

Université de Lausanne, Faculté des lettres
Universität Bern, Historisches Institut

CAS ALIS, Studiengang 2020–2022

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
Schachenstr. 15c
6010 Kriens
leonie.fritz@bluewin.ch

Betreuer: Georg Büchler
Eingereicht am: 28.7.2021

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Art. 36 Abs. 1 Buchst. r des Gesetzes über die Universität Bern und Art. 69 des Universitätsstatuts zum Entzug des aufgrund dieser Arbeiten verliehenen Titels berechtigt ist.

Kriens, 28.7.2021

A handwritten signature in blue ink, appearing to be 'L. Fritz', is centered on the page. The signature is written in a cursive style with a large initial 'L' and a stylized 'Fritz'.

Unterschrift

Inhaltsverzeichnis

Selbstständigkeitserklärung	1
1. Einleitung	4
1.1. Fragestellungen	4
1.2. Forschungsstand	5
1.3. Methode	7
1.4. Aufbau	7
2. Theoretische Grundlagen zur Übernahme digitaler Dateiablagen	8
2.1. Definition	8
2.2. Beispiele von Übernahmen digitaler Dateiablagen	9
2.3. Einsatz von Tools	10
3. Grundlagen zu Workflow und Bestand	12
3.1. Musterworkflow der KOST-Arbeitsgruppe «Dateiablage»	12
3.1.1. Beratung	12
3.1.2. Sichtung	12
3.1.3. Übernahme	13
3.1.4. Analyse	13
3.1.5. Aufbereitung	13
3.1.6. Erschliessung	14
3.1.7. Einsatz von Tools	14
3.2. Der Nachlass von Heinz Stefan Herzka	15
3.3. Nachlass Herzka im Workflow der KOST	16
4. Test der Tools	17
4.1. Archifiltre	17
4.1.1. Funktionalitäten von Archifiltre	17
4.1.2. Analyse des Nachlasses Herzka mit Archifiltre	18
4.2. DROID	21
4.2.1. Funktionalitäten von DROID	21
4.2.2. Analyse des Nachlasses Herzka mit DROID	22
4.3. TreeSize Professional	24
4.3.1. Funktionalitäten von TreeSize Professional	24
4.3.2. Analyse des Nachlasses von Herzka mit TreeSize Professional	25

5.	Funktionalitäten und Gebrauchstauglichkeit der Tools	28
5.1.	Vergleich der Funktionalitäten	28
5.1.1.	Umfang und Mengengerüst	28
5.1.2.	Dateiformate und Pfadlängen	29
5.1.3.	Doppelte Dateien	29
5.1.4.	Metadaten	30
5.2.	Gebrauchstauglichkeit der Tools	30
5.2.1.	System Usability Scale (SUS)	30
5.2.2.	Auswertung der Tools nach der SUS	31
5.2.3.	Vergleich der Tools anhand des Kriteriums der Gebrauchstauglichkeit	32
6.	Fazit und Empfehlungen fürs Archiv für Zeitgeschichte	33
6.1.	Schlussfolgerungen	33
6.2.	Empfehlungen fürs Archiv für Zeitgeschichte	34
7.	Bibliografie	35
7.1.	Literatur	35
7.2.	Internetquellen	37
8.	Anhang	39
8.1.	Abbildungsverzeichnis	39
8.2.	SUS-Auswertungen	40

