

Tools zur Übernahme digitaler Dateiablagen

Ein Test von Archifiltre, DROID und TreeSize Professional zur Umsetzung der technischen Analyse von Privatnachlässen anhand des Musterworkflows der KOST-Arbeitsgruppe «Dateiablage»

Leonie Fritz

*Thema des auf einer Zertifikatsarbeit basierenden Artikels ist der Prozessschritt der technischen Analyse bei der Übernahme digitaler Dateiablagen. Im praktischen Teil werden die drei Tools Archifiltre, DROID und TreeSize Professional auf ihre Eignung für die Ausführung der technischen Analyse anhand der digitalen Dateiablage eines Nachlasses aus dem Arbeitsalltag der Verfasserin geprüft. Nebst dem Test der erforderlichen Funktionalitäten wird die Frage geklärt, ob die Tools gebrauchstaugliche Arbeitsinstrumente für Anwender*innen ohne IT-Vorkenntnisse sind. Die Untersuchung ergab, dass Archifiltre und TreeSize Professional nach den definierten Kriterien als geeignete Tools für die technische Analyse als Ganzes gelten können. DROID hingegen ist auf einen der Arbeitsschritte innerhalb der Analyse spezialisiert.*

Le thème de cet article, basé sur un travail de certificat, est l'étape du processus d'analyse technique lors de la prise en charge de fichiers numériques. Dans la partie pratique, les trois outils Archifiltre, DROID et TreeSize Professional sont testés quant à leur aptitude à effectuer l'analyse technique de fichiers numériques provenant d'un fonds privé pris en charge par l'auteure. Outre le test des fonctionnalités requises, la question est de savoir si ces outils constituent des instruments de travail adaptés aux utilisatrices et utilisateurs ne bénéficiant pas de connaissances informatiques préalables. L'enquête a montré qu'Archifiltre et TreeSize Professional peuvent être considérés comme des outils adaptés à l'analyse technique dans son ensemble, selon les critères définis. DROID, en revanche, est spécialisé dans l'une des étapes de l'analyse.

The article (based on a certificate thesis) deals with the technical analysis process when digital files are handed over to an archive. In the practical part, the three tools Archifiltre, DROID and TreeSize Professional have been tested for their ability to perform the technical analysis of digital files from private archives for which the author has been responsible. In addition to testing the required functionality, the question is whether these tools are suitable working tools for users with no prior computer knowledge. The survey shows that Archifiltre and TreeSize Professional can be considered suitable tools for technical analysis as a whole, according to the criteria defined. DROID, on the other hand, specialises in one of the stages of the analysis.

1 Einleitung

Es ist ein vertrautes Bild für Archivar*innen: Einer übernommenen Ablieferung oder einem zu erschliessenden Bestand liegt – neben Ordnern, in Hängeregistern oder Schachteln mit analogem Material – eine Diskette, eine CD-ROM, ein Memorystick bei, oder man stösst auf eine Schachtel mit Audio- oder VHS-Kassetten. Zum Umgang mit diesen Datenträgern gibt es erprobte, in Workflows geregelte Strategien. Aktuell ist eine Veränderung auszumachen. Die Archivmitarbeitenden sind zunehmend mit der Ablieferung von «digital born»-Daten konfrontiert – das heisst mit Dokumenten, die digital produziert wurden und nur in digitaler Form vorliegen. Diese sind nicht in einer analogen Struktur verordnet. Das Ordnungssystem der Donatorin oder des Donators beziehungsweise der abliefernden Stelle ist eine digital gewachsene Struktur, die nicht ohne technischen Aufwand überblickt und bewertet werden kann. Bei der Ablieferung von Privatnachlässen begegnen Archivmitarbeitende zunehmend grossen und teils chaotischen elektronischen Ablagen, die Dokumente in verschiedensten Formaten enthalten. Bei diesen kann – im Gegensatz zu Übernahmen von Archivgut der öffentlichen Verwaltung, die der GEVER-Verordnung vom 3.4.2019¹ unterstehen – selten beeinflusst werden, wie die digitalen Unterlagen strukturiert werden.

1 Vgl. <https://fedlex.data.admin.ch/filestore/fedlex.data.admin.ch/eli/cc/2019/253/20210101/de/pdf-a/fedlex-data-admin-ch-eli-cc-2019-253-20210101-de-pdf-a.pdf> (Zugriff 10.2.2023).

1.1 Fragestellung

Informationen die ausschliesslich digital erstellt und veröffentlicht werden, stellen für alle Archive eine grosse und hochaktuelle Herausforderung dar. Aus diesem Grund initiierte die «Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen» (KOST), als Kompetenzzentrum der Schweizer und Liechtensteiner Archive, 2020 die Arbeitsgruppe «Dateiablage». Diese beschäftigte sich mit der Erarbeitung eines Workflows für die Übernahme digitaler Dateiablagen und erprobte Tools für den Übernahmeprozess.² Die Zertifikatsarbeit, auf welcher dieser Artikel basiert, orientiert sich an den Resultaten der Arbeitsgruppe. Anhand der digitalen Dateiablage des Nachlasses von Heinz Stefan Herzka, Prof. Dr. med. Kinder- und Jugendpsychiater (1935-2021) im Archiv für Zeitgeschichte der ETH Zürich (AfZ), wird die technische Analyse mit verschiedenen Tools erprobt und ausgewertet. Die Arbeit geht folgenden Fragestellungen nach:

- Ist der Prozessschritt der technischen Analyse im von der KOST-Arbeitsgruppe «Dateiablage» erarbeiteten Workflow für Personen ohne Vorkenntnisse im IT-Bereich und in der digitalen Langzeitarchivierung umsetzbar?
- Welche Erfahrungen wurden mit Tools gemacht, die für die Umsetzung des Prozessschrittes der technischen Analyse infrage kommen? Wie gebrauchstauglich sind die vorgeschlagenen Tools?
- Welche der untersuchten Tools eignen sich für die technische Analyse des Nachlasses Herzka beziehungsweise von privaten Beständen, wie wir sie im AfZ übernehmen?³

1.2 Forschungsstand

Um die Jahrtausendwende wurden international erste Konzepte und Standards zur digitalen Langzeitarchivierung entworfen.⁴ 2002 wurde das Referenzmodell «Open Archival Information System» (OAIS) veröffentlicht und zum ISO-Standard erhoben. Dank seiner Dynamik, Anpassungsfähigkeit und Offenheit ist

