <http://www.mlat.uzh.ch/> [29.06.2019].

85 <http://www.brepols.net/Pages/BrowseBySeries.aspx?TreeSeries=LLT-O> [29.06.2019].

angefertigte Transkriptionen und offen verfügbare Texte von anderen Projekten wie Perseus oder der MGH. Grundsätzlich können alle Forscherinnen Texte beisteuern, die Datenbank kann also auch als kostenloses Repositorium für Forschungsdaten von Latinisten genutzt werden. Beispiel einer gelungenen Zusammenarbeit sind die *Carmina* von Michel de l’Hospital, die Loris Petris der Forschung nicht nur als gedruckte Edition, sondern auch als Text im Corpus Corporum veröffentlichte.⁸⁶

Die Ausführung der Plattform ist sehr einfach: TEI-annotierte XML-Files werden via Plattform-Verantwortlichen hochgeladen.⁸⁷ Wenn nötig passt dieser auch das XML den Vorgaben an. Da die online dargestellten Texte sowieso nur eine beschränkte Anzahl von TEI-Tags anzeigen, ist die Anzahl benötigter TEI-Tags überschaubar. Für die Speicherung werden jedoch auch andere XML-Elemente berücksichtigt: Die Texte können (soweit es die Urheber erlauben) als TXT oder ursprüngliches XML heruntergeladen und so je nach Bedürfnissen weiter bearbeitet oder ausgewertet werden.

Trotz einfachem Design bietet die Onlinemaske einige starke Funktionen. Neben der mächtigen Suchfunktion ist dies auch die Verknüpfung mit anderen Projekten: Z. B. wurden Wörterbücher so verlinkt, dass die Benutzerin auf ein beliebiges Wort klicken kann und darauf, zusammen mit der Erklärung des Wortes, eine morphologische Analyse sowie Lemmatisierung angezeigt werden. Auch sind die Autoren der Texte, soweit bekannt, mit verschiedenen Datenbanken verlinkt: VIAF (*Virtual International Authority File*), DNB (Deutsche Nationalbibliothek), Mirabile und Wikidata.⁸⁸ Dadurch steht die Datenbank nicht nur im Geiste der Open-Source-Bewegung, sondern auch von Linked Open Data (LOD).

Da die Plattform grösstenteils in Eigenregie von einer einzelnen Person vorangetrieben wird, ist sie auch stark davon abhängig, wie viel Zeit und Ressourcen diese einsetzen kann. Ideen zur Weiterentwicklung hat Philipp Roelli viele, momentan sind finanzielle Mittel aber nur für die Serverkosten vorhanden. Der Rest wird in Form von Freiwilligenarbeit geleistet.⁸⁹ Die Gefahr bei einer solchen Konstellation ist gross, dass das Projekt irgendwann aus dem Interessensfokus gerät und technisch überholt sein wird: Beispiele von verwaisten Editionsplattformen gibt es viele.⁹⁰ Gerade weil das Projekt immer nur mit wenigen personellen und finanziellen Ressourcen auskommen musste, hat es aber auch gute Voraussetzungen, in einem anderen Rahmen weiterverwendet zu werden. Entscheidend dafür sind der einfache Aufbau der Website und die Nutzung von Open-Source-Software sowie des verbreiteten TEI-Standards. Die Textgrundlage kann problemlos in eine andere Umgebung integriert werden.

86 Roelli 2014.

87 <http://www.mlat.uzh.ch/MLS/impressum/impressum.php> [29.06.2019].

88 Technische Erklärung dazu bei Roelli 2014, S. 294-295.

89 Telefongespräch mit Philipp Roelli vom 16.05.2018.

90 Vgl. Sahle 2013, Bd. 2, S. 78.

Ein ähnliches Konzept verfolgend, gewissermassen aber auch ein Gegenteil, sind die *Digital Editions* der *Romantic Circles*, die sich (englischen) Texten aus dem 18. und 19. Jahrhundert verschrieben haben. Die Ähnlichkeit liegt darin, dass die *Digital Editions* auch bereits fertig edierte Texte aufnehmen und diese bei Bedarf mit TEI annotieren. Wie bei Corpus Corporum ist die Speicherung von dazugehörigen Bilddateien (Digitalisate) nicht vorgesehen. Während Corpus Corporum jedoch das Ziel verfolgt, eine möglichst umfangreiche Text-Datenbank anzubieten, gehen die Beiträge der *Digital Editions* durch einen Peer-Reviewing-Prozess, bevor sie auf die Plattform aufgenommen werden. Das Design ist zwar aufwändiger gemacht, technisch ist jedoch das Corpus Corporum klar weiter. So existieren z. B. bei den Digital Editions keine Verlinkungen im Sinne von LOD.⁹¹

NIE-INE und Freizo

NIE-INE (*Nationale Infrastruktur für Editionen - Infrastructure nationale pour les éditions*) ist von den untersuchten Projekten sicherlich das ambitionierteste. Der Aufbau einer «Arbeits- und Publikationsplattform» für wissenschaftliche Editionsprojekte der Schweiz wird von swissuniversities im Rahmen des Förderprogramms «Wissenschaftliche Information: Zugang, Verarbeitung und Speicherung» bis im September 2019 finanziell unterstützt.⁹² Angebunden ist NIE-INE einerseits an das *Forum für Edition und Erschliessung* (FEE), das das editorische Knowhow mitbringt, und andererseits an das *Data and Service Center for the Humanities* (DaSCH), das die technische Infrastruktur liefert. Da das Projekt vom FEE koordiniert wird, ist der Schwerpunkt der erarbeiteten Plattform klar die kritisch wissenschaftliche Onlineedition: Sie soll den «Bedürfnissen umfangreicher und komplexer Editionsprojekte» gerecht werden, aber auch für kleinere Editionsvorhaben verwendet werden können.⁹³

Die Plattform ist noch im Aufbau und es kann nur bruchstückhaft beurteilt werden, ob die gesteckten Ziele zu einer guten Umsetzung führen werden. Bereits implementiert ist die Datenarchitektur, die sich ganz der Linked Data verschrieben hat: (Text)-Daten werden in einem RDF-Triplestore abgelegt und Bilddateien auf einem IIIF-Bildserver. Die benutzten Technologien werden vom DaSCH als Open-Source-Software entwickelt und bestehen aus dem IIIF-Bildserver Sipi, der Serverapplikation Knora und der Benutzeroberfläche Salsah. Knora hält quasi die

91 <https://www.rc.umd.edu/editions> [29.06.2019] und <https://www.rc.umd.edu/pubinfo/policy.html> [29.06.2019].

92 <https://www.swissuniversities.ch/de/organisation/projekte-und-programme/p-5/> [29.06.2019].

93 <https://fee.unibas.ch/de/nie-ine/> [29.06.2019] und NIE-INE 2018. Vgl. Raemy 2017b, S. 26-29.

verschiedenen Komponenten zusammen und ermöglicht es, Daten versioniert zu speichern, zu teilen und zu annotieren.⁹⁴

Ziel ist einerseits eine Forschungsumgebung zu entwickeln, in der neue Editionen von Grund auf erstellt werden können, andererseits aber auch Datenimport-Schnittstellen zu anderen Forschungsumgebungen wie der FuD bereitzustellen, damit bereits bestehende Editionen importiert und auf Dauer archiviert und via Internet zugänglich gemacht werden können. Wie FuD orientiert sich auch NIE-INE in erster Linie an den Bedürfnissen von mit ihnen zusammenarbeitenden Projekten, um für diese eine geeignete Plattform zu entwickeln. Da auch fertige digitale Editionen übernommen werden, die eine dauerhafte Weblösung brauchen, ist der Import und Export in TEI-XML selbstverständlich.⁹⁵ Ein besonders weit fortgeschrittenes Pilotprojekt ist die Historisch-kritische Online-Edition von Kuno Raebers Lyrik. Dort werden zwar bereits die neuen Technologien von IIF und LD angewandt, jedoch ist dies noch nicht sichtbar; die IIF-Manifeste sind nicht als solche erkennbar und können deshalb nicht in anderen Viewern betrachtet werden.⁹⁶

Wie bei FuD werden kostenpflichtige Dienstleistungen angeboten. Dies ist neben der allgemeinen Beratung und Parametrisierung der Software für projektspezifische Bedürfnisse auch das Erstellen von Data Management Plans oder von geeigneten Ontologien für das RDF-Datenmodell.⁹⁷ Diese werden zusammen mit der Software auf der Open-Source-Software-Plattform Github in den öffentlichen Besitz übergeben. Momentan ist der Weg über den persönlichen Kontakt bzw. die angebotenen Dienstleistungen der einzige Weg, eine eigene Onlineedition zu erstellen: Eine Plattform im eigentlichen Sinne mit einer Standard-Parametrisierung existiert noch nicht, Überlegungen dazu sind jedoch vorhanden. Für eine oder mehrere Standardontologien würden die für die Plattform entwickelten Ontologien eine gute Grundlage bieten.⁹⁸

Ein ähnliches auf IIF und LD aufbauendes Angebot bietet *Freizo* von *Data Futures*.⁹⁹ *Data Futures* hat gegenüber NIE-INE einen fünfjährigen Vorsprung und hat schon über 100 Projekte realisiert. Trotzdem ist immer noch Voraussetzung für das Erstellen einer Onlineedition, dass mit dem Team zuerst kostenpflichtig eine dem Projekt angepasste Umgebung eingerichtet wird. Da sie bereits kleinere Editionsprojekte erstellt haben, wie z. B. die im Rahmen einer Masterarbeit transkribierten Falkner Stammbücher, haben sie jedoch auch bereits Erfahrungen und vorgefertigte Workflows, die für ähnliche Projekte wiederverwendet werden können. Für ein solches

94 <https://dhlbasel.github.io/Sipi/documentation/overview.html> [29.06.2019], <https://docs.knora.org/paradox/01-introduction/what-is-knora.html> [29.06.2019] und <https://dhlbasel.github.io/Salsah/> [29.06.2019]. Zur gegenwärtigen Umsetzung, vgl. NIE-INE 2018 S. 3.