1. Einleitung

Es ist ein vertrautes Bild für Archivarinnen und Archivare: Bei 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 oder man stösst auf eine Schachtel mit Audio- oder VHS-Kassetten. Zum Umgang mit diesen Datenträgern gibt es erprobte, in Workflows geregelte Strategien. Oft handelt es sich bei den Daten auf der Diskette nach genauerem Hinschauen um eine digitale Kopie von schriftlichen Unterlagen, die auch in gedruckter Form vorliegen und bei den Bildern auf dem Memorystick um klar einem Dossier zuordenbare Illustrationen einer Veranstaltung. Aktuell ist eine Veränderung auszumachen. Die Archivmitarbeitenden sehen sich zunehmend konfrontiert mit der Ablieferung von «digital born»-Daten – das heisst von 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. In der öffentlichen Verwaltung ist die Archivierung von Fileablagen spätestens seit der Einführung des digitalen Primats im Rahmen der GEVER-Verordnung vom 3.4.2019¹ ein drängendes Thema. Doch auch bei der Ablieferung von Privatnachlässen sehen sich Archivmitarbeitende zunehmend mit grossen und teils chaotischen elektronischen Ablagen konfrontiert, die Dokumente in verschiedensten Formaten enthalten.

1.1. Fragestellungen

Informationen die ausschliesslich digital erstellt und veröffentlicht werden, stellen für alle Archive eine grosse und hochaktuelle Herausforderung dar. Daher hat die «Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen» (KOST), als Kompetenzzentrum der Schweizer und Liechtensteiner Archive, 2020 die Arbeitsgruppe «20-039 Dateiablage» initiiert. Diese beschäftigt sich mit der Erarbeitung eines Workflows für die Übernahme digitaler Dateiablagen und erprobt für den Übernahmeprozess geeignete Tools.² Die vorliegende Zertifikatsarbeit orientiert sich an den Resultaten dieser 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 orientiert sich an folgenden Fragestellungen:

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 29.6.2021).

² Vgl. <https://kost-ceco.ch/cms/20-039-dateiablage.html> (Zugriff 29.6.2021).

- 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 in Frage 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 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 schreibt Stephan Lenartz in einer 2020 erschienenen Publikation, 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

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.

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: Jörg Filthaut (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.

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 Zugänglichmachung 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 allgemein gültiger Workflow dafür beschrieben wird. Die aktuellste und umfangreichste Publikation, welche (nebst dem ausführlich beschriebenen Fallbeispiel der Aufbereitung und Bewertung einer digitalen Fotosammlung insbesondere mithilfe der Programmiersprache Python) die Grundlagen der Übernahme digitaler Dateiablagen thematisiert, ist die eingangs erwähnte Studie von Lenartz.¹³

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, Annkathrin; 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.

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.

1.3. Methode

Der im Rahmen des KOST-Projekts «20-039 Dateiablage» erarbeitete und noch unpublizierte Musterworkflow beschreibt einen idealtypischen Archivierungsprozess digitaler Dateiablagen anhand von sechs Prozessschritten. In der vorliegenden Arbeit wird die technische Analyse als ein Teilschritt 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 geeignet. Zur Bestimmung ihrer Usability wird die empirische Methode der «System Usability Scale» (SUS) angewandt.

1.4. Aufbau

Die Zertifikatsarbeit ist in sechs Kapitel gegliedert. Im ersten Kapitel wurden die Fragestellungen, der Forschungsstand und die Methode erläutert. Im zweiten Kapitel werden die theoretischen Grundlagen zur Übernahme digitaler Dateiablagen besprochen. Hier wird auch der Einsatz von unterstützenden Tools thematisiert. Das dritte Kapitel legt einen Schwerpunkt auf die Beschreibung des Musterworkflows der KOST-Arbeitsgruppe «Dateiablage» allgemein und der technischen Analyse im Besonderen. Zudem wird die Akzession des als Beispiel verwendeten Nachlasses von Herzka umrissen und im Workflow verortet. Im vierten Kapitel werden die erprobten Tools beschrieben und die Erfahrungen mit den, im Rahmen der technischen Analyse, geprüften Funktionalitäten reflektiert. Das fünfte Kapitel umfasst den Vergleich der Funktionalitäten der Tools und den Vergleich ihrer Gebrauchstauglichkeit anhand der SUS. Das sechste und abschliessende Kapitel widmet sich der Frage, welche der Tools in der praktischen Umsetzung besonders gut abgeschnitten haben. Es umfasst ebenfalls die Frage, was die einzelnen Tools leisten können und wo ihre Grenzen liegen beziehungsweise welche Prozessschritte des Musterworkflows zur Übernahme digitaler Dateiablagen sie abdecken (können). Die Arbeit schliesst mit Empfehlungen von Tools, welche die Übernahme digitaler Dateiablagen im Archiv für Zeitgeschichte unterstützen könnten.

