

Data Cleaning als digitale Quellenkritik

VD17 und das Genre der katholischen Dissertation im Alten Reich

Heßbrüggen-Walter, Stefan

early.modern.thought.online@gmail.com
HSE University, Russian Federation

Die forschende Nachnutzung von Metadaten analoger und digitaler Artefakte ist ein wesentlicher Bestandteil der Arbeit am kulturellen Gedächtnis. Die digitalen Geisteswissenschaften, insbesondere die digitale Literaturwissenschaft, haben hierzu schon beachtliche Beiträge geleistet, insbesondere in der Betrachtung der Geschichte literarischer Gattungen (Fischer 2018, Gittel 2021). Auch in der Wissenschafts- und Geistesgeschichte ist die Relevanz von Metadaten mittlerweile erkannt worden (Sangiaco 2020, Scholz 2021). Historische Arbeit an und mit Metadaten, so die erste hier zu verteidigende Teilthese, erfordert aber, wie die Bearbeitung jeder anderer digitalen Quelle auch, „Quellenkritik“ (entsprechende Überlegungen in den Geschichtswissenschaften konzentrieren sich eher auf Archivalien und digital entstandenes Material, siehe Friederich 2018). Denn anders als mancher Kritiker meint, bleibt es Utopie, „in der elektronisch schaltbaren Präsenz des Wissens zugleich seine absolute Verfügbarkeit“ herzustellen (Jochum 1998, 29f). Metadaten des kulturellen Erbes sind vielmehr menschengemachte und deswegen außerordentlich fragile Gebilde.

Dies gilt erst recht, wenn mehrere Institutionen über einen längeren Zeitraum einen komplexen Bestand an Schrifttum zu verzeichnen haben. Meine Fallstudie betrifft einen zahlenmäßig unbedeutenden Ausschnitt der deutschen Nationalbibliografie für Drucke des 17. Jahrhunderts, VD17 (Anonym 2020), nämlich Metadaten zu philosophischen Dissertationen, die im Erfassungszeitraum von VD17 (1601-1700) an katholischen Institutionen angefertigt wurden.¹

Ich hoffe zu zeigen, und dies ist meine zweite Teilthese, dass die angesprochene Quellenkritik der Metadaten solcher Dissertationen bereits auf der Ebene der „Arbeitsvorbereitung“ stattzufinden hat, also in jenem Projektschritt, den wir in den digitalen Geisteswissenschaften englisch als *data cleaning* bezeichnen. Ein deutscher Ausdruck hat sich hierfür bislang nicht eingebürgert (CrankyPhilosopher 2021). Kurz könnte man beide Teilthesen auch so zusammenfassen: *Data cleaning* ermöglicht nicht nur Forschung, *data cleaning* erfordert auch Forschung oder kann selbst zu Forschungseinsichten führen.

Es wird gelegentlich die Auffassung vertreten, dass die Produzenten von Metadaten, im vorliegenden Fall also Bibliotheken, selbst dafür Sorge zu tragen hätten, dass ihre Aufnahmen den Anforderungen genügen, die von Forschenden v. a. in den Digital Humanities an solche Datensätze gestellt werden (Király / Brase 2021, 358-359). Inwiefern ein solcher Anspruch realistisch ist, wäre wohl zuerst aus bibliothekswissenschaftlicher Sicht zu klären. Wenn man ihn erhebt, muss man sich jedoch im Lichte des folgenden darüber im Klaren sein, dass dann Erschließungsleistungen in der Katalogisierung bspw. Alter Drucke selbst schon Forschungsleistungen darstellen, diese nicht bloß ermöglichen.

Die hauptsächliche Forschungsfrage, um die es im folgenden geht, lautet: wie viele katholische philosophische Dissertationen sind in VD17 verzeichnet? Diese auf den ersten Blick vielleicht nicht besonders dringlich erscheinende Frage gilt es zu beantworten, wenn man beispielsweise die Relevanz des Genres Dissertation im katholischen Raum in Beziehung setzen will zur Rolle der Gattung in der protestantischen Universität. Drei Anmerkungen sind zum Verständnis ihrer Voraussetzungen hilfreich. Erstens sind die „örtlichen Kriterien“ für VD17, also das von der Bibliografie abzudeckende Territorium notorisch vage (Stäcker 2004, 214f). Zweitens haben nicht alle deutschen Bibliotheken an diesem Projekt mitgewirkt, es steht also anzunehmen, dass nicht alle zwischen 1601 und 1700 in diesem vagen Raum erschienenen katholischen Dissertationen in der Bibliografie verzeichnet sind. Drittens zählt VD17 kleinere Abweichungen im Druck als eigenständige Editionen, so dass das Zählen von „Werken“ („geistigen Gegenständen“, Heßbrüggen-Walter 2020) hier eigene Herausforderungen birgt, um die es zunächst noch nicht gehen soll. In operationalisierbarer Form würde unsere Forschungsfrage also lauten: wie viele VD17-Zitiernummern beziehen sich auf eine an einer katholischen Bildungseinrichtung angefertigte Dissertation? Anhand eines kurzen Beispiels werde ich zudem erläutern, welche Herausforderungen bei der schlüssigen Erfassung von mit Dissertationen verbundenen Körperschaften zu bewältigen sind.

