

Vom Finden, Filtern und Auswerten der relevanten Daten im digitalen Nachlass von Friedrich Kittler im Deutschen Literaturarchiv Marbach

Holz, Alex

alex.holz@dla-marbach.de
Deutsches Literaturarchiv Marbach, Deutschland

Çakir, Dîlan Canan

dilan.cakir@dla-marbach.de
Deutsches Literaturarchiv Marbach, Deutschland

Auf unserem Poster zeigen wir, wie wir von einem digitalen Nachlass mit 3,3 Millionen Dateien zu einem Data-Set mit etwa 30.000 Dateien gelangt sind, mit dem wir sinnvoll zum Born-digital-Nachlass des Literaturwissenschaftlers und Medientheoretikers Friedrich Kittler (1943–2011), der im Deutschen Literaturarchiv Marbach (DLA) aufbewahrt wird, forschen können und dürfen. Mit unserem Fallbeispiel wollen wir zeigen, wie umfangreich ein Born-digital-Nachlass im DLA sein kann, wie man mit diesem technisch, konservatorisch und rechtlich umgeht und welches Potential darin liegt. Anhand unseres kuratierten Arbeitskorpus können wir sodann beispielsweise an der statistischen Auswertung und Visualisierung der (technischen¹) Metadaten arbeiten und damit zur Erschließung des Nachlasses beitragen.²

Digitale Vor- und Nachlässe im DLA

Mit Kittlers Nachlass kam 2011 der bisher umfangreichste und komplexeste digitale Nachlass ins DLA. Immer noch gehört dieser mit 762 Datenträgern (648 Disketten, 100 optische Medien, 1 USB Speicher, 12 Festplatten) und einer Größe von 1,6 TB zum quantitativ größten digitalen Bestand im DLA. Berücksichtigt man Kittler nicht, umfassen die gesamten digitalen Vor- und Nachlässe von etwa 70 Bestandsbildner*innen im DLA (also im Archiv, nicht in der Bibliothek) derzeit mit etwa 1600 Datenträgern insgesamt knapp 5,2 TB.

In den Jahren 2012 bis 2013 galt es, diesen neuen Bestand analog zu sichten, zu sortieren und zu verzeichnen. Dieser Workflow wurde bereits dokumentiert und publiziert (Enge/Kramski 2017). Die PCs, Laptops und

physischen Datenträger zu sortieren, reicht (anders als beispielsweise das Sortieren von Papiermanuskripten) allerdings nicht aus, um den Nachlass zu katalogisieren. Im DLA haben wir für digitale Vor- und Nachlässe eigens sogenannte „Digital Curator[s]“, die man auch Data Librarian oder „Datenarchäolog[innen]“ nennen könnte (Bülow/Kramski 2011: 159), die seit Mitte der 2000er Jahre immer wichtiger werden (Jaillant 2022: 8). Diese bereiten ‚rohe‘ digitale Nachlässe für verschiedene Anliegen auf. Bei digitalen Archivobjekten „muss die Lese- und Interpretationsfähigkeit [nämlich] zuerst aufwendig hergestellt werden.“ (Bülow/Kramski 2011: 159) Zudem ist eine manuelle Sichtung der Dateien bei einem kleineren digitalen Nachlass zwar vorstellbar, für Kittlers Nachlass allerdings wegen des Umfangs und der Art der Dateien eher unpraktisch.

Kittlers digitaler Nachlass: Von 3,3 Millionen zu ca. 30.000 Dateien

Bei erst kürzlich verstorbenen Autor*innen, wie Kittler, muss immer mitbedacht werden, dass selten der ganze Nachlass benutzbar ist und als Ganzes gar nicht oder zu einem späteren Zeitpunkt erforscht werden kann – bei Born-digital-Nachlässen spricht man so auch von „dark archives“, denn selten kann alles (online) leicht zugänglich gemacht werden, obwohl es bereits digital vorliegt. „Nobody would reasonably claim that all born-digital data should be unlocked and openly accessible. Yet, it is important to recognize that ‚dark‘ archives contain vast amounts of data essential to scholars [...]“ (Jaillant 2022: 7).

Der aktuelle Kittler-Bestand umfasst ca. 3,3 Millionen Dateien. Der erste Schritt besteht in dem ‚Identifizieren und Aussortieren‘ von hunderttausenden nicht-unikaler³ Dateien, also Dateien, die nicht von Kittler selbst erstellt wurden. Dies erfolgte automatisiert vor allem über einen Abgleich mit der National Software Reference Library (NSRL) des NIST⁴. Für solche Vorgänge sind auch KI-gestützte Methoden denkbar; allerdings werden solche derzeit eher für „low-level tasks“ herangezogen, etwa bei der Identifizierung von sensiblen persönlichen Informationen (Jaillant 2022: 14). Damit gelangten wir zu etwa 2,25 Millionen Dateien.