¹⁴ Nestor: Toolbox, vgl. <https://wiki.dnb.de/pages/viewpage.action?pageId=134715087> (Zugriff 24.7.2021).

2. Theoretische Grundlagen zur Übernahme digitaler Dateiablagen

In diesem Kapitel werden die Grundlagen der Übernahme digitaler Dateiablagen vermittelt. Anhand von drei, der Literatur entnommenen, Beispielen solcher Übernahmen wird die Heterogenität der jeweiligen Projektausgangslagen aufgezeigt. Ebenfalls wird der Einsatz von unterstützenden Tools diskutiert.

2.1. Definition

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 nicht-amtliche Überlieferungen, beispielsweise für 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 umfangsreich sind.¹⁸

Daraus ergeben sich die Charakteristika,

- dass digitale Dateiablagen in der Regel eine unübersichtliche Anzahl an Einzel-Dateien enthalten;
- dass die Dokumente in zahlreichen, unterschiedlichen Formaten abgespeichert sind;
- dass archivwürdige und nicht-archivwürdige Dateien nebeneinanderliegen;

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.

- und 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

2.2. Beispiele von Übernahmen digitaler Dateiablagen

Entsprechend heterogen gestalten sich die Ausgangslagen der in Werkstattberichten und Pilotprojekten beschriebenen Übernahmen digitaler Dateiablagen. Im Folgenden werden drei diesbezüglich interessante Beispiele herausgegriffen und kurz beschrieben.

Jaeger und Kobold beschreiben in einem 2017 erschienenen Artikel die Übernahme des Bestands einer Privatschule vom Hessischen Staatsarchiv Darmstadt. Die Odenwaldschule – einst reformpädagogische Vorzeigeanstalt – musste ihren Betrieb einstellen, nachdem 2015 der systematische Missbrauch von Kindern und Jugendlichen bekannt geworden war. Aufgrund der Brisanz dieser Ereignisse standen die Übernahme und Aufbereitung der Dokumente der Odenwaldschule von Beginn an im Fokus des öffentlichen Interesses. «Die Sicherung und der Nachweis der Integrität und Authentizität der Unterlagen» seien «zwingende Vorgabe für den Prozess der Bearbeitung bis hin zur Aufbewahrung» gewesen. Es galt eine grosse Menge (insgesamt rund 0.9 TB Daten bzw. 1'453'113 Dateien) zu sichern, aufzubereiten, zu bewerten und – soweit archivwürdig – zu erschliessen. Dies habe «zu den bisher größten Herausforderungen in fachlicher und technischer Hinsicht» gehört.²² Schritt für Schritt wird im Artikel beschrieben, mit welchen Problematiken die Archivarinnen und Archivare konfrontiert waren, wie sie diesen begegneten und welche Lösungsansätze erprobt und umgesetzt wurden. Abschliessend wird festgehalten, dass «das Projekt Odenwaldschule aufgrund seiner Größe und Heterogenität von Beginn an viele Möglichkeiten bot, Erfahrungen mit der Sichtung kreativer Dateiablagen und Programmen zur Bearbeitung dieser zu sammeln.»²³

Taylor reflektiert in einem 2018 erschienenen Beitrag die Übernahme des Bestandes eines Staatlichen Schulamtes durch das Generallandesarchiv Karlsruhe im Jahr 2015. Es handelte sich dabei um 18 GB Daten bzw. 57'000 Dateien. Die Ablage des Staatlichen Schulamtes stellte sich als chaotisch heraus. Es existierten etwa zahlreiche Ordner mit identischer Be-

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 Jaeger; Kobold: Werkstattbericht zum Umgang mit unstrukturierten Dateisammlungen, 2017, S. 107f.