Man kann sich berechtigterweise auf den Standpunkt stellen, dass die Beantwortung der Frage nach der Zahl katholischer Dissertationen in VD17 mit *data cleaning* alleine nicht zu leisten ist, selbst wenn die auf diesen Arbeitsschritt folgende Auswertung der aufbereiteten Daten in einer Zeile Programmcode bestehen würde. Das stimmt auch. Um jedoch die Daten in einer Form zur Verfügung zu stellen, die eine solche Auswertung möglich macht, sind weitere unter das *data cleaning* zu subsumierende Arbeitsschritte notwendig, die ihrerseits, wie nun zu zeigen ist, Forschungsleistungen darstellen bzw. voraussetzen.

Zunächst kurz zum Begriff der Dissertation und der Relevanz dieses Genres nicht nur für die philosophiegeschichtliche Forschung im engeren Sinne, sondern auch für die Erforschung frühneuzeitlicher Bildungsgeschichte im weitesten Sinne. Wesentliches Merkmal dieser Texte ist ihre Bindung an eine Einrichtung höherer Bildung – neben Universitäten auch akademische Gymnasien ohne Promotionsrecht oder an Klöster gebundene Ordensstudien. Eine genauere Erfassung dieser Dissertationen und der mit ihnen verbundenen Personen und Institutionen ist nicht nur eine Vorbedingung ihrer in weiten Teilen noch ausstehenden philosophiegeschichtlichen Erschließung (siehe jedoch für Dillingen Leinsle 2006), sondern schon allein auf der Ebene der Prosopographie von Belang. Denn nicht nur haben Geistliche und Mönche an katholischen Universitäten studiert, auch die Ordensstudien selbst haben Dissertationen hervorgebracht. Damit sind Dissertationen wertvolle Datenquellen für die Identifizierung konkreter Individuen in monastischen Gemeinschaften und für deren Praxis philosophischen Unterrichts.

Philosophische Dissertationen sind als Genre in den AAD *Gattungsgenres* (Anonym o. J.) enthalten, die für die Formalerschließung in VD17 zugrundegelegt wurden (Anonym 2020). Daneben enthalten die *Gattungsgenres* auch die Kategorie „Ordensliteratur“ mit 22 Unterkategorien. Da die akademische Lehre im katholischen Deutschland des 17. Jahrhunderts Ordensgemeinschaften und den mit diesen verbundenen Institutionen oblag, kann man zunächst davon ausgehen, dass katholische philosophische Dissertationen in der Schnittmenge beider Kategorien zu finden sind.

Rahmenbedingungen

VD17-Daten wurden über die zur Verfügung gestellte SRU-Schnittstelle im MODS-Format abgefragt und in nach Suchbegriff unterschiedenen XML-Dateien gespeichert. Für unsere Analyse sind zwei im Rahmen von MODS als Namen bzw. Namens-Ids erfasste Felder in den Katalogisaten von Belang: ‚name-corporate‘ mit dem Attribut ‚oth‘ erfasst die für die Dissertation verantwortliche Bildungseinrichtung. Und das sogenannte *statement of responsibility* erfasst weitere für Dissertationen einschlägige Informationen. Die beschriebenen Tags wurden mit Hilfe eines Jupyter-Notebooks ausgelesen und weiterverarbeitet.²