2 Vgl. <https://www.kost-ceco.ch/kostwiki/doku.php> (Zugriff 10.2.2023).

3 Die Verfasserin arbeitet im Archiv für Zeitgeschichte der ETH Zürich.

4 Lenartz, Stephan: Digital ist besser? Möglichkeiten der automatisierten Aufbereitung und Bewertung von Fileablagen mit Python am Beispiel einer digitalen Fotosammlung. Werkhefte Digital 1. Landesarchiv Baden-Württemberg 2020, S. 15.

es ein weiterhin aktuelles und zukunftsfähiges Modell geblieben.⁵ Während der umfassende Bereich der digitalen Langzeitarchivierung ein etabliertes Forschungsgebiet ist, rückte das spezifische Forschungsfeld der Übernahme digitaler Dateiablagen erst seit Mitte der 2010er-Jahre in den Fokus. So schrieb Stephan Lenartz 2020, es gebe «bisher nur wenige konkrete Erfahrungen im Umgang mit unstrukturierten Fileablagen, auf die zurückgegriffen werden könnte». Dies gelte nicht nur für den deutschen, sondern auch für den europäischen und internationalen Kontext.⁶

Obwohl bereits erste Fallstudien und Erfahrungsberichte publiziert wurden⁷, fand eine breitere Auseinandersetzung mit Dateisammlungen erst in einer 2016 durchgeführten Tagung der «Deutschen Konferenz der Leiterinnen und Leiter der Archivverwaltungen des Bundes und der Länder» zum Thema «Kreative digitale Ablagen und die Archive» statt, deren Ergebnisse als Sammelband vorliegen.⁸ Dieser enthält Beiträge zu allgemeinen Fragen der Übernahme digitaler Dateiablagen⁹ sowie Erfahrungsberichte durchgeführter Pilotprojekte¹⁰. Die seither entstandenen Publikationen können in zwei Kategorien eingeteilt werden. Es sind einerseits Projektbeschreibungen durchgeführter Übernahmen, in denen der Übernahmeprozess von der Akzession bis zur Zugänglichma-

5 Neuroth, Heike: Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.3, Boizenburg 2010, S. 76, 79, 81.

6 Lenartz: Digital ist besser?, 2020, S. 19f.

7 Vgl. u. a. Uglean Jackson, Laura; McKinley, Matthew: It's How Many Terabytes?! A Case Study on Managing Large Born Digital Audio-Visual Acquisitions, in: International Journal of Digital Curation (IJDC), Volume 11/2016, Edinburgh 2016, S. 64-75; Enge, Jürgen; Kramski, Heinz Werner: «Arme Nachlassverwalter...» Herausforderungen, Erkenntnisse und Lösungsansätze bei der Aufbereitung komplexer digitaler Dateisammlungen, in: Filthaut, Jörg (Hg.), Von der Übernahme zur Benutzung. Aktuelle Entwicklungen in der digitalen Archivierung. 18. Tagung des Arbeitskreises «Archivierung von Unterlagen aus digitalen Systemen», am 11. und 12. März 2014 in Weimar (Schriften des Thüringischen Hauptstaatsarchivs Weimar 6), Weimar 2015, S. 53-62.

8 Naumann, Kai; Puchta, Michael (Hg.): Kreative digitale Ablagen und die Archive. Ergebnisse eines Workshops des KLA-Ausschusses Digitale Archive am 22./23. November 2016 in der Generaldirektion der Staatlichen Archive Bayerns, München 2017.

9 Vgl. u. a. Miegel, Annekathrin; Schieber, Sigrid; Schmidt, Christoph: Vom richtigen Umgang mit kreativen digitalen Ablagen, in: Naumann, Kai; Puchta, Michael (Hg.), Kreative digitale Ablagen und die Archive. Ergebnisse eines Workshops des KLA-Ausschusses Digitale Archive am 22./23. November 2016 in der Generaldirektion der Staatlichen Archive Bayerns, München 2017, S. 7-16.

10 Vgl. u. a. Birn, Marco: Analyse und Datenaufbereitung von digitalen Ablagen mit TreeSize Professional und Total Commander, in: Naumann, Kai; Puchta, Michael (Hg.), Kreative digitale Ablagen und die Archive. Ergebnisse eines Workshops des KLA-Ausschusses Digitale Archive am 22./23. November 2016 in der Generaldirektion der Staatlichen Archive Bayerns, München 2017, S. 61-70; Naumann, Kai: Welche Schritte erfordert die Aufbereitung von Dateisammlungen und welche Querschnitts- und Spezialwerkzeuge werden gebraucht?, in: Naumann, Kai; Puchta, Michael (Hg.), Kreative digitale Ablagen und die Archive. Ergebnisse eines Workshops des KLA-Ausschusses Digitale Archive am 22./23. November 2016 in der Generaldirektion der Staatlichen Archive Bayerns, München 2017, S. 44-60.

chung der Bestände beschrieben werden. Dabei werden verwendete Tools vorgestellt und deren Funktionalitäten kritisch ausgewertet.¹¹ Andererseits handelt es sich um Texte, in denen Strategien einzelner Archivinstitutionen im Umgang mit digitalen Objekten vorgestellt und Erfahrungen im Umgang mit diesen diskutiert werden.¹² Allgemein lässt sich festhalten, dass es bislang kein Standardwerk gibt, in dem die Übernahme digitaler Dateiablagen ausführlich thematisiert oder ein allgemeingültiger Workflow dafür beschrieben wird. Die aktuellste und umfangreichste Publikation, welche (nebst dem ausführlich beschriebenen Fallbeispiel der Aufbereitung und Bewertung einer digitalen Fotosammlung mithilfe der Programmiersprache Python) die Grundlagen der Übernahme digitaler Dateiablagen thematisiert, ist die eingangs erwähnte Studie von Lenartz.¹³

1.3 Methode

Der im Rahmen des KOST-Projekts «20-039 Dateiablage» erarbeitete Musterworkflow beschreibt einen idealtypischen Archivierungsprozess digitaler Dateiablagen anhand von sechs Prozessschritten. Im vorliegenden Artikel wird die technische Analyse fokussiert. Es wird danach gefragt, welche Tools dafür geeignet und benutzerfreundlich sind. Untersucht und verglichen werden die Tools Archifiltre, DROID und TreeSize Professional (nachfolgend TreeSize genannt). Diese drei Tools werden sowohl beim Projekt der KOST als auch in der nestor-Toolbox¹⁴ für die Umsetzung der praktischen Analyse vorgeschlagen; sie sind von ihrem Einsatzgebiet her also ähnlich und damit für einen Vergleich geeig-