95 NIE-INE 2018.

96 <http://raeber.nie-ine.ch/init> [29.06.2019].

97 NIE-INE 2018. Eine kostenlose allgemeine Beratung gibt es bei FEE.

98 Telefongespräch vom 04.07.2018 mit Sascha Kaufmann, Chefentwickler bei NIE-INE.

99 <https://www.data-futures.org/freizo.html> [29.06.2019].

Vorhaben ist mit Kosten von zirka 3'000-10'000 Euro zu rechnen, je nachdem wie viel von bereits bestehenden Projekten übernommen werden kann. Momentan in der Dienstleistung nicht inbegriffen ist eine langfristige Archivierung bzw. Zugänglichkeit zu den Daten – die Forscherin muss sich selber um einen Server kümmern und diesen auch selber zahlen, falls ihr nicht in einem universitären Rahmen eine entsprechende Infrastruktur zur Verfügung gestellt wird. Zukünftig will *Data Futures* jedoch ein Netzwerk von Repositorien in ihr Angebot mit einbeziehen.¹⁰⁰

Juxta Editions

Juxta Editions ist von den untersuchten Tools dasjenige, bei dem man am schnellsten starten und einen Einblick bekommen kann: Es handelt sich um eine echte Cloud-Lösung, bei der man ausser der Registrierung (E-Mail und Passwort) keinerlei Hürden hat, um ein erstes Editionsprojekt zu beginnen. Nachdem man sich ein kostenloses Konto eröffnet hat, bekommt man auch Zugang zu verschiedenen Video-Tutorials, die die Grundfunktionen der Software aufzeigen.¹⁰¹

Juxta Editions kann Bilder und TEI-XML-Dateien importieren, wobei nur diejenigen TEI-Tags für die Bearbeitung und Darstellung berücksichtigt werden, die die Software kennt.¹⁰² In einer Paralleldarstellung von Bild und Text kann eine Transkription des Digitalisats erstellt werden. Im Gegensatz zu Ansätzen wie diejenigen von Transcribo oder der Annotationsfunktion von Mirador geschieht dies ohne genaueren Zusammenhang zwischen Bild und Text. Der ursprüngliche Teil von Juxta Editions war die Kollationierung, die relativ einfach gebaut ist: Die Texte werden verglichen und Unterschiede in verschiedenen Darstellungsarten hervorgehoben.¹⁰³ Das Tool bietet einen einfachen TEI-XML-Export und ein Standard «Scientific Archiving»-Design, um das Projekt als Website zu veröffentlichen und anderen Wissenschaftlerinnen zugänglich zu machen. Bei Bedarf kann das Design auch geringfügig angepasst werden.

Mit dem Gratiskonto ist man allerdings relativ beschränkt, da man nur ein TEI-XML erstellen kann. Kostenpflichtige Features sind das Publizieren der Edition im Internet, die kollaborative Nutzung oder das Erstellen von mehreren Editionen. Wie bei Cloud-Software üblich, müssen monatliche Gebühren bezahlt werden.¹⁰⁴ Dies ist für ein Projekt, das langfristig archiviert werden soll, jedoch nicht langfristig finanziert wird, ein grosser Nachteil. Noch negativer ins Gewicht fällt bei Juxta Editions allerdings, dass offensichtlich kein Service garantiert ist: Als ich mich nach ersten

100 <https://www.data-futures.org/falkner.html> [29.06.2019], <https://www.data-futures.org/services.html> [29.06.2019] und Telefongespräch vom 25. Juli 2018 mit Peter Cornwell von Data Futures.

101 <https://www.juxtaeditions.com/editions> [29.06.2019].

102 Juxta Editions orientiert sich an TEI-Lite: <http://www.juxtaeditions.com/features> [29.06.2019].

103 Andrews 2014, S. 183.

104 <http://www.juxtaeditions.com/plans> [29.06.2019].

Tests nicht mehr einloggen und auch kein neues Konto eröffnen konnte, blieb eine entsprechende E-Mail-Anfrage unbeantwortet.¹⁰⁵

Zwischenfazit

Auch nach intensiver Beschäftigung mit den verschiedenen Angeboten fällt es schwer, eine Entscheidung für eine Software-Lösung zu treffen. FuD fällt klar schon aus Kostengründen weg: Das System ist für eine Klein- oder Teiledition eindeutig eine Nummer zu gross. Corpus Corporum ist zwar ein interessantes Projekt, für mich jedoch nicht brauchbar, da ich mit deutschen und nicht mit lateinischen Texten arbeitete. NIE-INE ist erst im Aufbau; in Gesprächen mit den Verantwortlichen ist zwar von vielen Möglichkeiten die Rede, es ist aber schwierig zu eruieren, wie gross der Aufwand für die Umsetzung eines eigenen Projekts sein würde. Ein mit NIE-INE erstelltes Projekt müsste als Zusammenarbeit für deren Weiterentwicklung aufgegleist werden. Juxta Editions hingegen bietet nur sehr grundlegende Funktionalitäten und, noch schwerwiegender, hat sich nicht als vertrauenswürdig erwiesen.

Teilweise lassen sich diese Schwierigkeiten dadurch erklären, dass FuD, NIE-INE und die weiteren Forschungsumgebungen auf mittlere und grössere Editionsprojekte ausgerichtet sind. Für eine Klein- oder Teiledition sind die zentralen Funktionen des kollaborativen Arbeitens und der Workflowdefinition unwichtiger, dafür fallen wenig berücksichtigte Anforderungen wie eine einfache Einarbeitung und eine umfassende technische Betreuung viel stärker ins Gewicht.

In einem Beschaffungsverfahren müsste spätestens an dieser Stelle die Eigenentwicklung in Betracht gezogen werden. Für eine Klein- bzw. Teiledition ist dies kaum eine Lösung, da die Entwicklung einer Softwarelösung noch viel mehr Ressourcen verschlingt als das Einkaufen eines bereits bestehenden Produkts. Ein Blick in die Forschungsliteratur zeigt, dass meine Anforderungen für viele ähnliche Projekte genauso gelten. Deshalb soll in den folgenden Abschnitten eine Auflistung von Anforderungen gesammelt werden, die als Anstoss für die Neu- oder Weiterentwicklung einer für Klein- und Teileditionen interessanten Editionsplattform gelten können. Neben der dort aufgeführten Einbettung der Anforderungen in den Kontext von bereits vorhandenen Technologien gibt es im Anhang einen Anforderungskatalog, der die Anforderungen technologieneutral in Kann- und Muss-Kriterien unterscheidet und als Ausgangslage für eine Ausschreibung dienen könnte. Da keine spezifische Lösung angestrebt wird, ist der Katalog relativ generisch auf hoher Flughöhe geschrieben.

105 E-Mail vom 29.06.2018.

Anforderungen an eine Editionsplattform

Hürden beseitigen

Auf dem Weg zu einer Onlineedition gibt es drei grosse Hürden für die Forscherin: sie muss *langfristig* in der Lage sein, nötiges *technisches Knowhow* sowie *Geld* in das Projekt zu investieren. Die Edition soll ja nicht nur aufgebaut, sondern auf Dauer zugänglich und nutzbar sein. Das erste Ziel einer Editionsplattform muss sein, diese Hürden zu beseitigen. Eine ganz grundlegende Anforderung ist deshalb die Benutzerfreundlichkeit.¹⁰⁶ Im Idealfall heisst das, dass sich die Plattform am Web 2.0 orientiert bzw. der Idee der sozialen Medien entspricht: Ein Login und das nötige editorische Grundwissen sollten die einzigen Voraussetzungen sein, um eine einfache Onlineedition zu veröffentlichen.