In einem nächsten Schritt wurde die Menge auf die Dateien begrenzt, die von Seiten der Nachlassverwaltung begutachtet und mit einem Status versehen wurden (freigegeben, vorläufig gesperrt, gesperrt). Unser Arbeitskorpus ist also als Momentaufnahme zu sehen, da er nur auf den bereits bewerteten Dateien basiert. Ohne die bislang unbekanntes bzw. noch nicht bewerteten Dateien kommen wir auf 219.989 Dateien.

Aus dieser Menge wurden zuletzt nun die für die Forschung vollständig freigegebenen Dateien (Metadaten und Inhalt) herausgefiltert, die zudem von Kittler erstellt wurden. Wir landeten bei etwa 30.000 Dateien, also etwa 0,88% von den 3,3 Millionen Dateien.

Bei den meisten Dateien in unserem Arbeitskorpus handelt es sich um Textdateien in unterschiedlichsten Formaten. Kittler hat „nicht nur Texte, Bilder und Vi-

deos hinterlassen, sondern auch Relikte seiner programmierenden Tätigkeiten.“ (Enge/Kramski 2017). Er interessierte sich neben der Literatur, Musik und Philosophie vor allem für Technologien, Medien und das Programmieren (Winthrop-Young, 2017: 210). „[W]ie ungezählte Teenager in dieser Zeit“ habe er vor allem mit Codes gespielt und autodidaktisch gelernt (Pias 2014: 39–44). Das spiegelt sich in gewisser Weise auch in seinem Nachlass wider. So befinden sich im Arbeitskorpus beispielsweise auch mit Kommentaren versehene Dateien der Programmiersprache C/C++ (insbesondere *.c und *.h-Dateien).

2014.“ Merkur. Deutsche Zeitschrift für europäisches Denken 69, Nr. 791 (2015): 31–43.

Riley, Jenn. Understanding metadata. What is metadata and what is it for? Baltimore 2017.

Fußnoten

1. Wir orientieren uns bei den Metadaten an der u.a. von Jenn Riley (2017) beschriebenen Kategorisierung. Die Technischen Metadaten sind eine Untergruppe der administrativen Metadaten. Es geht also um Informationen zu den Dateien selbst, nicht ihren Inhalt (deskriptive Metadaten). Neben FileSize und mimeType sind hier für uns besonders die Zeitstempel oder auch Informationen aus den Dateipfaden wichtig.

2. Dies ist Teil des Projekts „Archivierung, Erschließung und Erforschung von Born-digitals“ des Forschungsverbundes Marbach, Weimar, Wolfenbüttel, <https://www.mww-forschung.de/born-digitals> (letzter Zugriff 08.07.2022).

3. Als unikale Dateien bezeichnen wir Dateien, die von dem/der Bestandsbilder*in erstellt wurden und die außerhalb dieses digitalen Vor- oder Nachlasses nicht verfügbar sind.

4. National Software Reference Library (NSRL) des National Institute of Standards and Technology. <https://www.nist.gov/itl/ssd/software-quality-group/national-software-reference-library-nsrl> (letzter Zugriff am 22.07.2022).

Bibliographie

Bülow, Ulrich von/Kramski, Heinz Werner. „Es füllt sich der Speicher mit köstlicher Habe‘ – Erfahrungen mit digitalen Archivmaterialien im Deutschen Literaturarchiv Marbach.“ In: Neues Erbe: Aspekte, Perspektiven und Konsequenzen der digitalen Überlieferung, hg. v. Caroline Y., Hauser Robert, Robertson-von Trotha, 141–62. Kulturelle Überlieferung - digital 1. Karlsruhe: KIT Scientific Publishing, 2011.

Enge, Jürgen/Kramski, Heinz Werner. „Friedrich Kittler’s Digital Legacy – PART I - Challenges, Insights and Problem-Solving Approaches in the Editing of Complex Digital Data Collections.“ Digital Humanities Quarterly 11, Nr. 2 (22. Mai 2017). <http://www.digitalhumanities.org/dhq/vol/11/2/000307/000307.html>.

Jaillant, Lise. „Introduction.“ In: Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections, hg. v. Jaillant, Lise, 7–28. Bielefeld, Germany: Bielefeld University Press / transcript Verlag, 2022.

Pias, Claus. „Friedrich Kittler und der ‚Mißbrauch von Heeresgerät‘. Zur Situation eines Denkbildes 1964–1984–