²³ Ebd., S. 311.

nennung und/oder identischem Inhalt. Die Ordnerstruktur war unübersichtlich und innerhalb der Ordner lagen zusammengehörige Dokumente nicht beieinander.²⁴ Aufgrund dieser Komplexität wurde schnell klar, dass die Übernahme dieses Bestandes hohe Kosten verursachen würde. Daher wurde die Grundsatzdiskussion geführt, ob auf die inhaltliche Bewertung der Unterlagen verzichtet werden könne, um damit personelle Kosten einzusparen. Taylor entschied sich schliesslich aus Gründen der Benutzerfreundlichkeit für eine Bewertung.²⁵ Sie resümiert: «The time investment was justified because this was a pilot project, but in the future the effort must be significantly reduced».²⁶ Potential dafür liegt nach Taylor insbesondere in der Entwicklung von Tools, die Funktionalitäten zur Ausführung verschiedenster Arbeitsschritte vereinen und dabei stets die Authentizität der Daten gewährleisten.²⁷

Steigerwald beschreibt in einem 2019 publizierten Artikel das Pilotprojekt der Übernahme einer digitalen Dateiablage des Ministeriums der Finanzen im Bundesland Sachsen-Anhalt. Anders als bei den ersten beiden Beispielen, in denen ein abgeschlossener Bestand als Schenkung ins Archiv kam, war in diesem Fall das Landesarchiv Sachsen-Anhalt schon intensiv an den Vorbereitungen der Übernahme beteiligt. Die archivwürdigen Daten konnten definiert und so eine Vorauswahl getroffen werden. Dadurch konnte die Datenmenge bereits etwa um die Hälfte verringert werden. Die 2018 erfolgte Ablieferung umfasste schliesslich 14.5 GB bzw. 44'597 Dateien.²⁸ Im Archiv erfolgte dann die Feinbewertung der Dokumente mithilfe verschiedener Tools. Steigerwald hält fest, dass diese Werkzeuge zur Unterstützung der archivfachlichen Bewertung sinnvoll seien, diese aber nicht ersetzen.²⁹

Diese Beispiele illustrieren, dass ein standardisierter Workflow und Tools, welche die Bearbeitung digitaler Dateiablagen unterstützen, notwendig und gewünscht sind, doch aufgrund der Heterogenität digitaler Dateiablagen nicht allgemeingültig sein können.

2.3. Einsatz von Tools

Die anhand des Beitrags von Taylor diskutierte Frage, ob digitale Dateiablagen auch ohne Bewertung auskämen, wird von Lenartz in einer 2020 erschienenen Publikation ebenfalls aufgegriffen. Wie Taylor kommt auch er zum Schluss, dass die archivische Bewertung von Dateiablagen zur Informationsverdichtung notwendig sei. Ziel sollte seiner Einschätzung nach «eine softwaregestützte Vorgehensweise sein, die den Archivar entlastet, indem Bewertungs- und weitere Aufbereitungsprozesse so weit wie möglich verkürzt und wo möglich automatisiert werden». Dies jedoch immer unter der Voraussetzung, dass sämtliche Schritte

²⁴ Taylor: A Hydra-like Russian Doll, 2018, S. 150f.

²⁵ Taylor: A Hydra-like Russian Doll, 2018, S. 152–154.

²⁶ Ebd., S. 158f.

²⁷ Ebd.

²⁸ Steigerwald: Erste Übernahmen digitaler Daten aus der Landesverwaltung, 2019, S. 10.