Die Software ist idealerweise einfach auffindbar mit einer möglichst nützlichen Beschreibung der wichtigsten Funktionalitäten. Weiter sollte es möglich sein, ohne Softwareinstallation, Kontaktaufnahme oder vorhergehende Definition von Workflow oder Ontologie eine einfache Version der Software zu benutzen. Für komplexere Editionsprojekte sind diese Voraussetzungen überwindbar und sogar wichtig, da eine gute Parametrisierung für das erfolgreiche Abschliessen des Vorhabens zentral sein kann. Für eine Klein- oder Teiledition ist jedoch eine schnelle und einfache Benutzung wichtiger als eine ideal parametrisierte Lösung.

Bei der bestehenden Editionssoftware handelt es sich, verallgemeinernd, um komplexe Programme, denen die Komplexität auf den ersten Blick anzusehen ist.¹⁰⁷ Es ist zwar positiv, dass sie gute Hilfsmittel für den Einstieg bieten, sei es als Benutzerhandbuch, als frei editierbares Wiki oder als Video.¹⁰⁸ Die Benutzeroberfläche sollte jedoch trotz Komplexität zumindest für einen Einstieg einfach gestaltet sein, indem die zentralen Funktionen hervorgehoben werden, während erweiterte, weniger zentrale Funktionen versteckter sein dürfen. Um dieses Ziel zu erreichen, müssen sich die Entwicklerinnen im Klaren sein, welche der Funktionen für die Forscherin die wichtigsten sind.

Damit die Software ohne Installation nutzbar ist, braucht es eine technische Infrastruktur, die auch das Server-Hosting anbietet (Cloud Service). Auf einem Cloud-Server gehostete Software wird direkt vom Softwareanbieter mit den neusten Updates versehen. Dass die Software immer verfügbar ist und die Server-Hardware periodisch ausgewechselt wird, ist heutzutage selbstverständlich. Die Aktualisierung von Hardware und Software, besonders im Hinblick auf neue Browser-Technologien,

106 Kaden/Rieger 2015, S. 63-75.

107 Vgl. Kaden/Rieger 2015, S. 74 und Bender 2016, S. 206-208.

108 <https://fud.uni-trier.de/docs/handbuch/> [29.06.2019] und <https://wiki.de.dariah.eu/display/TextGrid/Main+Page> [29.06.2019].

ist eine wichtige Grundlage für die nachhaltige Nutzung der Onlineedition und auch Teil der FAIR-Prinzipien.¹⁰⁹

Der Nachteil von Cloud-Lösungen ist, dass es meistens weniger Parametrisierungsmöglichkeiten gibt. Da für Klein- oder Teileditionen die Installation und Parametrisierung einer komplexen Software eine grosse personelle und finanzielle Hürde darstellen, muss bei diesen Projekten die Bereitschaft bestehen, diesen Nachteil zu akzeptieren und mit einigen technischen Einschränkungen zu leben. Problematischer für kleine Projekte sind die periodischen Kosten: Eine Editionsplattform sollte diese nicht auf die Nutzerin abwälzen, sondern ein alternatives Finanzierungsmodell finden. Für viele kleine Projekte wird dies den Ausschlag geben, ob eine Onlineedition überhaupt in Frage kommt.

Wenn nicht die Nutzerinnen für die langfristigen Kosten aufkommen können, gibt es verschiedene Finanzierungsmodelle: Der Ansatz, dass die Plattform komplett von einem Unternehmen getragen wird, ist unrealistisch. Im Vergleich zu erfolgreichen Plattformen wie Facebook, das mit den Daten der Benutzerinnen viel Geld machen kann, haben die Daten einer Editionsplattform in der Wirtschaft viel weniger Wert. Es gibt zwar z. B. mit dem *Bentham Project* den Ansatz, per Crowdsourcing Transkriptionen als Basis für Editionen herzustellen, die dann nur noch kostenpflichtig zugänglich sind. Nutzerinnen sind aber viel weniger bereit, kostenlose Transkriptionsarbeit zu leisten, wenn sie keinen direkten Nutzen davon haben bzw. unter Umständen sogar den kostenlosen Zugang zu den von ihnen erstellten Daten verlieren.¹¹⁰

Viel eher kommt das alternative Modell der Open-Source-Software in Frage. Als Beispiel sei WordPress genannt: Die Software ist öffentlich zugänglich und jede kann sie herunterladen, abändern oder auf einem beliebigen Server installieren. Mit dem Code selber darf zwar kein Geld verdient werden, die Software kann aber zusammen mit kostenpflichtigen Dienstleistungen zur Verfügung gestellt werden. Bei WordPress sind das z. B. bereits auf einem Server installierte Instanzen, die von der Benutzerin als Software as a Service benutzt werden können.¹¹¹

Abgesehen von Juxta Editions sind alle der besprochenen bestehenden Software-Lösungen für Online-Editionsplattformen quelloffen. Wenn die Serverkosten von einer universitären Einrichtung als Teil der Open-Access-Strategie gedeckt werden, muss die Editionsplattform keine periodischen Grundgebühren erheben.¹¹² Als Finanzierungsmodell für die Weiterentwicklung werden im Bereich der Editionsplattformen Forschungsgelder eingesetzt oder Dienstleistungen zum Produkt verkauft. Bei

109 Wilkinson et al. 2016. Zur Anforderung der Verfügbarkeit vgl. Wieder 2015, S. 184-187.

110 Robinson 2016b.

111 <https://wordpress.com/> [29.06.2019]. Zu den Vorteilen von Open Source für Forschungsumgebungen vgl. Franke 2018, S. 64.

112 Vgl. Apollon/Régnier/Bélisle S. 24-25 und Drott 2006, S. 94-96. Zu verschiedenen Finanzierungsmodellen für Open-Access-Publikationen vgl. Meroni 2016.

den Onlineeditionen sind diese Dienstleistungen meistens im Rahmen von Workflow-Definition, Datenimport, Datendarstellung und projektspezifischen Weiterentwicklungen.

Um die finanziellen Bürden gerechter zu verteilen, wäre auch ein Anwenderinnenverein ein sinnvolles Modell. In einem solchen Verein tauschen sich Anwenderinnen darüber aus, welche Features für die Zukunft ihrer Anwendungen relevant sind. Wenn sich verbreitete Bedürfnisse herauskristallisieren, können sich kleinere oder grössere Gruppen die Kosten teilen, um gewünschte Erweiterungen programmieren zu lassen.¹¹³ Damit die kleinen Projekte von Weiterentwicklungen profitieren können, soll die Plattform also auch für mittlere, falls nicht auch für grössere Projekte attraktiv sein, die eher die entsprechenden Mittel zur Verfügung haben. Trotzdem dürfen die Bedürfnisse von Kleinprojekten nicht aus dem Blick verloren gehen.

Funktionalitäten – die Minimalanforderungen

Die Editionspraxis ist sowohl für gedruckte als auch digitale Editionen extrem heterogen. Dementsprechend umfangreich fallen die gewünschten Anforderungen aus. Gleichzeitig gibt es aber auch ganz eindeutige Minimalanforderungen an die Funktionalitäten einer Editionsplattform. Die wichtigsten hängen von zwei Faktoren ab: von den Anforderungen an eine FAIR-kompatible Sicherung der Forschungsdaten und von den gängigen Standards im Bereich der Onlineedition.

Für die FAIR-Kompatibilität muss eine Onlineedition in erster Linie den technischen Aspekten einer langfristig nutzbaren Datenbank genügen. Bereits angesprochen wurden die periodische Migration auf neue Servertechnologie sowie eine periodische Anpassung der Darstellungstechnologie an die neusten gängigen Browser. Auch ein FAIR-Prinzip ist die Zitierfähigkeit einer Onlineressource. Dazu wird ein Persistent Identifier (PID) vergeben. Die Editionsplattform kann entweder die Registrierung für eine PID als Service anbieten oder zumindest auf geeignete Ressourcen verweisen, um der Forscherin die Meldung zu vereinfachen.¹¹⁴

Als Forderung der FAIR-Prinzipien müssen die Lizenzangaben für die weitere Benutzung ersichtlich sein, ebenso wie die nötigen Metadaten für das Verständnis der Daten.¹¹⁵ Dazu muss es möglich sein, zwei Arten von Inhaltsseiten anzuzeigen: einerseits die Seiten der Edition selber (Bild, Text und Metadaten) und andererseits «normale» Webseiten, auf denen zusätzliche Informationen zum Editionsprojekt publiziert werden können. Diese können in einem bereits bestehenden Content Management System (CMS) wie WordPress oder Typo3 dargestellt und integriert werden. Im Fall

113 Ein Beispiel ist der Verein OneGov.ch, bei dem sich verschiedene Kantone, Gemeinden und Fachpartner zusammengeschlossen haben, um Software im Bereich e-Government als Open-Source-Software zu entwickeln. <https://www.onegov.ch/> [29.06.2019].