Data Cleaning I: protestantische Dissertationen über katholische Orden

Die Hoffnung, dass die Schnittmenge der Kategorien ‚philosophische Ordensliteratur‘ und ‚philosophische Dissertation‘ alleine Arbeiten enthält, die an katholischen Institutionen entstanden sind, erfüllte sich nicht. Von den 1245 Titeln, die beide Kriterien erfüllen, sind vielmehr neun an protestantischen Institutionen, nämlich den Universitäten Wittenberg, Leipzig, Jena, Halle und Gießen sowie den Gymnasien in Ulm und Bayreuth entstanden. Es handelt sich bei diesen Arbeiten allerdings nicht um Beiträge katholischer Gastprofessoren, sondern vielmehr um Arbeiten, die aus protestantischer Sicht über katholische Orden verfasst worden sind. Sie behandeln zum Beispiel die *reservatio mentalis* der Jesuiten oder das Leben des Heiligen Norbert von Xanten, Erzbischof von Magdeburg und Stifter des Prämostratenserordens. Damit verbleiben 1236 katholische Dissertationen. Die Kategorie ‚Ordensliteratur‘ ist also nicht eindeutig bestimmt: es kann sich hier sowohl um Literatur handeln, die aus einem Orden hervorgeht, als auch um Literatur, die über einen Orden oder dessen Mitglieder verfasst wurde.

Data Cleaning II: Augsburg

VD17 enthält jedoch auch 30 Datensätze zu Dissertationen katholischer Bildungseinrichtungen, bei denen die Klassifikation als Ordensliteratur fehlt. Diese Datensätze wurden identifiziert, indem alle mit einer als katholisch ausgewiesenen Institution in Zusammenhang stehenden philosophischen Dissertationen in einem zweiten Anlauf darauf hin durchsucht wurden, ob für sie auch nicht als Ordensliteratur ausgezeichnete Titel in VD17 verzeichnet sind.

Dazu mussten in den im ersten Arbeitsschritt die in den schon erfassten 1235 Datensätzen verzeichneten Bildungsinstitutionen eindeutig identifiziert werden. Körperschaften werden in VD17, anders als ein Teil der Autoren, jedoch nicht mit GND-Identifikatoren versehen und haben auch keine vereinheitlichte Ansetzungsform. Hier mussten also Daten bereinigt und ergänzt werden. Dies warf besondere Schwierigkeiten auf, wenn in einer Stadt sowohl eine katholische wie eine protestantische Bildungseinrichtung vorhanden sind. Dies soll am Beispiel Augsburgs erklärt werden.

Für das ‚Jesuitenkolleg (Augsburg)‘ (GND 4222329-5) enthält VD17 insgesamt 14 Ansetzungsformen für 37 Dissertatio-

nen. Acht dieser Ansetzungsformen sind eindeutig identifizierbar, denn sie enthalten das Patrozinium der Lehranstalt oder nehmen eindeutig Bezug auf ihren Träger, den Jesuitenorden: Lyceum S. Salvatoris Augsburg, Lyceum S. Salvatoris Augustae Vindel, S. Salvatoris Lyceum Augustanum, Lyceum S. Salvatoris Augustanum, S. Salvatoris Lyceum Augustanum [sic!], Gymnasium S. Salvatoris Augsburg, Gymnasium ad S. Salvatoris Augsburg, Gymnasium ad S. Salvatoris. Sie erfassen 56,8% der an dieser Institution verfassten und in VD17 überlieferten Dissertationen.

Neben dem dem Erlöser (*salvator*) gewidmeten Jesuitenkolleg existierte in Augsburg im 17. Jahrhundert jedoch auch ein protestantisches Gymnasium. Die Ansetzungsformen ‚Gymnasium Augsburg‘ und ‚Lyceum Augustanum‘ sind also nicht eindeutig. Das protestantische Gymnasium der Stadt befand sich zu seiner Gründung in den Räumlichkeiten eines vor der Reformation der Hl. Anna geweihten Klosters und leitete daraus seinen Namen ab (GND 2012843-5). Jedoch finden sich auch von Jesuiten veranstaltete Dissertationen, deren Ansetzung in VD17 ebenfalls auf ein der Hl. Anna geweihtes Kloster Bezug nehmen, und zwar sechs: Sankt Anna, Augsburg, Augustae Vindelicorum ad S. Annam, Kloster St. Anna, Augsburg, S. Anna, Augsburg, Gymnasium St. Anna, Augsburg, Collegium Philosophicum ad S. Annam, Augsburg. Der dem Kloster gewidmete Wikipedia-Artikel gibt näheren Aufschluss: das protestantische Gymnasium bei St. Anna bezog schon 1613 ein neues Gebäude, behielt aber den alten Namen bei. Zugleich zog das dem Erlöser gewidmete Jesuitenkolleg 1635 in die Räumlichkeiten des Klosters St. Anna (Wikipedia 2021) und nutzte also anscheinend dessen Bezeichnung auch für die eigene Institution. Diese streng genommen irreführenden Zuschreibungen machen 43,2% der jesuitischen Dissertationen Augsburgs aus. Die Dissertation mit der VD17- Zitiernummer 23:241965E ist nur als am – ja streng genommen nicht existenten ‚Gymnasium Augsburg‘ entstandene - ‚Dissertation:phil.‘ ausgezeichnet. Ohne die beschriebenen Klärungen hätte sie damit nicht als katholische Dissertation identifiziert werden können.