²⁹ Ebd., S. 11.

intellektuell nachvollziehbar und korrigierbar bleiben.³⁰ Insgesamt zeige sich, dass sich dank des breiteren Erfahrungsaustausches die Übernahme digitaler Dateiablagen betreffend gewisse bewährte Tools wie TreeSize durchzusetzen begännen.³¹

Knobloch beschreibt in einem 2019 erschienenen Artikel den Versuch, in einer überschaubaren Zeit die Bewertung und archivische Aufbereitung einer digitalen Dateiablage vorzunehmen. Es handelte sich dabei um den Bestand eines staatlichen Internats für hörgeschädigte Kinder, der Johannes-Wagner-Schule in Nürtingen, der dem Staatsarchiv Ludwigsburg 2016 übergeben wurde. Dieser umfasste 677 GB bzw. 65'057 Dateien.³² Knobloch legte einen Schwerpunkt auf die Beschreibung der Bewertung. Gewisse Schritte wurden manuell vorgenommen, andere sollten mithilfe einer Software ausgeführt werden. Um herauszufinden, welche Software die gestellten Anforderungen am besten erfüllte, wurden zehn verschiedene Tools erprobt. Dabei schloss sowohl bei Tests mit Ordnern als auch mit Dateien die Software TreeSize am besten ab. Sie habe durch umfangreiche Einstellmöglichkeiten für die Suche und Löschung überzeugt und sehr zuverlässig gearbeitet.³³

Auch Miegel, Schieber und Schmidt betonen die Bedeutung von Analysetools. Schon für die Bestimmung der Archivwürdigkeit sei es unerlässlich, einen guten Überblick über den Bestand zu gewinnen.³⁴ Wenn sich bei der ersten Sichtung herausstelle, dass eine Kassation einzelner Teile «fachlich sinnvoll und ökonomisch möglich» sei, könne die softwareunterstützte Auswahl nach technischen Kriterien eine Alternative oder Ergänzung zur klassischen, intellektuellen Bewertung darstellen.³⁵

Im französischen Magazin «Archimag» erschien im März dieses Jahres ein Interview zu praktischen Fragen der Aufbereitung digitaler Dateiablagen. Darin wird betont, dass es sich lohne, den (mitunter hohen) Aufwand der Bestandsanalyse auf sich zu nehmen. «Une fois effectué, ce travail permet de dégager les tendances et les grandes orientations de travail pour le traitement du vrac.»³⁶ Es wird erwähnt, dass für verschiedene Prozessschritte effiziente Tools existieren.³⁷ Auch die Erfahrungs- und Werkstattberichte machen deutlich, dass eine grosse Nachfrage nach unterstützenden Tools besteht – nicht nur für die Bewertung, sondern für sämtliche Prozessschritte.

³⁰ Lenartz: Digital ist besser?, 2020, S. 18.

³¹ Lenartz: Aufbereitung und Bewertung von Fileablagen, 2018, S. 5.

³² Knobloch: Archivischer Umgang mit digitalen Sammlungen, 2019, S. 76.

³³ Ebd., S. 78f.

³⁴ Miegel; Schieber; Schmidt: Vom richtigen Umgang mit kreativen digitalen Ablagen, 2017, S. 10.

³⁵ Ebd., S. 11.

³⁶ Texier, Bruno: Traiter un vrac numérique – mode d'emploi, in: archimag, Nr. 342/mars 2021, Varennes-Vauzelles 2021, S. 17.

³⁷ Ebd.

3. Grundlagen zu Workflow und Bestand

In diesem Kapitel werden die Prozessschritte der Übernahme digitaler Dateiablagen thematisiert. Diese werden anhand des im Rahmen der KOST-Arbeitsgruppe «Dateiablage» erarbeiteten Musterworkflows im ersten Unterkapitel aufgezeigt. Im zweiten Unterkapitel wird ein konkretes Übernahmebeispiel erläutert, nämlich die Akzession des Nachlasses von Herzka, dessen digitaler Bestandteil Übungsfeld dieser Arbeit ist. Abschliessend wird dieser im Musterworkflow verordnet.