114 Wilkinson et al. 2016.

115 Wilkinson et al. 2016.

meiner Edition würden hier zusätzlich zu den Lizenzangaben eine Beschreibung des Projekts, die Editionsrichtlinien und die inhaltliche Auswertung der Quellen dazugehören.

Das letzte FAIR-Prinzip geht bereits in den zweiten Punkt der gängigen Standards über: Es muss Exportmöglichkeiten ohne Datenverlust geben. Dies ist einerseits wichtig, falls die Forscherin die Plattform wechseln möchte, andererseits ist es aber auch ein Teil der Exitstrategie, falls die Plattform aus irgendwelchen Gründen nicht mehr weiterbetrieben wird. Die Exportmöglichkeit ist grundlegend, um die langfristige Aufbewahrung zu garantieren. Da normalerweise sowohl Text- als auch Bilddateien zu den Forschungsdaten einer Edition gehören, muss der Export für beide Arten von Dateien möglich sein.

In der Langzeitarchivierung hat sich das Prinzip durchgesetzt, Dateien in Formaten abzuspeichern, die möglichst verbreitet und offen sind. Bei verbreiteten Formaten kann davon ausgegangen werden, dass eine qualitativ hohe Konversion in ein neues gängiges Format möglich sein wird, da sich der Aufwand lohnt, die entsprechende Software zu programmieren. Offen sollen die Formate sein, weil die Spezifikation der Formate öffentlich zugänglich sind, und dadurch neue Software für die Darstellung der Dokumente bzw. für die Konversion in andere Formate geschrieben werden kann.¹¹⁶

Für eine Onlineedition ist dabei die Arbeit mit Sonderzeichen ganz grundlegend. Dafür bietet sich die durchgehende Speicherung in UTF-8 an. Für die Darstellung wird idealerweise die Möglichkeit geboten, via JavaScript einen beliebigen Webfont zu integrieren, damit die Forscherin selber entscheiden kann, welcher Schriftsatz die für ihre Edition nötigen Zeichen darstellen kann.¹¹⁷ Zusätzlich sollte zumindest als Import- und Exportdateiformat TEI-XML unterstützt werden, da sich dieses Format im Editionsreich als Standard etabliert hat. Wie bereits oben ausführlich besprochen, bieten XML die Syntax und TEI das Vokabular. Eine fertige formale Ontologie bieten die Standards jedoch nicht.

Während für grössere Editionsprojekte im Vorfeld eine geeignete Ontologie erstellt wird, ist diese Herangehensweise für eine Klein- oder Teiledition kaum eine Option. Es muss also eine umfangreiche oder mehrere auf spezifische Projekttypen zugeschnittene kleine Standardontologien geben, auf die von diesen Projekten zugegriffen werden kann.¹¹⁸ Dabei ist zwischen zwei Aspekten zu unterscheiden: der gespeicherten und der dargestellten Information. Auch nicht direkt vom System verwendete Tags müssen im System gespeichert werden können, da die Information eventuell

116 Vgl. Toebak 2007, S. 504-516 und Schweizerisches Bundesarchiv 2018. Gleich argumentiert auch Pierazzo 2015, S. 171-173.

117 Z. B. über <http://fonts.com> [29.06.2019].

118 Vgl. Bender 2016, S. 231-232.

zu einem späteren Zeitpunkt oder durch eine andere Applikation benötigt wird.¹¹⁹ Das heisst konkret, dass es in einem integrierten Editor möglich sein muss, Tags manuell einzutippen und nicht nur eine vorgegebene Auswahl zu tätigen.

Die Standardontologie muss hingegen hauptsächlich diejenigen Metadaten definieren, die in der Editionsplattform auch ausgewertet werden. Die genaue Definition dieser Ontologie ist im Grunde für eine Klein- oder Teiledition nicht besonders wichtig, da sowieso mit Einschränkungen der Funktionalität gegenüber der optimalen Lösung zu rechnen ist. Ein Orientierungspunkt kann hier der Standard einer Printedition bieten. Für meine Edition gehören dazu die Auszeichnung folgender Elemente: 1) Archivisch-beschreibende Metadaten zum Werk; im Idealfall wären das zu jedem einzelnen Brief oder juristischen Schreiben Angaben über Format und Beschaffenheit des Originals, Datumsangabe, Angaben zum Siegel und zum Schreiber sowie die Möglichkeit einer Beschreibung des Werks (Regest). 2) Die Auszeichnung von Personen- und Ortsnamen, die eine Basis für das Register bzw. eine Suchfunktion bieten. Zumindest für die Tabelle mit Angaben zu den Hintersassen sollte 3) die Möglichkeit einer Verlinkung von einem Werk zu einem anderen bestehen, da sie sich stark an anderen Teilen der Edition orientiert. Spezifisch für die Briefe sind 4) auch Sender und Empfänger zu annotieren.

Um eine Standardisierung und damit auch Interoperabilität anzustreben, sollte sich diese Standardontologie theoretisch an bereits bestehenden Modellen orientieren. In Punkto TEI gibt es bereits verschiedene leichtere Versionen, die einerseits weniger kompliziert für die Benutzung sind und andererseits auf einen verbesserten Datenaustausch zielen. Grundsätzlich ist das bereits im modularen Ansatz vorgesehen: Meine Bedürfnisse können alle mit den Tags in den vier Standard-Modulen abgedeckt werden. Da diese Auswahl offenbar jedoch nicht auf breiter Basis zufriedenstellend ist, gab es weitere Versuche zur Vereinfachung, z. B. TEI Lite, TEI Tite und TEI Simple, die alle grundsätzlich das gleiche Ziel verfolgen. Das von *Juxta-Editions* unterstützte TEI-Lite soll «90% of the needs of 90% of the TEI user community» abdecken.¹²⁰ Bei TEI Simple wurde in statistischen Verfahren die von Editionsprojekten am häufigsten benutzten Tags ausgewertet.¹²¹ TextGrid und Corpus Corporum haben ihre eigenen minimalen Kodierungsstandards, die eine Interoperabilität über die verschiedenen Projekte bzw. Corpora ermöglichen.¹²² Bei dieser grossen Auswahl an Standardisierungsversuchen kann kaum von einem Standard gesprochen werden, der Interoperabilität fördert, weshalb es schwierig ist, in dieser Hinsicht eine Anforderung zu formulieren.

119 Vgl. Rosselli Del Turco 2016, S. 234.

120 <http://www.tei-c.org/Guidelines/Customization/Lite/> [29.06.2019].

121 Ciotti/Tomasi 2017.

122 Aschenbrenner 2015, S. 205-206 und Telefongespräch mit Philipp Roelli vom 16.05.2018.

Funktionalitäten – weitere Anforderungen

Bereits die Diskussion der Minimalanforderungen hat gezeigt, dass es teilweise schwierig ist, konkrete Lösungen zu finden, weil das Feld der Onlineedition wenig standardisiert ist. Das gilt noch in grösserem Masse für neuere Standards oder für Anforderungen, die nicht unbedingt universal akzeptiert sind. Trotzdem ist klar, dass eine Editionsplattform mehr als nur die Minimalanforderungen erfüllen sollte, um eine bessere Dienstleistung zu bieten. Je mehr Anforderungen abgedeckt werden, desto breiter einsetzbar ist die Software. Die Komplexität der Software darf aber die Grundanforderungen sowie die Benutzerfreundlichkeit nicht überschatten. Auch gibt es Anforderungen, die für eine Klein- oder Teiledition im Gegensatz zu Editionsprojekten nur von geringer Bedeutung sind, wie zum Beispiel das kollaborative Arbeiten sowie die Kollationierung von Texten.

Grundlegend ist zunächst der Import der Daten. In meinem Beispiel liegt der Text der druckreifen Edition als LaTeX-Datei vor, was eher die Ausnahme sein dürfte. Da für jedes importfähige Format eine eigene Schnittstelle gebaut werden muss, können nie alle möglichen Formate in Betracht gezogen werden; es muss eine Auswahl zugunsten häufig benutzter Formate getroffen werden. Für den Text kommen hauptsächlich das Word-File und das TEI-XML in Betracht: das Word-File, weil die Arbeit mit Word unter Geisteswissenschaftlern die Norm ist und dieses Format oft auch von Verlagen verlangt wird; das TEI-XML, weil es das weitest verbreitete Standardformat für digitale Editionen ist. Falls die Editionsplattform ihre Daten mit dem *Presentation-API* von IIF zur Verfügung stellt, soll auch JSON-LD importiert werden können. Da diese Dateiformate weit verbreitet sind, gibt es Transformationstools, die Texte in diese Zielformate konvertieren können: Mit der Open-Source-Konvertierungssoftware TeX4ht ist es mir z. B. möglich, meine LaTeX-Datei in eine Word-Datei oder eine TEI-XML-Datei zu konvertieren.¹²³

Eine Onlineedition bietet den Vorteil, dass zusätzlich zum Text ohne grosse finanzielle Hürden auch ein Faksimile in Form von Digitalisaten veröffentlicht werden kann. Wenn Digitalisate via IIF *Image-API* oder einer anderen API zur Verfügung gestellt werden, können sie ohne weitere Speicherung in eine Editionsplattform integriert werden. Sonst müssen die Bilddateien importiert werden können. Für mein Projekt ist das Bildmaterial grösstenteils im JPEG-Format, da ich die Archivalien selber mit meiner Kamera fotografiert habe. Hochauflösende Scans werden hingegen oft als TIFF oder neuerdings JPEG-2000 abgelegt: Diese beiden Formate werden für die digitale Langzeitarchivierung bevorzugt, weil sie verlustfrei bzw. weitestgehend verlustfrei komprimiert sind. Auch hier gilt, dass diese Formate so weit verbreitet sind, dass quellfreie Möglichkeiten zur Konvertierung einfach auffindbar sind. Es dürfte

123 <https://www.tug.org/tex4ht/> [29.06.2019].

nicht schwierig sein, ein Tool wie *imagemagick*¹²⁴ vor dem Import anzuhängen, um das gewünschte Format zu erhalten, falls auf dem Server nur ein spezifisches Format abgelegt werden sollte.