11 Vgl. u. a. Jaeger, Karina; Kobold, Maria: Zwischen Datenwust und arbeitsökonomischer Bewertung. Ein Werkstattbericht zum Umgang mit unstrukturierten Dateisammlungen am Beispiel des Bestandes der Odenwaldschule, in: *Archivar* 70 (2017), Heft 3, S. 307-311; Knobloch, Corinna: Archivischer Umgang mit digitalen Sammlungen am Beispiel der Johannes-Wagner-Schule Nürtingen, in: Stumpf, Marcus; Tiemann, Katharina, *Erziehung und Bildung als kommunalarchivische Überlieferungsfelder*. Beiträge des 27. Fortbildungsseminars, November 2019, Münster 2019, S. 76-86; Lenartz, Stephan: *Aufbereitung und Bewertung von Fileablagen mit Python am Beispiel einer digitalen Fotosammlung*, St. Gallen 2018; Steigerwald, Jelena: *Erste Übernahmen digitaler Daten aus der Landesverwaltung*, in: *Archive in Sachsen-Anhalt* 2019, Magdeburg 2019, S. 10-12; Taylor, Isabel: *A Hydra-like Russian Doll. Appraising and Describing the Shared Drive of a Staatliches Schulamt*, in: *Jaarboeken Stichting Archiefpublicaties* 2018, S. 150-159.

12 Vgl. u. a. Birn, Marco; Naumann, Kai: *Bewertung schwach strukturierter Unterlagen. Berichte und Thesen aus Baden-Württemberg*, in: *Brandenburgischen Landeshauptarchiv* (Hg.), *Brandenburgische Archive. Berichte und Mitteilungen aus den Archiven des Landes Brandenburg*, Nr. 36/2019, Potsdam 2019, S. 8-14; Heizmann, Uwe: *Strategie zur Archivierung digitaler Objekte im Kreisarchiv Reutlingen*, Reutlingen 2017.

13 Lenartz: *Digital ist besser?*, 2020.

14 Vgl. <https://wiki.dnb.de/pages/viewpage.action?pageId=134715087> (Zugriff 10.2.2023).

net. Zur Bestimmung ihrer Benutzerfreundlichkeit wird die empirische Methode der «System Usability Scale» (SUS) angewandt.

2 Theoretische Grundlagen zu digitalen Dateiablagen

Unter einer digitalen Dateiablage – in der Literatur auch als Dateisammlung oder Fileablage bezeichnet¹⁵ – versteht man nach Miegel, Schieber und Schmidt eine «Menge von Einzeldateien, die von einem oder mehreren Bearbeitern zur Erledigung einer oder mehrerer Aufgaben über einen bestimmten Zeitraum erstellt und nach individuellen Ordnungskriterien zusammengestellt wurden. Die Dateien liegen auf einer Ebene und/oder hierarchisch in einer Verzeichnisstruktur vor. Es können in einer Dateisammlung unterschiedlichste Dateiformate enthalten sein».¹⁶ Lenartz fügt hinzu, dass digitale Dateiablagen «dem Archiv in ihrer Gesamtheit [...] angeboten werden». Auch merkt er an, dass solche Ablagen vor allem typisch für nichtamtliche Überlieferungen, wie digitale Nachlässe seien, aber auch in Behörden abseits der geregelten Aktenführung entstünden.¹⁷ Während die Ablage analoger Dokumente mit zunehmender Menge an Unterlagen an physische Grenzen stösst, ist bei digitalen Ablagen sowohl die Zahl der Dokumente als auch der Anzahl Ordner auf einer Ebene nicht begrenzt. Auch lassen sich neue Ordner so einfach anlegen, dass dies oft schneller geht, als sich zu überlegen, in welche bestehende Position der Ablagestruktur ein Dokument einzuordnen wäre. Zudem besteht grundsätzlich kein Zwang mehr, die Dokumente überhaupt einem Ordner zuzuweisen, denn das System erlaubt die Ablage auf jeder Ebene und an jeder Stelle in der Ordnerhierarchie. All dies führt dazu, dass digitale Ablagestrukturen besonders umfangreich sind.¹⁸

Daraus ergeben sich die Charakteristika,

- dass digitale Dateiablagen in der Regel eine unübersichtliche Anzahl an Einzeldateien enthalten;
- dass die Dokumente in zahlreichen, unterschiedlichen Formaten abgespeichert sind;

15 Naumann: *Aufbereitung von Dateisammlungen*, 2017, S. 44.

16 Miegel; Schieber; Schmidt: *Vom richtigen Umgang mit kreativen digitalen Ablagen*, 2017, S. 7.

17 Lenartz: *Digital ist besser?*, 2020, S. 16.

18 Schludi, Ulrich: *Zwischen Records Management und digitaler Archivierung. Das Dateisystem als Basis von Schriftgutverwaltung und Überlieferungsbildung*, in: Kai Naumann, Peter Müller (Hg.), *Das neue Handwerk – Digitales Arbeiten in kleinen und mittleren Archiven*, Stuttgart 2013, S. 24.

- dass archivwürdige und nicht-archivwürdige Dateien nebeneinanderliegen;
- dass die Ablagestruktur meist thematischen Kriterien folgt.¹⁹

Diese Besonderheiten führen dazu, dass jede digitale Dateiablage in Grösse, Zusammensetzung und Struktur einzigartig ist. Das macht es – zusammen mit der meist schwachen Strukturierung der Daten²⁰ – so schwierig, für die Übernahme dieser Ablagen Werkzeuge, Strategien oder gar einen standardisierten Bearbeitungsprozess zu erarbeiten.²¹

3 Musterworkflow der KOST-Arbeitsgruppe «Dateiablage»

Der Musterworkflow wurde von einer Arbeitsgruppe aus Mitarbeitenden zehn verschiedener schweizerischer Archivinstitutionen unter Leitung der KOST, auf bislang gesammelten Erfahrungen beruhend, erarbeitet. Er beschreibt die Übernahme digitaler Dateiablagen anhand von sechs Prozessschritten. Der Workflow ist in die Prozessschritte Beratung, Sichtung, Übernahme, Analyse, Aufbereitung und Erschliessung unterteilt. Diese sind jeweils in bis zu fünf Unterschritten aufgliedert.²² Er steht im Sinne einer Empfehlung allen Institutionen zur Verfügung.²³

Der vorliegende Artikel beschäftigt sich mit dem Prozessschritt der Analyse. Diese umfasst drei Arbeitsschritte. Im Rahmen der «technischen Analyse» wird der Umfang und das Mengengerüst der Ablieferung bestimmt, die Dateiformate eruiert und inventarisiert, Dubletten ermittelt und die Metadaten ausgelesen. Dies dient der Identifizierung von technischen Problemen und Herausforderungen sowie der Definition von Sofortmassnahmen. Die «inhaltliche/archivfachliche Analyse» dient der Analyse von Ordnungsstruktur und Mengengerüst sowie der Identifizierung sensibler Daten. Auf dieser Grundlage kann eine Top-Down-Bewertung der Ablieferung beschlossen werden. Der «Abschluss der Analysephase» umfasst das Verfassen eines Analyseberichts und den Beschluss von Sofortmassnahmen und Top-Down-Kassationen.