3.1. Musterworkflow der KOST-Arbeitsgruppe «Dateiablage»

Der bislang unveröffentlichte Musterworkflow wurde von einer Arbeitsgruppe aus Mitarbeitenden zehn verschiedener schweizerischer Archivinstitutionen unter Leitung der KOST gemeinschaftlich und auf bislang gesammelten Erfahrungen beruhend, erarbeitet. Er beschreibt die Übernahme digitaler Dateiablagen anhand von sechs Prozessschritten. Ein wichtiger Bestandteil sind Empfehlungen von Tools, die einen oder mehrere Arbeitsschritte unterstützen können. Der Workflow soll in den kommenden Monaten erprobt werden. Ziel ist es, abschliessend ein Set an Best Practice zusammenzustellen und zusammen mit dem Musterworkflow im Sinne von Empfehlungen allen Institutionen zugänglich zu machen. Der Workflow ist in die Prozessschritte Beratung, Sichtung, Übernahme, Analyse, Aufbereitung und Erschliessung unterteilt. Diese sind jeweils in bis zu fünf Unterschritten aufgegliedert.

3.1.1. Beratung

Die Beratung potenzieller Anbieter (juristische und natürliche Personen oder Amtsstellen) erfolgt vor dem Zeitpunkt des Aktenangebots. Ziel der Beratung ist es unter anderem, die Anbieter mit den Grundprinzipien des Records Management vertraut zu machen, damit das Geschäftsarchiv nach diesen Prinzipien aufgebaut werden kann. Hybride Ablagen sollten nach Möglichkeit vermieden werden. Auch gilt es, den Inhalt und den Ordnungszustand der zentralen Ablage sowie geltende Aufbewahrungs- und Datenschutzfristen zu kennen. Abschliessend können Zieltermine für weitere Besprechungen beziehungsweise das Angebot definiert werden.

3.1.2. Sichtung

Die Sichtung ist in fünf Schritte unterteilt.³⁸ Im Rahmen der *Vorbereitung des Aktenangebots* geht es darum, Inhalt und Struktur der abzuliefernden Daten zu besprechen und sicherzustellen, dass die Inhalte nicht passwortgeschützt oder verschlüsselt sind. Die *archivfachliche Sichtung* umfasst die Vorbereitungen für den Transfer. Teil davon ist es, die Provenienz der Unterlagen zu klären, einen groben Überblick über die Ordnerstruktur und die Ordnerinhalte zu gewinnen, die Qualität der Metadaten zu beurteilen und ein Sichtungsverzeichnis zu erstellen. Auf dieser Grundlage können bereits eine Vorbewertung des Materials und erste Kassationen stattfinden. Die *technische Sichtung* umfasst die technischen Aspekte der Vor-

³⁸ Die Unterschritte werden nachfolgend in kursiver Schrift kenntlich gemacht.

bereitungen für den Transfer. Es gilt die Gesamtgrösse der zu übernehmenden Daten zu definieren und sich eine Übersicht über vorhandene Formate und Sonderformate sowie Pfadlängen zu verschaffen. Nach einer *Rücksprache mit dem Aktenproduzenten*, bei der man über Fragen der Bewertung und den Umgang mit unerwünschten Datenformaten übereingekommen ist, folgt der *Übernahmeentscheid*.

3.1.3. Übernahme

Der Prozessschritt der Übernahme beinhaltet drei Unterpunkte. Die *Übergabe* umfasst den Transfer der Daten ins Archiv, die Sicherung (und notfalls Rettung) der elektronischen Daten und damit die Trennung von den ursprünglichen Datenträgern. Im Archiv wird ein Virenscheck durchgeführt, die Checksummen berechnet, ein Container erstellt und ein Backup generiert. Die Checksummen als eindeutige Kennung jeder Datei sind nach Naumann der Nachweis, dass keine der abgelieferten Daten im Nachhinein versehentlich oder bewusst gelöscht oder verändert wurden.³⁹ Die *Sicherung* umfasst eine Integritäts- und Vollständigkeitsprüfung und eine zentrale Sicherung der Daten im Archiv. Die *Dokumentation* umfasst unter anderem das Festhalten der bisher durchgeführten Arbeitsschritte, ein Eintrag ins Archivinformationssystem und die Übernahmebestätigung beim Aktenproduzenten. Auch soll hier geregelt werden, dass die abliefernde Person oder Stelle die übergebenen Daten bei sich löscht oder passiviert, um damit zu verhindern, dass diese dem Archiv zu einem späteren Zeitpunkt erneut angeboten werden.