Für die Transkription von Texten kann die Editionsplattform einen Beitrag leisten, indem sie eine automatische bzw. halbautomatische Texterkennungs-Software, einen auf Transkription und Annotation optimierten Editor oder die Möglichkeit einer kollaborativen Transkription zur Verfügung stellt. Die gängige Texterkennungs-Software liefert momentan nur bei gedruckten Schriften gute Resultate. Juxta Editions bietet jedoch auch schon jetzt als Dienstleistung OCR-Erkennung mit ABBYY FineReader.¹²⁵ Besonders für Editionen moderner oder bereits im Druck erschienener Texte kann mit dieser Software eine gute Grundlage für die Transkription erstellt werden. Für Handschriften kommt zur Zeit hauptsächlich halbautomatische Texterkennungs-Software in Betracht, bei der die Software zuerst trainiert werden muss. Dies ist der Fall beim Projekt *Venice Time Machine*, das eine Software entwickelt, die mehrere Instanzen des gleichen Worts erkennen kann. Die Forscherin muss dann im Idealfall jedes Wort nur noch einmal transkribieren.¹²⁶ Da meine Texte von sehr vielen verschiedenen Händen verfasst worden sind, wird eine solche Software anhand der kleinen Datengrundlage nicht besonders gute Resultate erzielen. Je mehr die Datenbank mit bereits transkribierten Texten gefüllt wird, desto besser wird jedoch das Resultat, weshalb dieses Modell durchaus auch für Kleineditionsprojekte einen interessanten Ansatz bietet.

Wenn die Transkription vollständig von Hand angefertigt wird, kann die Editionsplattform einen entsprechenden Editor zur Verfügung stellen. Ein integriertes System, das auf die von der Plattform unterstützten TEI-Tags optimiert ist und somit besonders benutzerfreundlich gestaltet werden kann, ist sinnvoll. Besonders hilfreich ist Software wie *Transcribo* oder die Annotationsfunktion von *IIIF*, bei der die Texte nicht abgelöst vom Digitalisat transkribiert werden. Das zu transkribierende Wort oder der Textabschnitt wird zuerst im Digitalisat markiert und danach erst wird die Transkription für den markierten Bereich eingegeben. Es kann also jederzeit nachvollzogen werden, welcher Text mit welchem Ausschnitt des Digitalisats übereinstimmt. Bei *Transcribo* geht diese Verbindung jedoch verloren, sobald die Transkription exportiert wird; die Transkription ist in einem proprietären XML abgespeichert, beim Export wird jedoch ein TEI-XML generiert, bei dem diese Verbindung nicht vorgesehen ist.¹²⁷

124 <https://www.imagemagick.org/> [29.06.2019].

125 <http://www.juxtaeditions.com/features> [29.06.2019].

126 Abbott 2017, S. 343. Zur Problematik von automatischer Texterkennung von Manuskripten vgl. Andrews 2014, S. 178-179.

127 <http://transcribo.org/de/ueber-transcribo/technische-standards/> [29.06.2019]. Vgl. Pierazzo 2015, S. 30-31.

Die Annotationsfunktion von IIF geht noch einen Schritt weiter, indem jede Benutzerin der Edition selber weitere Annotationen bzw. Transkriptionen hinzufügen kann. Für eine Teiledition ist dies ein spannender Ansatz, da die Ressource von weiteren Forscherinnen verbessert werden kann. Dies geht auch in das Modell des Crowdsourcing über: Während beim bewussten Outsourcen der Transkriptionsarbeit an interessierte Freiwillige oder Billiglöhnerinnen die abschliessenden Entscheidungen über die korrekte Transkription im Normalfall von einer ausgebildeten Person geleistet wird, bleiben bei der Annotationsfunktion verschiedene Transkriptionen nebeneinander bestehen.¹²⁸

Für Klein- bzw. Teileditionen durchaus sinnvoll ist auch die vollständige Öffnung der Daten für weitere Bearbeitung im Sinne eines Wiki:¹²⁹ Bei meinen juristischen Texten wäre es z. B. denkbar, dass eine andere Forscherin anhand der Digitalisate eine Volledition anfertigt oder die Regesten verbessert. Die einzelne Forscherin würde dadurch zwar Kontrolle über den Inhalt verlieren, die Wissenschaft aber davon profitieren, da die gesamte Arbeit an einem Ort gespeichert wäre. Damit dieses Modell wissenschaftlich haltbar ist, sind eine sinnvolle Versionierung und Rollbackmöglichkeiten unabdingbar.¹³⁰

Neben der Speicherung spielt natürlich auch die Darstellung der Daten eine wichtige Rolle: Peter Robinson argumentiert zwar, dass sie im Grunde genommen nur Kosten verursachen und ein einfaches Interface wie bei Corpus Corporum für die wissenschaftliche Nutzung reichen würde.¹³¹ Wichtiger als die Darstellung sind dabei Erschliessungsmöglichkeiten wie Suchfunktion und/oder Register.

Mit der neuen Technologie von IIF besteht jedoch die Möglichkeit, ohne grossen Aufwand viele Darstellungsmöglichkeiten einer traditionellen gedruckten Edition in einer benutzerfreundlichen Form anzubieten. Im Mirador-Viewer können z. B. die Transkription jedes Digitalisats genau mit dem Digitalisat verknüpft und mehrere Digitalisate des gleichen Texts sowie mehrere Transkriptionen parallel zueinander dargestellt werden.¹³² Für Onlineeditionen wäre nützlich, wenn Unterschiede zwischen den Textzeugnissen visuell hervorgehoben werden könnten.¹³³ Idealerweise werden unterschiedliche diakritische Zeichen oder Doppelkonsonanten nach Wunsch nicht als Textvariationen hervorgehoben, so wie ich es für meine Edition als Transkriptionsrichtlinie festgelegt habe; dadurch werden gewichtigere Variationen besser visualisiert. Diese Funktionalitäten sind nicht innovativ, garantieren jedoch, dass mit der Onlineedition genauso gut gearbeitet werden kann wie mit einer Printedition. IIF bietet

128 Vgl. Pierazzo 2015, S. 21-25 und Snyderman/Sanderson/Cramer 2015, S. 19.

129 OCDE 2007, S. 37.

130 Vgl. Van Zundert 2014, S. 342 und Bender 2016, S. 282.

131 Zusammengefasst bei Pierazzo 2015, S. 167-168.

132 Snyderman/Sanderson/Cramer 2015, S. 20 und Raemy 2017b, S. 34-36.

133 Vgl. Mombert 2017, S. 281-282.

auch den Vorteil, dass die Editionsnutzerin die Wahl hat, welche dieser Darstellungsmöglichkeiten sie sehen will und dass die Edition auch auf leichten Geräten wie einem Tablett oder Smartphone nutzbar ist.¹³⁴

Da bei IIIF die Daten per Definition via API zur Verfügung gestellt werden, bietet diese Form auch ganz grundlegend den Vorteil, dass sie auf unterschiedliche Art ausgewertet bzw. dargestellt werden kann. Peter Robinson plädiert stark dafür, dass alle für Onlineeditionen erstellten Rohdaten nicht nur per API der Forschung zur Verfügung gestellt werden, sondern auch unter einer Creative-Commons-Lizenz, die eine unbeschränkte Weiterverwendung ermöglicht. Die computergesteuerte Weiterverarbeitung von Texten ist einer der grundlegendsten Vorteile gegenüber der Printedition und ist auch im Sinne der Open-Access-Bewegung.¹³⁵

Auch die beste Onlineedition bringt wenig, wenn das Resultat nicht einfach im Internet auffindbar ist. Grundsätzlich ist es für die Sichtbarkeit des eigenen Projekts wünschenswert, mit Metaplattformen wie Europeana, Swissbib oder Archives Portal Europe zu arbeiten. Diese Plattformen sammeln bibliographische Angaben via OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) für ihre Kataloge. Obwohl sie alle mit dem gleichen Transferprotokoll und mit der gleichen Markupsprache (XML) arbeiten, müssen die Metadaten je nach Portal unterschiedlich aufbereitet werden, um vom Portal berücksichtigt zu werden. Dafür kommen eine Reihe von Vokabularen in Frage, z. B. Dublin Core, EAD (*Encoded Archival Description*) oder die verbreitete Katalogisierungssprache MACR21.¹³⁶ Es ist also schwierig, für alle Plattformen die dafür formatierten Daten zur Verfügung zu stellen. Eine Klein- oder Teiledition wird nicht grosse Ansprüche an das Format dieser beschreibenden Metadaten stellen können, da die ideale Metaplattform stark vom angebotenen Inhalt abhängig ist.