19 Lenartz: Digital ist besser?, 2020, S. 17.

20 Heizmann, Uwe: Strategie zur Archivierung digitaler Objekte im Kreisarchiv Reutlingen, in: Archivar 70 (2017), Heft 3, S. 306.

21 Lenartz: Aufbereitung und Bewertung von Fileablagen, 2018, S. 1.

22 Der Musterworkflow wird in der Zertifikatsarbeit vorgestellt. Im vorliegenden Artikel wird auf diese detaillierte Erläuterung verzichtet.

23 Vgl. <https://www.kost-ceco.ch/kostwiki/doku.php> (Zugriff 10.2.2023).

4 Getestete Tools

Der Bestand von Heinz Stefan Herzka - Kinder- und Jugendpsychiater sowie Professor an der Universität Zürich, dessen Nachlass im Archiv für Zeitgeschichte liegt – bildet die Grundlage für die Tests. Anhand der 175 GB umfassenden und schwach strukturierten digitalen Nachlieferung zum bereits erschlossenen, analogen Bestand wurde die technische Analyse mit den Tools Archifiltre, DROID und TreeSize durchgeführt. Nachfolgend werden die drei Tools vorgestellt.

4.1 Archifiltre

Archifiltre ist ein Werkzeug zur Visualisierung digitaler Dateiablagen. Es ist eine kostenlose Open-Source-Software. Sie ermöglicht die Analyse der Bestandsstruktur sowie der vorhandenen Dateien. Ebenfalls können die Dateien geprüft, sortiert und bearbeitet werden.²⁴ In Archifiltre vorgenommene Änderungen greifen dabei nicht direkt in die ursprüngliche Struktur ein. Hat man das Tool heruntergeladen – für die Benutzung ist keine Installation erforderlich – kann die zu analysierende digitale Dateiablage durch Drag and Drop unkompliziert geöffnet werden. Auf der allgemeinen Benutzeroberfläche erscheint eine in drei Hauptelemente gegliederte Ansicht. Auf der linken oberen Seite befindet sich ein Feld, in dem die Haupteigenschaften der Dateiablage beschrieben werden. Es sind dies der Titel, die Grösse, die Anzahl enthaltener Ordner und Dateien sowie die Zeitspanne der letzten Bearbeitung. Im Feld rechts oben erscheinen Metadaten des ausgewählten Elements innerhalb der Dateiablage wie der Titel, die Grösse, der Dateityp, der Hashwert und das letzte Änderungsdatum. Im Hauptbereich wird die zur Analyse ausgewählte Ablage als Balkendiagramm visualisiert. Eine Farbskala gibt Aufschluss über die vorhandenen Dateitypen. Fährt man mit der Maus ins Balkendiagramm und verharret auf einem Unterelement, wird unterhalb der Grafik sein Anteil an der Gesamtgrösse der Ablage in Prozent angegeben. Per Doppelklick auf ein Unterelement wird dieses herangezoomt. Archifiltre ermöglicht das Erstellen von Reports, etwa in ein Excel-Dokument.

24 Vgl. <https://github.com/SocialGouv/archifiltre/wiki/Wiki-Archifiltre> (Zugriff 13.7.2021).

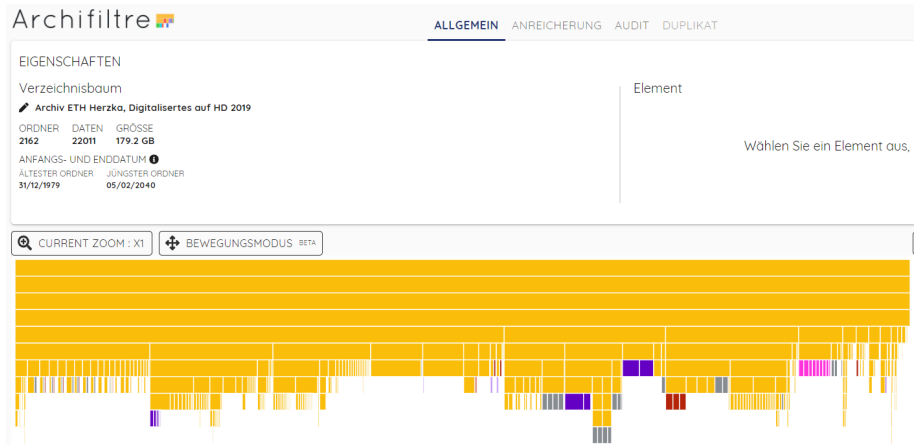


Abb. 1. Benutzeroberfläche mit Übersicht über die digitale Dateiablage in Archifiltre

Es gibt drei weitere Ebenen. Unter «Anreicherung» besteht unter anderem die Möglichkeit, die Elemente zu bearbeiten und Kommentare einzufügen. Ändert man den Titel eines Ordners oder einer Datei, werden sowohl der neue als auch der ursprüngliche Titel gespeichert und angezeigt. Die Ergänzungen sind in Archifiltre sichtbar und können als Reports gespeichert werden. Sie werden nicht als Primärdaten übernommen. Die Ebene «Audit» gibt Aufschluss über die vorhandenen Dateiformate unter Angabe der Anzahl und der Grösse der Summe aller Dateien desselben Formats. Im Reiter «Duplikat» besteht die Möglichkeit, mehrfach vorhandene Dateien zu suchen. Diese werden anhand der automatisch erstellten Hashwerte aufgrund der MD5-Prüfsumme identifiziert. Das Tool ermöglicht die Auflistung der Dubletten mit der Option, diese als Listen zu exportieren oder direkt zu löschen.

4.2 DROID

DROID steht als Abkürzung für «Digital Record Object Identification». Es ist eine kostenlose und quellenoffene Software. Die wichtigste Funktion von DROID ist die Formatidentifizierung innerhalb einer digitalen Dateiablage. Zudem ist das Tool imstande, eine grosse Anzahl von Metadaten zu erheben. So kann es etwa Dateigrößen, letzte Änderungsdaten und den Pfad wiedergeben. Es besteht die Möglichkeit innerhalb des Tools mit sogenannten Reports im Graphical User Interface (GUI) zu arbeiten. Ebenfalls können die gesamten Informationen in ein Excel-Dokument exportiert und dort analysiert werden. Um das Excel-Dokument

kument interpretieren zu können, sind gewisse Einstellungen bei DROID und in Excel notwendig.