Data Cleaning III: Keine Institution

In 68 Katalogisaten von Dissertationen finden wir keine Angabe einer Institution, die für die Dissertation verantwortlich zeichnen würde, obwohl die Titel sowohl als Ordensliteratur wie auch als philosophische Dissertation klassifiziert worden sind. In einem Fall handelt es sich tatsächlich nicht um eine Dissertation im formalen Sinne einer Hochschulschrift (VD17 14:697769A). In sieben Fällen ist auf dem Titelblatt zwar der Dissertationscharakter ersichtlich, aber keine verantwortliche Institution angegeben. Sie könnte vermutlich unter Zuhilfenahme von Druckort und Autor erschlossen werden, doch der Status solcher erschlossener Metadaten wäre eigens zu bedenken und liegt außerhalb der Zielsetzung dieses Papers. Es verbleiben 60 Katalogeinträge, für die zwar keine Institution eingetragen ist, diese aber aus den vorliegenden Metadaten maschinell oder durch Inspektion bspw. der Schlüsselseiten festgestellt werden können. Wesentliches Hilfsmittel war hier das bereits erwähnte MODS-Tag *statement of responsibility*, das die bei der Katalogisierung für wesentlich erachteten Elemente des Untertitels der Dissertation, darunter zumindest gelegentlich eben auch die verantwortliche Körperschaft, erfasst. Die Suche nach 26 Zeichenketten, die jeweils eine Körperschaft eindeutig identifizieren, erlaubte die Klärung der institutionellen Verantwortlichkeit von 46 der 60 ungeklärten Dissertationen. Für die weiteren 14 Titel war eine manuelle Inspektion der Schlüsselseiten erforderlich. Es erweist sich in diesem Zusammenhang als misslich, dass in VD17, anders als noch in VD16, bewusst auf die

vollständige, aber eben „umständliche“ (Stäcker 2004, 214) Erfassung der Titelseiten verzichtet wurde.

Zusammenfassung und Ausblick

Die im Rahmen des *data cleaning* der Katalogisate katholischer Dissertationen erzielten quellenkritischen Forschungsergebnisse betreffen zunächst den im VD17 verwendeten Gattungsbegriff „Ordensliteratur“: es ist nicht deutlich, ob dieser nur die literarischen Hervorbringungen von Ordensgemeinschaften oder auch Schriften über Ordensgemeinschaften erfassen soll. Der inklusive Gebrauch der Kategorie in VD17 ist nirgends dokumentiert und für den unvoreingenommenen Betrachter auch nicht sofort offensichtlich. Die Ansetzung von Institutionen schließt nicht den Gebrauch von Normdaten ein und ist somit zunächst nicht nachvollziehbar und unzuverlässig. Zwar hatte dies zumindest in unserem Fall keinen Einfluss auf die Zählung katholischer Dissertationen, wirkt aber dennoch auf den unbefangenen Nutzer irreführend.

Eingangs war die Frage aufgeworfen worden, wieviele katholische Dissertationen in VD17 verzeichnet sind. In der Kombination der Gattungsbegriffe für Ordensliteratur und philosophische Dissertation wurden zunächst 1245 einschlägige Katalogisate identifiziert. Davon waren neun jedoch protestantischen Institutionen zuzuordnen. Diesen 1236 Dissertationen sind 30 Titel zuzuschlagen, die zwar an einer katholischen Institution entstanden sind, aber nicht als Ordensliteratur ausgezeichnet wurden. Ein Titel, der als Dissertation erfasst wurde, aber keine Hochschulschrift darstellt, ist abzuziehen. Damit kommen wir auf 1265 in VD17 verzeichnete katholische Dissertationen. Die Frage, wie viele katholische Dissertationen insgesamt in deutschen Bibliotheken und darüber hinaus vorhanden sind, würde weitere Recherchen in Verbundkatalogen und Altbestandsbibliotheken ohne VD17-Katalogisierung voraussetzen.