Eine Editionsplattform kann dieses Problem umgehen, indem auf Sichtbarkeit bei allgemeinen Suchmaschinen wie Google gesetzt wird. Dafür bietet sich an, zumindest bibliographische Angaben auch mit dem Vokabular von schema.org zu annotieren; diese Ontologie wurde von den grössten Suchmaschinen gemeinsam lanciert und wird von diesen im Suchalgorithmus und in der Darstellungsform der Suchresultate berücksichtigt.¹³⁷ Visibilität der eigenen Ressourcen ist besonders für kleinere Projekte extrem wichtig. Eine gute Auffindbarkeit über Suchmaschinen ist sicherlich

134 Vgl. Bender 2016, S. 222-223, Rosselli Del Turco 2016, S. 235-236 und Pierazzo 2015, S. 137.

135 Vgl. Robinson 2016a und Drott 2007, S. 79-80.

136 Grundlagen bei <https://www.openarchives.org/OAI/2.0/guidelines-repository.htm> [29.06.2019]. Informationen zu den benötigten Daten für Europeana <https://pro.europeana.eu/resources/standardization-tools/edm-documentation> [29.06.2019], Swissbib <http://www.swissbib.org/wiki/index.php?title=Harvesting> [29.06.2019], und Archives Portal Europe http://wiki.archivesportal.europa.net/index.php/OAI_manual_Introduction [29.06.2019].

137 <http://schema.org> [29.06.2019]. Zur Wichtigkeit von Visibilität bei Suchmaschinen vgl. Van Zundert 2014, S. 336.

einer der Hauptgründe, wieso Open-Access-Ressourcen generell mehr und nicht weniger zitiert werden als ihre gedruckten Gegenstücke.¹³⁸

Die Annotation mit dem Vokabular von schema.org geschieht (meistens) anhand von LD-Triples. Dieser Ansatz kann für die Sichtbarkeit und zur gleichzeitigen Tiefenerschliessung der Quellen noch weiter verfolgt werden. Bei einer Verknüpfung auf der Basis von LD stehen nicht das Dokument oder das Werk im Vordergrund, sondern die eigentlichen Inhalte. Wie bei der klassischen Edition, bei der durch Literaturangaben Personen, Orte oder Sachverhalte identifiziert und in den Zusammenhang der Forschungslandschaft gestellt werden, wird bei LD auf URLs bzw. URIs verwiesen, die den gleichen Zweck verfolgen.¹³⁹ Die URL kann auf weitverbreitete Ressourcen wie VIAF, DNB oder Wikidata verweisen, wie es z. B. bei Corpus Corporum umgesetzt ist. Grundsätzlich kann aber auf jede Website verlinkt werden, und im Fall einer URI kann die Verknüpfung sogar ohne dahinterstehende Ressource hergestellt werden. Wenn es sich um Personen handelt, die noch nicht in den Standarddatenbanken vorhanden sind, kann dies ein sinnvolles Vorgehen sein. Andererseits können offene Datenbanken wie Wikidata von allen mit neuen Inhaltsseiten ergänzt werden, was einen Mehrwert gegenüber einer inhaltslosen URI darstellt.

Die Verlinkung als Linked Data bietet in mehrerer Hinsicht einen Mehrwert gegenüber der klassischen Quellenangabe. Erstens ist es möglich, anhand der eindeutig identifizierten Personen einfach ein Personenregister zu erstellen. Es entstehen aber auch neue Möglichkeiten zur Darstellung der Daten, wie z. B. bei der Carl-Maria-von-Weber-Gesamtausgabe, bei der die Korrespondenzbeziehungen der in der Edition erwähnten Personen als facebookartige Profile dargestellt wurden.¹⁴⁰ Ein grundlegender Vorteil des LD-Modells ist jedoch, dass die Verlinkung in beide Richtungen abfragbar ist. Das heisst, jedes mit der VIAF verlinkte Projekt kann theoretisch auch via VIAF gefunden werden. Wenn sich eine Forscherin genau für den Herzog Sigmund von Österreich interessiert, der in meiner Edition eine relativ prominente Rolle spielt, kann sie via LD auch mein Projekt finden.¹⁴¹ Die Verknüpfung auf der inhaltlichen Ebene ermöglicht eine gezieltere Recherche. Solche Vorteile der LD-Technologie werden jedoch bisher erst sehr punktuell für benutzerfreundliche Anwendungen genutzt. So beschränkt sich die Carl-Maria-von-Weber-Gesamtausgabe auch auf diejenigen Daten, die sie selber zur Verfügung stellt. Ein Vorteil einer Editionsplattform mit einer Vielfalt von Projekten kann sein, dass Möglichkeiten geschaffen werden, um LD einfach über die gesamte Plattform hinweg zu benutzen.

138 Pierazzo 2015, S. 131 und Drott 2006, S. 86.

139 Vgl. de la Iglesia/Moretto/Brodhun 2015, S. 101.

140 Stadler 2013, S. 38-39. Die Edition befindet sich unter <https://weber-gesamtausgabe.de/de/Index> [29.06.2019].

141 Dafür brauchen beide Datensets einen SPARQL-Endpoint. Sikos 2015, S. 181.

Eine Onlineresource ermöglicht auch einen ganz anderen Zugang zur Auswertung der Interaktion mit der Edition. Indem die Benutzerinteraktion in einem Log aufgezeichnet wird und mit einem geeigneten (Webstatistik-)Tool ausgewertet wird, können viel mehr Informationen über die Nutzung gesammelt werden, die auch in die Verbesserung der Ressource fließen können.¹⁴² Dies ist keine zwingende Anforderung, aber da sie sehr einfach umsetzbar ist, sollte nicht darauf verzichtet werden.

Fazit

Peter Robinson schrieb 2016: «[I]t should be possible for editing projects (and indeed, single editors) to set up a system to support a wider range of editing tasks, creating data fully adequate for scholarly use, without prodigies of programming and massive grant funding (indeed, without any programming and no grant funding)»¹⁴³ Im Zusammenhang mit Onlineeditionen wird oft begeistert über die Möglichkeiten des Web 2.0 geschrieben. Pichler und Bruvik sehen eine Demokratisierung der Edition rein dadurch, dass sie übers Internet (kostenlos) zur Verfügung gestellt wird. Als weiteren Schritt nennen sie dann die Möglichkeit des User Created Content bzw. Projekte, die mit Crowdsourcing Transkriptionen anfertigen.¹⁴⁴ Zur Frage, inwiefern auch die Auswahl des Inhalts demokratisiert werden kann, fehlt die Debatte grösstenteils. Dementsprechend ist die Bereitstellung von Infrastruktur für Klein- und Teileditionen auch bei der Entwicklung von entsprechender Software nicht im Fokus der Hersteller oder Betreiber.

Wenn die Onlineedition als Teil der Diskussion und aktuellen Entwicklung im Bereich der Aufbewahrung und des Zugänglichmachens von Forschungsdaten betrachtet wird, ist die Frage zur Herkunft des Inhalts von zentraler Bedeutung: Solange es keine Editionsplattform gibt, auf der Forscherinnen unabhängig von ihrer Zugehörigkeit zu einer bestimmten Institution Inhalte hochladen können, bietet die Onlineedition nur ansatzweise eine Möglichkeit für die Archivierung von Forschungsdaten. Die vorliegende Arbeit ist der Frage nachgegangen, welche Grundlagen geschaffen werden müssen, damit die Onlineedition für eine breitere Gruppe von Wissenschaftlerinnen oder wissenschaftlich interessierten Personen und Institutionen eine interessante und umsetzbare Alternative zur gedruckten Edition wird.

Als Fazit kann festgestellt werden, dass es einerseits bereits einige Software gibt, die für eine Klein- oder Teiledition eine spannende Ausgangslage bieten. Besonders positiv ist, dass ein grosser Teil davon quellfrei ist und die Projekte so gegenseitig voneinander profitieren können. Die geplante Schnittstelle zwischen FuD und NIE-

142 Vgl. Rosselli Del Turco 2016, S. 224-245.

143 Robertson 2016b. Vgl. auch Rosselli Del Turco 2016, S. 232-234.

144 Pichler/Bruvik 2017, S. 224.

INE ist in dieser Hinsicht sicherlich ein Highlight. Auch Corpus Corporum ist ein Vorzeigeprojekt, da die Plattform Editionen unabhängig vom Wohnsitz oder Bezug zu einer Hochschule annimmt.