Die grafische Benutzeroberfläche von DROID ist schlicht. Hat man die zu analysierende digitale Dateiablage importiert, wird diese in ihrer Explorer-Struktur wiedergegeben. Klickt man sich bis zur gewünschten Datei durch, erscheinen in den Spalten neben dem Titel weitere Metadaten. Die Ansicht gleicht der Darstellung einer Tabelle in Excel. Die Angaben umfassen unter anderem die Dateiendung, die Grösse, das Datum der letzten Bearbeitung, das Format und die PUID²⁵. Exportiert man die Daten in ein Excel-Dokument werden zusätzliche Angaben sichtbar, etwa der URI und der Pfad.

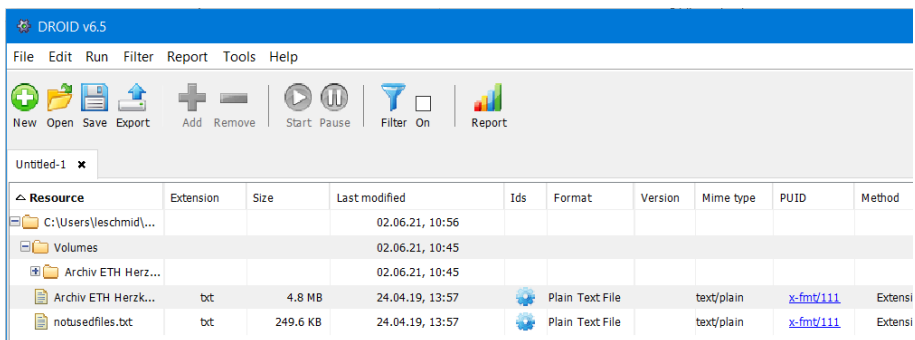


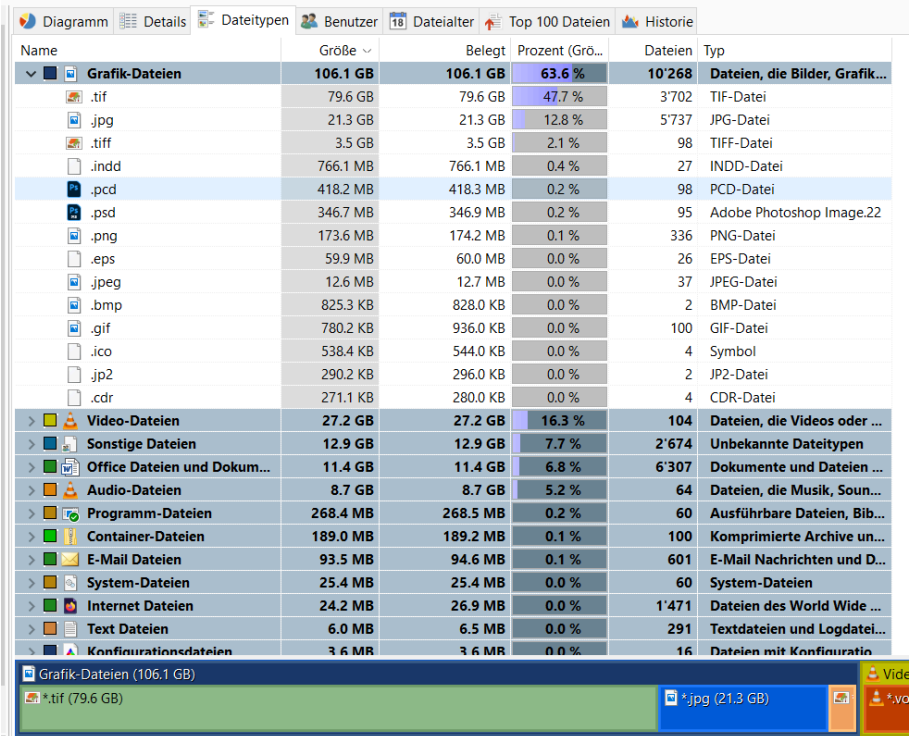
Abb. 2. Benutzeroberfläche mit Übersicht über Bestandsstruktur in DROID

4.3 TreeSize Professional

TreeSize ist eine lizenz- und damit kostenpflichtige Software zur Analyse von digitalen Dateiablagen, dem Identifizieren und Löschen von Dubletten und unerwünschten Dateien wie etwa Systemdateien. Sie bietet eine übersichtliche Darstellung über das zu analysierende Verzeichnis inklusive aller Unterverzeichnisse. Die Benutzeroberfläche ist an den Explorer angelehnt und daher relativ intuitiv zu bedienen. In der linken oberen Hälfte wird der Verzeichnisbaum des gescannten Laufwerks oder Ordners angezeigt. In der Kopfzeile findet man die Eckdaten zum ausgewählten Pfad wie etwa die Grösse, die Anzahl Dateien und Verzeichnisse oder der Speicherplatzbedarf. Der Hauptbereich bietet verschiedene Ansichten der zu analysierenden digitalen Dateiablage. Die «Detail»-Ansicht zeigt Informationen zu allen Dateien und Ordnern des ausgewählten Ele-

25 DROID arbeitet mit einer PRONOM-Datenbank. Darin wird jedem Format eine sogenannte PUID (PRONOM Unique Identifier) zur eindeutigen Identifizierung desselben vergeben.

ments, analog zu den Eckdaten des Pfades. Die Ansicht nach «Dateitypen» bietet einen Überblick über die vorhandenen Dateitypen. Es lassen sich die vorhandenen Formate anzeigen, inklusive der Angabe, wie viel Speicherplatz diese insgesamt in Anspruch nehmen. Die «Diagramm»-Ansicht ermöglicht eine Visualisierung der Grösseninformationen, wahlweise etwa nach Platzbedarf oder Anzahl der Dateien. Aufschlussreich ist ausserdem die Ansicht der «top 100 Dateien», bei der man einen Überblick über die 100 grössten Dateien innerhalb der Dateiablage gewinnt. In jeder Ansicht besteht die Möglichkeit zum Export und Drucken detaillierter Berichte und Diagramme sowie zum Export als PDF, HTML, XML, in eine Textdatei oder in ein Excel-Dokument.