1950 Wörter

Fußnoten

1. Diese Beschränkung erschien auch aus rechtlichen Gründen angezeigt, weil größere Ausschnitte von Datenbanken urheberrechtlichem Schutz unterliegen können. VD17 macht keine Angaben zu den rechtlichen Rahmenbedingungen einer etwaigen Nachnutzung der zur Verfügung gestellten Daten.
2. Daten und das Python-Notebook sind unter <https://gitlab.com/shessbru/catholic-dissertations-in-vd17/-/tree/main> einsehbar.

Bibliographie

- Anonym** (o. J.): *AAD Gattungsgenres*, URL: <http://uri.gbv.de/terminology/aadgenres/>
- Anonym** (2020): *VD17 - Das Verzeichnis der im deutschen Sprachraum erschienenen Drucke*. URL: <http://archive.md/ngh9x>
- CrankyPhilosopher** (2021): „Dumme Frage in den Saal: wie heißt ‚data cleaning‘ eigentlich auf deutsch? #DH“. Tweet. *@FrueheNeuzeit* (Blog), URL: <https://twitter.com/FrueheNeuzeit/status/1414988624874070023>.
- Fischer, Frank / Jäschke, Robert** (2018): „Liebe und Tod in der Deutschen Nationalbibliothek: Der DNB-Katalog als Forschungsobjekt der digitalen Literaturwissenschaft.“ in: *DHd2018: „Kritik der digitalen Vernunft“, Digital Humanities im deutsch-*

sprachigen Raum, Feb. 2018, Köln, Deutschland, 261-266. URL: <https://hal.archives-ouvertes.fr/hal-01787558>

Friederich, Christine (2018): Tagungsbericht: "HT 2018: Quo vadis Quellenkritik? Digitale Perspektiven", 25. 09. 2018–28. 09. 2018 Münster, in: *H-Soz-Kult*, 23. 11. 2018, URL: <http://www.h-sozkult.de/conferencereport/id/tagungsberichte-7977>.

Gittel, Benjamin (2021): „An Institutional Perspective on Genres: Generic Subtitles in German Literature from 1500-2020“, in: *Journal of Cultural Analytics* 10.22148/001c.22086.

Heßbrüggen-Walter, Stefan (2013): "Tatsachen im semantischen Web: Nanopublikationen in den digitalen Geisteswissenschaften?", in: Haber, Peter, Pfanzelter, Eva (eds.): *Historyblogosphere*, München: Oldenbourg Wissenschaftsverlag, 149-160. 10.1524/9783486755732.149

Heßbrüggen-Walter, Stefan (2020): "Positivismus der geistigen Gegenstände: Carnap und die Digital Humanities", in: *DHd 2020 Spielräume: Digital Humanities zwischen Modellierung und Interpretation. 7. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum" (DHd 2020)*, Paderborn: Zenodo. 10.5281/zenodo.4621804

Jochum, Uwe (1998): „Die Bibliothek als locus communis“, in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte*, 14-30 10.1007/BF03375514.

Király, Péter / Brase, Jan (2021): "Qualitätsmanagement", in: Markus Putnigs, Heike Neuroth, Janna Neumann (eds.): *Praxishandbuch Forschungsdatenmanagement*, Berlin, Boston: De Gruyter Saur, 2021, 357-380. 10.1515/9783110657807-020

Leinsle, Ulrich G. (2006): *Dilinganae disputationes: der Lehrinhalt der gedruckten Disputationen an der Philosophischen Fakultät der Universität Dillingen 1555 - 1648*. (Jesuitica 11). Regensburg: Schnell Steiner.

Sangiacomo, Andrea / Beers, Daan (2020): „Divide et Impera: Modeling the Relationship between Canonical and Noncanonical Authors in the Early Modern Natural Philosophy Network“, in: *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, 365-413 10.1086/710178.

Scholz, Luca (2021): „A Distant Reading of Legal Dissertations from German Universities in the Seventeenth Century“, in: *The Historical Journal*, 1-31, 10.1017/S0018246X2100011X.

Stäcker, Thomas (2004): „VD 17 – mehr als eine Zwischenbilanz“, in: *Zeitschrift für Bibliothekswesen und Bibliographie* 51: 213-21.

Wikipedia (2021): „Karmelitenkloster Augsburg“, in: *Wikipedia*, 25. April 2021. https://de.wikipedia.org/w/index.php?title=Karmelitenkloster_Augsburg&oldid=211306498.