Gleichzeitig wurde jedoch auch schnell klar, dass von den bestehenden Lösungen keine eine ideale Umgebung für ein Klein- bzw. Teileditionsprojekt bietet. Viele Editionsprojekte entwickeln noch immer ihre eigenen, massgeschneiderten Plattformen, die sie nur mit den von ihnen bearbeiteten Texten füllen. Tara Andrews fordert in diesem Zusammenhang das erhöhte Engagement von Verlagen bei der Bereitstellung von Editionsplattformen, einerseits, um dem verbreiteten Selfhosting und Selfpublishing ein Ende zu setzen, aber andererseits auch, weil Verlage Erfahrung in der Verbreitung und Sichtbarmachung wissenschaftlicher Ressourcen mitbringen.¹⁴⁵ Gleichermassen wäre es auch wünschenswert, wenn Editionsprojekte wie die Sammlung Schweizerischer Rechtsquellen¹⁴⁶ ihre Infrastruktur für thematisch verwandte Kleineditionen bereitstellen würden.

Obwohl die bestehenden Editionsplattformen diesen Problemen entgegenwirken wollen, gibt es noch grosse Hürden: Bei der von mir untersuchten Software waren die Hauptprobleme die thematische Eingrenzung (Corpus Corporum), Kostspieligkeit (FuD), nötiger Entwicklungsaufwand für die Forscherin (NIE-INE) oder keine garantierte Verfügbarkeit (Juxta Editions). Diese Probleme zeigen bereits einige der wichtigsten Anforderungen an eine Editionsplattform: Nicht nur muss sie für externe Forscherinnen zur Verfügung stehen, sondern sie soll auch eine möglichst grosse thematische Vielfalt beherbergen können. Kosten und Entwicklungsaufwand sollten sich für die Forscherin in Grenzen halten. Wenn die Plattform nicht stabil und langfristig zur Verfügung steht, kann nicht von einer echten Alternative zur Printedition gesprochen werden. Entwicklungsbedarf besteht im Bereich der Benutzerfreundlichkeit, nicht nur bei der Bedienung der Software, sondern auch bei den noch mangelnden Cloudangeboten. Auch die langfristige Benutzbarkeit der Daten nach dem FAIR-Prinzip ist ein Problem, das sich erst langsam ins Bewusstsein der Anbieterinnen drängt.

Die Diskussion mit Verantwortlichen von Projekten wie FuD und NIE-INE hat gezeigt, dass der Bedarf für eine Editionsplattform, die mit wenigen technischen und finanziellen Mitteln benutzt werden kann, durchaus bekannt ist. In beiden Projekten sind auch bereits Überlegungen für eine Weiterentwicklung in diese Richtung vorhanden, sie werden jedoch nicht prioritär behandelt. Eine Erklärung dafür ist, dass bei beiden Projekten die Software nach den Bedürfnissen der mit ihnen zusammenarbeitenden Editionsprojekte entwickelt wird. Wegen der aktuellen technischen und finanziellen Hürden sind dabei Klein- und Teileditionsprojekte nicht vertreten bzw. bei FuD nur im Rahmen der Trier Universitätsmitglieder, die das FuD-Repository

145 Andrews 2014, S. 192.

146 <https://www.ssrq-sds-fds.ch/home/> [29.06.2019].

benutzen dürfen. Entsprechend gross war das Interesse an meiner Arbeit und den von mir beschriebenen Anforderungen, da sie eine für sie wenig bekannte Dimension aufzeigen. Besonders im Zusammenhang mit der intensivierten Diskussion über die Speicherung von Forschungsdaten ist zu hoffen, dass die Anforderungen auf Gehör stossen.

Bibliographie

Abbott, Alison: The ‘time machine’ reconstructing ancient Venice’s social networks. In: *Nature* 546 (2017), S. 341-344. <https://www.nature.com/news/the-time-machine-reconstructing-ancient-venice-s-social-networks-1.22147> [29.06.2019].

Ackermann, Christa: Drei Ritter auf einem Vulkan. Eine Mikrogeschichte über eine spätmittelalterliche Niederadel-Korrespondenz aus dem deutschen Südwesten. Doktorarbeit Universität Jena 2014. DOI: <https://doi.org/10.22032/dbt.37644>

Aliprand, Joan M.: The Unicode™ Standard. An overview with emphasis on bidirectionality. In: Byrum, John D. Jr.; Madison, Olivia (Hgg): *Multi-Script, Multilingual, Multi-character Issues for the Online Environment*. München 1998, S. 95-111.

Andrews, Tara: Digital Techniques for Critical Edition. In: Calzolari, Valentina, *Armenian Philology in the Modern Era. From Manuscript to Digital Text*. 2014 Leiden/Boston, S. 175-198.

Andrews, Tara; Macé, Caroline (Hgg): *Analysis of Ancient and Medieval Texts and Manuscripts. Digital Approaches*. Turnhout 2014.

Apollon, Daniel; Régnier Philippe; Bélisle, Claire: Alors que les textes deviennent numériques. In: Apollon/Régnier/Bélisle (Hgg.) 2017, S. 9-51.

Apollon, Daniel, Régnier Philippe; Bélisle, Claire (Hgg.): *L’Edition critique à l’ère du numérique*. Paris 2017.

Aschenbrenner, Andreas: Share It. Kollaboratives Arbeiten in TextGrid. In: Neuroth/Rapp/Sörig (Hgg.) 2015, S. 201-210.

Bender, Michael: *Forschungsumgebungen in den Digital Humanities. Nutzerbedarf – Wissenstransfer – Textualität*. Berlin/Boston 2016.

Brumfield, Ben: Text Beyond Annotations at IIF-Vatican, 18.06.2017. <https://content.fromthepage.com/text-beyond-annotations-at-iiif-vatican/> [29.06.2019].

Ciotti, Fabio; Tomasi Francesca: Formal Ontologies, Linked Data, and TEI Semantics. In: *Journal of the Text Encoding Initiative* 9 (2017). <http://journals.openedition.org/jtei/1480> [29.06.2019].

de la Iglesia, Martin; Moretto, Nicolas; Brodhun, Maximilian: Metadaten, LOD und der Mehrwert standardisierter und vernetzter Daten. In: Neuroth/Rapp/Sörig (Hgg.) 2015, S. 91-102.

Dickel, Sascha; Franzen, Martina: Digitale Inklusion. Zur sozialen Öffnung des Wissenschaftssystems. In: Zeitschrift für Soziologie 44, 5 (2015), S. 330–347.

Drott, M. Carl: Open Access. Philadelphia 2006.

Franke, Michael: Erfolgsfaktoren für virtuelle Forschungsumgebungen. In: Seng, Eva-Maria; Keil, Reinhard; Oevel, Gudrun (Hgg): Studiolo. Kooperative Forschungsumgebungen in den eHumanities. Berlin/Boston 2018, S. 57-65.

FuD: Leistungsbeschreibung Version 1.5 (Mai 2018). https://fud.uni-trier.de/wp-content/uploads/2018/06/FuD_Leistungsbeschreibung-1.5.pdf [29.06.2019].

Gillam, Richard: Unicode Demystified. A Practical Programmer's Guide to the Encoding Standard. Boston 2003.

Hörschemeyer, Jörg: Textgenetische Prozesse in digitalen Editionen. Doktorarbeit Universität Köln 2013. URN: <http://nbn-resolving.de/urn:nbn:de:hbz:38-75446>

Huitfeld, Claus: Système de balisage de textes et éditions critiques. In: Apollon/Régnier/Bélisle (Hgg.) 2017, S. 179-202.

Kaden, Ben; Rieger, Simone: Usability in Forschungsinfrastrukturen für die Geisteswissenschaften. Erfahrungen und Einsichten aus TextGrid III. In: Neuroth/Rapp/Sörig (Hgg.) 2015, S. 63-75.

Kurz, Susanne: Digital Humanities. Grundlagen und Technologien für die Praxis. Wiesbaden 2017.

Meroni, Silvia: Définition d'une méthode de suivi des coûts de la publication en Open Access pour l'Université de Genève. Masterarbeit Haute Ecole de Gestion de Genève 15. August 2016. <http://doc.rero.ch/record/277983?ln=fr> [29.06.2019].

Minn, Gisela et al.: FuD2015 – Eine virtuelle Forschungsumgebung für die Geistes- und Sozialwissenschaften auf dem Weg in den Regelbetrieb. Akteure, Arbeitsfelder, Organisations- und Finanzierungsstrukturen. Trier Oktober 2016. http://ubt.opus.hbz-nrw.de/volltexte/2016/1010/pdf/WP_Nr_01_FuD_GeschAftsmo-dell.pdf [29.06.2019].