Name	Größe	Belegt	Prozent (Grö...	Dateien	Typ
Grafik-Dateien	106.1 GB	106.1 GB	63.6 %	10'268	Dateien, die Bilder, Grafik...
.tif	79.6 GB	79.6 GB	47.7 %	3'702	TIF-Datei
.jpg	21.3 GB	21.3 GB	12.8 %	5'737	JPG-Datei
.tiff	3.5 GB	3.5 GB	2.1 %	98	TIFF-Datei
.indd	766.1 MB	766.1 MB	0.4 %	27	INDD-Datei
.pcd	418.2 MB	418.3 MB	0.2 %	98	PCD-Datei
.psd	346.7 MB	346.9 MB	0.2 %	95	Adobe Photoshop Image.22
.png	173.6 MB	174.2 MB	0.1 %	336	PNG-Datei
.eps	59.9 MB	60.0 MB	0.0 %	26	EPS-Datei
.jpeg	12.6 MB	12.7 MB	0.0 %	37	JPEG-Datei
.bmp	825.3 KB	828.0 KB	0.0 %	2	BMP-Datei
.gif	780.2 KB	936.0 KB	0.0 %	100	GIF-Datei
.ico	538.4 KB	544.0 KB	0.0 %	4	Symbol
.jp2	290.2 KB	296.0 KB	0.0 %	2	JP2-Datei
.cdr	271.1 KB	280.0 KB	0.0 %	4	CDR-Datei
Video-Dateien	27.2 GB	27.2 GB	16.3 %	104	Dateien, die Videos oder ...
Sonstige Dateien	12.9 GB	12.9 GB	7.7 %	2'674	Unbekannte Dateitypen
Office Dateien und Dokum...	11.4 GB	11.4 GB	6.8 %	6'307	Dokumente und Dateien ...
Audio-Dateien	8.7 GB	8.7 GB	5.2 %	64	Dateien, die Musik, Soun...
Programm-Dateien	268.4 MB	268.5 MB	0.2 %	60	Ausführbare Dateien, Bib...
Container-Dateien	189.0 MB	189.2 MB	0.1 %	100	Komprimierte Archive un...
E-Mail Dateien	93.5 MB	94.6 MB	0.1 %	601	E-Mail Nachrichten und D...
System-Dateien	25.4 MB	25.4 MB	0.0 %	60	System-Dateien
Internet Dateien	24.2 MB	26.9 MB	0.0 %	1'471	Dateien des World Wide ...
Text Dateien	6.0 MB	6.5 MB	0.0 %	291	Textdateien und Logdatei...
Konfigurationsdateien	3.6 MB	3.6 MB	0.0 %	16	Dateien mit Konfiguration

Abb. 3. «Dateitypen»-Ansicht unter Angabe der vorhandenen Dateiformate in TreeSize

Die «Dateisuche» bietet die Möglichkeit einer gezielten Suche nach definierten Kriterien. Es lässt sich etwa eine Auflistung der grössten oder der ältesten Dateien, von Dateien mit einem besonders langen Pfad, von temporären oder von doppelten Dateien generieren. Die Dubletten können anhand verschiedener Vergleichsmethoden gesucht werden, so etwa anhand von Name, Grösse und

Datum oder des genaueren Kriteriums der Prüfsumme. TreeSize bietet auch die Möglichkeit zur «Stapelumbenennung» (dem Umbenennen mehrerer Dateien auf einmal) und zum Löschen von Dubletten.

5 Funktionalitäten und Gebrauchstauglichkeit der Tools

Die detaillierte Auswertung der Funktionalitäten der einzelnen Tools werden in der Zertifikatsarbeit präsentiert. In diesem Artikel werden lediglich die Resultate der Auswertungen erläutert. Ebenfalls werden die drei Tools bezüglich ihrer Funktionalitäten zur Umsetzung der technischen Analyse verglichen. Die Gebrauchstauglichkeit von digitalen Tools wurde mithilfe der System Usability Scale (SUS) bestimmt.²⁶

5.1 Vergleich der Funktionalitäten

Beim Erproben der drei Tools Archifiltre, DROID und TreeSize wurde das Augenmerk auf die Arbeitsschritte gelegt, die (nach dem Musterworkflow der KOST Arbeitsgruppe «Dateiablage») zur technischen Analyse gehören. Es sind dies die Erhebung des Mengengerüsts der Ablieferung, die Bestimmung der Dateiformate, die Ermittlung von doppelten Dateien und das Auslesen von Metadaten sowie die Identifizierung von technischen Problemen und Herausforderungen.

5.1.1 Umfang und Mengengerüst

Alle drei Tools sind in der Lage, die Gesamtgrösse der analysierten Dateiablage zu eruieren. Während bei den Tools Archifiltre und TreeSize der Gesamtumfang prominent ersichtlich ist, muss er bei DROID mithilfe mehrerer Arbeitsschritte ermittelt werden. Interessant ist dabei, dass die Gesamtgrösse nicht bei allen Tools identisch ist. Archifiltre und DROID berechnen einen Gesamtumfang von 166.9 GB, wobei TreeSize (gleich wie der Explorer) eine Grösse von 179.2 GB angibt. Dieser Unterschied ist darauf zurückzuführen, dass die (identische) Anzahl von Bytes einmal binär und einmal dezimal in Gigabytes umgerechnet wird.

Die Erhebung des Mengengerüsts – also die Aufschlüsselung der Bestandsstruktur nach der Grösse der einzelnen Elemente oder dem Anteil einzelner Dateitypen – ist bei Archifiltre visuell erfassbar. Man kann jedes Element, ob

26 Die SUS wird in der Zertifikatsarbeit ausführlich besprochen.

Ordner oder Datei, anwählen und sich den Umfang als Prozentangabe des Gesamtvolumens anzeigen lassen. Zudem ist sichtbar, wie viele Dateien desselben Formats existieren.²⁷ Auch TreeSize bietet eine übersichtliche Darstellung des Mengengerüsts. Der Verzeichnisbaum listet nicht nur die Titel von Ordnern und Dateien auf, sondern auch deren Grösse. Die Darstellung nach Dateitypen gibt Aufschluss darüber, wie gross die Summe aller Dateien eines Dateityps ist und wie viel Prozent des Gesamtvolumens diese ausmachen. DROID besitzt keine Funktion, mithilfe derer die Erhebung des Mengengerüsts ohne Weiteres möglich wäre.

5.1.2 Dateiformate und Pfadlängen

Alle drei Tools können verschiedene Dateiformate zu identifizieren. Während Archifiltre dafür elf verschiedene Kategorien vorsieht, unterscheidet TreeSize nach sechzehn Dateitypen. Bei beiden Tools enthält die Kategorie der nicht identifizierbaren Sonderformate eine hohe Anzahl an Dateien. Archifiltre klassiert 4'722 Dateien als «andere», TreeSize eine Anzahl von 2'674 als «unbekannte Dateitypen». DROID besitzt diesbezüglich eine wesentlich höhere Präzision. Das auf die Formatidentifizierung spezialisierte Tool identifiziert 679 Dateien, die keine PUID besitzen.

Für das Eruiere von Pfadlängen ist insbesondere das Tool TreeSize geeignet. Die Dateisuche enthält die Option, gezielt nach Dateien mit Pfadlängen von über 255 Zeichen zu suchen. Diese Anzahl kann innerhalb einer benutzerdefinierten Dateisuche variiert werden. Exportiert man die von DROID erhobenen Daten in ein Excel-Dokument, wird der Pfad jeder einzelnen Datei angezeigt. Hier sind Excel-Skills gefragt, um die Pfadlängen nach gewünschten Kriterien zu filtern. Archifiltre ist zwar imstande, den Pfad einzelner Dateien und Ordner im Balkendiagramm abzubilden, zeigt diesen aber nicht als Wert an.

Diese Suchfunktionen sind deshalb so relevant, da sowohl spezielle Dateiformate als auch umfangreiche Pfadlängen technische Herausforderungen darstellen und zu technischen Problemen führen können.

27 Bei Archifiltre ist es wichtig zu beachten, dass dieses Tool nicht imstande ist, ZIP-Formate einzulesen. Diese müssen im Voraus entpackt werden. Ebenfalls kann Archifiltre – im Gegensatz zu den anderen beiden Tools – keine Laufwerke auswerten.

5.1.3 Doppelte Dateien

Alle drei Tools bieten die Möglichkeit, mehrfach vorhandene Dateien zu identifizieren. In den meisten Fällen sind dies Doppel, die Ablage enthält aber auch einige dreifach vorhandene Dateien. Überraschend ist die Erkenntnis, dass (obwohl bei allen Tools anhand des Hashwerts beziehungsweise der MD5-Prüfsumme analysiert wurde) die Anzahl von als Dubletten erkannten Dateien erheblich variiert. Während Archifiltre 4'580 Dubletten findet, sind es bei TreeSize 6'647 und bei DROID gar 7'785. Auf die Frage, wie diese grosse Differenz zu erklären ist, kann die Verfasserin zum aktuellen Zeitpunkt keine schlüssige Antwort geben. An der Definition von doppelten Dateien sollte es nicht liegen, da die Tools mit denselben Vergleichswerten arbeiten. Kommt eine Datei mit demselben Hashwert zweimal vor, schlägt sich dies bei allen drei Tools entsprechend in der Zählweise nieder. Es wird also jede einzelne Datei gezählt und nicht als «Päckchen» gerechnet. Möglicherweise könnten versteckte, leere oder Systemdateien der Grund sein.

Die unterschiedliche Anzahl von durch die Tools eruierten Dubletten macht die Frage dringlich, ob diese überhaupt gelöscht werden sollen. Diese Problematik wird kontrovers diskutiert. Geht man davon aus, dass doppelte Dateien innerhalb einer digitalen Dateiablage gewollt sind und einen Zweck erfüllen, wäre das Löschen von Dubletten ein Eingriff in die Bestandslogik und sollte vermieden werden. Enthält die digitale Dateiablage aber zum Beispiel inhaltlich identische Ordnerstrukturen, kann das Löschen doppelter Dateien durchaus Sinn machen. Entscheidet man sich bei der Bestanderschliessung für das Eliminieren von doppelten Dateien anhand von Dubletten-Listen, wie sie von den drei untersuchten Tools generiert werden, muss mit höchster Sorgfalt vorgegangen werden. Die Dateien zu identifizieren ist nur der erste Schritt. Danach muss entschieden werden, welche Dateien gelöscht werden sollen.

5.1.4 Metadaten

Alle drei Tools erheben gewisse Metadaten. Während DROID nur Metadaten zu den Dateien generiert, werden diese bei Archifiltre und TreeSize auch für Ordner erhoben. Bei den Dateien sind Titel, Grösse, Dateityp und Hashwert Angaben, die von allen Tools erkannt werden. Archifiltre und DROID geben zudem Auskunft über das letzte Änderungsdatum. TreeSize verfügt über keine entsprechende Funktion. DROID ist imstande, weitere technische Metadaten auszule-

sen, die allesamt zu einer genaueren Bestimmung der Dateiformate beitragen. Die Frage, welche Metadaten für die Erschließung einer digitalen Dateiablage erforderlich sind, muss sorgfältig abgewogen werden.

5.2 Auswertung der Tools nach der System Usability Scale

Für die vorliegende Untersuchung wurden Personen ohne Vorkenntnisse im IT-Bereich und in der digitalen Langzeitarchivierung als Benutzergruppe definiert. Die Auswertung der gesammelten Erfahrungen anhand der SUS soll Auskunft darüber geben, ob die untersuchten Tools für diese Anwender:innen verständlich und einfach zu nutzen sind.

Das Tool Archifiltre ist intuitiv erfassbar. Durch die Darstellung mit Balkendiagrammen und den Einsatz von Farben gewinnt man schnell einen visuellen Überblick des Mengengerüsts digitaler Dateiablagen. Eckdaten zu den einzelnen Elementen sind prominent platziert.

DROID ist für Anwender:innen ohne Vorkenntnisse (auch nach dem Studium des Benutzerhandbuchs) anspruchsvoll zu bedienen. Ohne die Unterstützung einer Fachperson braucht es viel Zeit und Geduld, bis man die Funktionalitäten des Tools nutzen kann.

Die Darstellung des Tools TreeSize ist sehr übersichtlich. Der Einsatz verschiedener Diagramme und die Detailangaben zu einzelnen Elementen sind visuell ansprechend. Die zahlreichen Funktionalitäten sind zumeist selbsterklärend oder können ohne Hilfe einer Fachperson erlernt werden.

5.3 Vergleich der Tools anhand des Kriteriums der Gebrauchstauglichkeit

Das Tool Archifiltre erhielt bei meiner Auswertung nach der Methode der SUS eine Punktzahl, die auf eine fast perfekte Usability hindeutet. TreeSize kommt ebenfalls auf eine Wertung, die auf eine sehr gute Usability hinweist. DROID dagegen erhielt eine Punktzahl, die dem Tool nach SUS eine ungenügende Bewertung einbringt. Welche Schlussfolgerungen können daraus gezogen werden? Es ist wichtig, noch einmal zu betonen, dass diese Bewertung keine Aussage über die Funktionalitäten macht. Sie gibt einzig einen Anhaltspunkt, ob die Funktionen des untersuchten Tools effektiv, effizient und zufriedenstellend genutzt werden können. Die erstellte Rangliste gibt Auskunft darüber, welche Tools von Nutzer:innen ohne Vorkenntnisse im IT-Bereich schnell und (weitgehend) ohne Support durch Fachpersonen erlernt und beherrscht werden können. Dies

trifft auf Archifiltre ebenso zu wie auf TreeSize. DROID setzt hingegen gewisse IT-Kenntnisse voraus.