Mombert, Sarah: Du livre à la collection. Éditions critiques de documents hétérogènes. In: Apollon/Régnier/Bélisle (Hgg.) 2017, S. 271-289.

Neuroth, Heike; Rapp, Andrea; Sörig, Sibylle (Hgg): TextGrid: Von der Community – für die Community. Eine Virtuelle Forschungsumgebung für die Geisteswissenschaften. Glückstadt 2015. DOI: <http://dx.doi.org/10.3249/webdoc-3947>

NIE-INE: Factsheet, 23.05.2018: https://docs.wixstatic.com/ugd/1b2bb4_7dfcda0ac013462597daa3d6a7345af8.pdf [29.06.2019].

OCDE (Organisation for Economic Co-operation and Development): Participative web and user-created content. Web 2.0, Wikis and social networking. Paris 2004.

Oßwald, Achim; Scheffel, Regine; Neuroth, Heike: Langzeitarchivierung von Forschungsdaten, Einführende Überlegungen. In: Neuroth, Heike et al. (Hgg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg 2012, S. 13-22.

Pichler, Alois; Bruvik, Tone Merete: Éditions critiques et séparation de la description et de la présentation. In: Apollon/Régnier/Bélisle (Hgg) 2017, S. 203-224.

Pierazzo, Elena: Digital Scholarly Editing. Theories, Models and Methods. Oxon/New York 2016.

Raemy, Julien: IIF ou comment enfin valoriser ses images numériques de manière optimale et interopérable. In: Hors-Texte 115 (2018), S. 22-32.

Raemy, Julien: Interopérabilité des images. De la nécessité des tests d'utilisabilité. In: Ressi 2017a. http://www.ressi.ch/num18/article_142 [21.07.2018].

Raemy, Julien: The International Image Interoperability Framework (IIIF). Raising awareness for the user benefits for scholarly editions. Bachelorarbeit Haute Ecole de Gestion de Genève, 16. Juli 2017b.

Robinson, Peter: Current issues in making digital editions of medieval texts. Or, do electronic scholarly editions have a future? In: Digital Medievalist 1 (2005). DOI: <http://doi.org/10.16995/dm.8>

Robinson Peter: Project-based digital humanities and social, digital, and scholarly editions. In: Digital Scholarship in the Humanities. Oxford 2016a. DOI: <http://doi.org/10.1093/llc/fqw020>

Robinson, Peter: Social editions, social editing, social texts. In: Digital Studies/Le champ numérique 6 (2016b). DOI: <http://doi.org/10.16995/dscn.6>

Roelli, Philipp: The Corpus Corporum, a new open Latin text repository and tool. In: Archivvm 72 (2014), S. 289-304.

Rosselli Del Turco, Roberto: The Battle We Forgot to Fight. Should We Make a Case for Digital Editions? In: Driscoll, Matthew James; Pierazzo, Elena (Hgg): Digital Scholarly Editing. Theories and Practices. Cambridge 2016, S. 219-238. DOI: <http://doi.org/10.11647/OBP.0095>

Sahle, Patrick: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels, 3 Bände. Norderstedt 2013.

Schweizerisches Bundesarchiv: Archivtaugliche Formate. April 2018. https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte_und_weisungen/archivtaugliche_dateiformate.1.pdf.download.pdf/archivtaugliche_dateiformate.pdf [29.06.2019].

Sikos, Leslie F.: Mastering Structured Data on the Semantic Web. From HTML5 Microdata to Linked Open Data. New York 2015.

Snydman, Stuart; Sanderson, Robert; Cramer, Tom: The International Image Interoperability Framework (IIIF). A community & technology approach for web-based images. In: Archiving Conference. Los Angeles, Mai 2015, S. 16-21. PURL: <https://purl.stanford.edu/df650pk4327>

Stadler, Peter: Die Grenzen meiner Textverarbeitung bedeuten die Grenzen meiner Edition. In: editio 35 (2013), S. 31-40.

TEI-Consortium: TEI P5. Guidelines for Electronic Text Encoding and Interchange. Version 3.3.0, 31.01.2018. <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> [17.06.2018].

Toebak, Peter M.: Records Management. Ein Handbuch. Baden 2007.

Tupman, Charlotte; Jordanous, Anna: Sharing Ancient Wisdoms across the Semantic Web Using TEI and Ontologies. In: Andrews/Macé (Hgg) 2014, S. 213-228.

Van Zundert, Joris J: By Way of Conclusion. Truly Scholarly, Digital, and Innovative Editions? In: Andrews/Macé (Hgg) 2014, S. 335-346.

Viglianti, Raffaele: Keep 'em Separated. Integrating TEI and IIIF without loss (paper). 13.11.2013. https://hcmc.uvic.ca/tei2017/abstracts/t_125_viglianti_tei-andiiif.html [29.06.2019].

Wegstein, Werner; Rapp, Andrea; Jannidis, Fotis: Textgrid – Eine Geschichte. In: Neuroth/Rapp/Sörig (Hgg.) 2015, S. 23-35.

Wilkinson et al.: The FAIR Guiding Principles for scientific data management and stewardship. In: Scientific Data 3 (2016). DOI: <http://doi.org/10.1038/sdata.2016.18>

Winkler-Nees: Stand der Diskussion und Aktivitäten. National. In: Neuroth, Heike et al. (Hgg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg 2012, S. 23-40.

Anhang: Anforderungskatalog

	Anforderung	Priorität (kann / muss)
0	Allgemeines	
0.01	Für Privatpersonen die Möglichkeit, Quellen hochzuladen	muss
0.02	Intuitive Benutzungsoberfläche des Backends	muss
0.03	Gute Dokumentation für die Benutzung des Backends	muss

0.04	Kontaktmöglichkeit für Hilfe zur Problembhebung	muss
0.05	Keine wiederkehrenden Gebühren	muss
0.06	Modulare Architektur	kann
0.07	Open Source	kann
1	Textverarbeitung	
1.01	Durchgehende UTF-8-Kompatibilität	muss
1.02	Import von Textgrundlagen (LaTeX, Word, TEI-XML, copy/paste)	muss
1.03	Halb- oder vollautomatische Texterkennung	kann
1.04	Integriertes Tool zur inhaltlichen Annotation nach TEI	kann
1.05	Integriertes Tool zur internen und externen Verlinkung nach dem Prinzip von LOD	kann
1.06	Kollaboratives Arbeiten an den Texten	kann
1.07	Transkription via Crowdsourcing	kann
1.08	Kollationierung und Stematisierung	kann
2	Speicherung	
2.01	Cloud-Angebot (Webserver wird zur Verfügung gestellt)	muss
2.02	Periodische Migration auf neuere Server-Technologie	muss
2.03	Dauerhafte Adressierung (z. B. DOI)	muss
2.04	Hohe Verfügbarkeit des Servers	muss
2.05	Exportmöglichkeiten des Projekts in ein nicht-proprietäres Format ohne Datenverlust	muss
2.06	Indizierung für Volltextsuche	muss
2.07	Speicherung von beschreibenden Metadaten	muss
2.08	Speicherung von TEI (inklusive Standardontologie, die von der Software ausgewertet wird)	muss
2.09	Speicherung von Bilddateien	muss
2.10	Speicherung von LOD	kann
2.11	IIIF-Kompatibilität	kann

2.12	OAI-PMH Schnittstelle	kann
2.13	Search Engine Optimization (z. B. Annotierung mit schema.org)	kann
2.14	Auswertung von Benutzerlogs via Webstatistik	kann
3	Publikation	
3.01	Übersichtliches und intuitives Webfrontend	muss
3.02	Darstellung von Digitalisaten (Bilddateien)	muss
3.03	Einbindung von Webfonts (z. B. für diakritische Zeichen)	muss
3.04	Einblenden, Ausblenden und Vergleichen von verschiedenen Textversionen	muss
3.05	Periodische Anpassung der Darstellungstechnologie an die neusten gängigen Browser	muss
3.06	Einbindung von Texten zum Verständnis der Edition (Editionsrichtlinien, Auswertungen, Beschreibung, ...)	muss
3.07	Angaben zu Lizenzen für die Weiterbenutzung	muss
3.08	Standard-Design, das ohne grossen Aufwand benutzt werden kann	muss
3.09	Anpassung des Designs via CSS ist möglich	kann
3.10	Responsive Design	kann
3.11	Unterscheidung Transkription / normalisierte Textversion	kann
3.12	Einblenden, Ausblenden und Vergleichen von Text- und Bilddarstellung	kann
3.13	Exportmöglichkeiten für Text und Bild via Webfrontend	kann
3.14	API für Weiterbearbeitung von Text und Bild	kann
3.15	Möglichkeit der Annotierung / Bearbeitung durch Benutzerinnen (User Created Content)	kann