6 Fazit und Empfehlungen für das Archiv für Zeitgeschichte

6.1 Schlussfolgerungen

Der vorliegende Artikel thematisiert die Übernahme digitaler Dateiablagen im Allgemeinen und die technische Analyse als spezifischer Prozessschritt derselben im Besonderen. Die Aufarbeitung des Forschungsstands zum Thema und die anhand von Publikationen recherchierten Grundlagen zeigen, dass die Übernahme digitaler Dateiablagen ein – obwohl hochaktueller – noch relativ wenig erforschter Bereich innerhalb der digitalen Langzeitarchivierung darstellt. Publiziert wurden einzelne Werkstattberichte und Beschreibungen von Pilotprojekten. In diesen werden die Bearbeitungsschritte der Übernahme dargestellt und die verwendeten Tools beschrieben. Die KOST initiierte 2020 ein Projekt mit dem Ziel, einen Musterworkflow für die Übernahme digitaler Dateiablagen zu erarbeiten sowie ein Set an Best Practices und eine Liste unterstützender Tools zusammenzustellen. Im Rahmen des vorliegenden Artikels wurde die technische Analyse als ein Prozessschritt dieses Musterworkflows anhand von drei Tools erprobt. Es wurde gefragt, ob die technische Analyse mithilfe der Tools durchführbar ist und ob diese für Anwender:innen ohne IT-Vorkenntnisse als benutzerfreundlich beurteilt werden können. Untersucht wurde die digitale Dateiablage des Nachlasses von Heinz Stefan Herzka mit den Tools Archifiltre, DROID und TreeSize Professional.

Es lässt sich festhalten, dass die meisten Bestandteile der technischen Analyse – namentlich die Bestimmung des Umfangs und der Dateiformate, die Ermittlung von doppelten Dateien und das Auslesen von Metadaten – von allen drei Tools geleistet werden können. Unterschiede zeigten sich dabei in der Präzision. So ist etwa DROID imstande, die Metadaten detaillierter zu erheben. Auch findet dieses Tool eine grössere Anzahl doppelter Dateien. Die Erhebung des Mengengerüsts der Ablieferung, die ebenfalls Teil der technischen Analyse ist, kann mithilfe von Archifiltre und TreeSize erhoben und visualisiert werden. DROID bietet keine entsprechende Möglichkeit. Was die Beurteilung der Benutzerfreundlichkeit der untersuchten Tools angeht, schliessen Archifiltre und TreeSize mit sehr guten Resultaten ab. Es kann davon ausgegangen werden, dass diese beiden Tools von Anwender:innen ohne Vorkenntnisse im IT-Bereich

schnell erlernt und angewendet werden können. DROID setzt Vorwissen voraus und schneidet daher beim Usability-Test anhand der SUS wesentlich schlechter ab. Zusammenfassend bedeutet dies, dass die Tools Archifiltre und TreeSize sowohl in der Lage sind, alle Einzelschritte der technischen Analyse durchzuführen, als auch für die definierte Nutzergruppe als geeignete Instrumente gelten können.

Nimmt man den Musterworkflow der Übernahme digitaler Dateiablagen nochmals als Ganzes in den Blick, lässt sich festhalten, dass sich die Anwendungsmöglichkeiten der Tools Archifiltre und TreeSize nicht auf die technische Analyse beschränken. Denkbar wäre auch ein Einsatz dieser Tools bei der Beratung, etwa zum Aufzeigen der Grundprinzipien des Records Management. Beide Tools könnten auch für die archivfachliche Sichtung und Analyse eingesetzt werden, so etwa für das Identifizieren verschlüsselter oder passwortgeschützter Inhalte, die Analyse der Ordnerstruktur sowie für eine erste Beurteilung der Qualität der Metadaten. Die Beurteilung, ob ein Bestand übernommen werden sollte und wenn ja ob integral oder teilweise, kann durch ein Tool, das einen schnellen Überblick über digitale Dateiablagen ermöglicht, erheblich erleichtert werden. Nach der Übernahme ist die archivfachliche zusammen mit der technischen Analyse ein relevanter Bestandteil für das Erstellen eines Analyseberichts sowie eines Erschliessungskonzepts.

6.2 Empfehlungen

Die im Rahmen der Zetrifikatsarbeit durchgeführten Untersuchungen haben gezeigt, dass der Einsatz unterstützender Tools für die Übernahme digitaler Dateiablagen unerlässlich ist. Aktuell gibt es noch kein Tool, das die für die verschiedenen Prozessschritte erforderlichen Funktionalitäten vereint. Es müssen also für die einzelnen Prozessschritte je geeignete Tools ermittelt werden. In der Arbeit wurden Archifiltre, DROID und TreeSize auf ihre Eignung zur Durchführung der technischen Analyse des Nachlasses Herzka untersucht.

DROID ist ein für die Formatidentifizierung gewinnbringendes Tool, das aber nicht alle Arbeitsschritte der technischen Analyse ausführen kann. Archifiltre und TreeSize können demgegenüber alle zur technischen Analyse gehörenden Arbeitsschritte abdecken. Beide haben sich zudem als besonders benutzerfreundlich für Anwender:innen ohne Vorkenntnisse im IT-Bereich herausgestellt. Dies erscheint der Verfasserin relevant, da die Mehrheit der Archivarinnen und Archivare Geschichtswissenschaften studiert haben und über keine Ausbildung in der Informatik verfügen.

Die Untersuchungsergebnisse zeigen, dass sowohl Archifiltre als auch TreeSize bei der Übernahme digitaler Dateiablagen gewinnbringend eingesetzt werden können. Während mit Archifiltre ein mühelos interpretierbarer, visueller Überblick über die digitale Dateiablage möglich ist, besticht TreeSize durch seine übersichtliche Darstellung der relevanten Angaben und die zahlreichen Suchfunktionen. Beide Tools eignen sich nicht nur für die technische, sondern auch für die archivfachliche Analyse von Beständen. Dies bedeutet etwa für die Erschließungsplanung einen Mehrwert. Denkbar ist der Einsatz der Tools zudem für die archivfachliche Sichtung von Übernahmeangeboten, womit der Übernahmeentscheid erleichtert werden kann. Die Verfasserin schlägt demnach vor, die beiden Tools Archifiltre und TreeSize Professional als Arbeitsinstrumente einzusetzen.