3.1.4. Analyse

Die Analyse 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 und der Definition von Sofortmassnahmen. Die *inhaltliche* oder *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.

3.1.5. Aufbereitung

Die Aufbereitung gliedert sich in zwei Unterpunkte. Es ist dies einerseits die *Umsetzung technischer Massnahmen*. Diese umfasst unter anderem die Kassation nicht für die Archivierung bestimmter Dateien und Dubletten, die Normalisierung von Dateinamen und Pfadlängen, die Validierung der Formate und die Konvertierung in für die digitale Langzeitarchivierung geeignete Formate⁴⁰. Andererseits das *Erarbeiten einer Erschliessungsplanung*, in der eine Struktur vorgeschlagen und die Erschliessungstiefe definiert wird, sowie eine Kostenschätzung von Erschliessung und Datenspeicherung erfolgt.

³⁹ Naumann: Welche Schritte erfordert die Aufbereitung von Dateisammlungen, 2017, S. 47.

⁴⁰ Vgl. dazu etwa Lindlar, Micky: Dateiformatidentifizierung. Was, warum, wie, womit, ..., Referat am nestor-Praktikertag 2021 «Formate – Erkennen, Validieren und Behandeln», 9.6.2021; Röthlisberger-Jourdan, Claire:

3.1.6. Erschliessung

Die abschliessende Prozessphase bildet diejenige der Erschliessung der übernommenen digitalen Dateiablage. Diese ist unterteilt in vier Schritte. Bei der *Paketbildung und Metadatering* werden die übernommenen Daten geordnet und strukturiert und danach «Submission Information Packages» (SIP) gebildet.⁴¹ Dies umfasst auch die Erfassung von technischen und beschreibenden Metadaten. Bei der *Magazinierung* der digitalen Daten, dem sogenannten *Ingest* werden die «Archival Information Packages» (AIP) gebildet und ins Repository für die Digitale Langzeitarchivierung verschoben. Die Metadaten werden ins Archivinformationssystem importiert und dort verlinkt. Bei der *Nacherschliessung* geht es in erster Linie darum, die Metadaten zu vervollständigen. Auf dieser Grundlage werden die Zugriffsrechte definiert. Abschliessend wird ein Findmittel generiert. Es folgen *Qualitätssicherung und Abschluss* der Erschliessungsphase, in der die Erschliessung aufgrund der Verzeichnungskontrolle bereinigt wird.

3.1.7. Einsatz von Tools

Die meisten dieser Arbeitsprozesse der Bewertung und Erschliessung digitaler Sammlungen kamen bereits bei der Bearbeitung analoger Materialien zum Einsatz. «Alleine der Virensan, die Erstellung eines Backups, Formatmigration, die Erzeugung von Primär- und Metadaten für das digitale Archivsystem und die Validierung kommen neu hinzu. Bei allen anderen Arbeitsschritten verändern sich nur die Arbeitsweisen und Abläufe», resümieren Birn und Naumann. Neu hinzu kämen zudem «die zahlreichen Tools, die den Archivar [...] unterstützen und auch zur Automatisierung mancher Teilprozesse beitragen».⁴² Wichtiger Bestandteil des detailliert beschriebenen Musterworkflows ist das Eruiieren solcher Tools. Jedem Prozessschritt ist eine Aufzählung von Werkzeugen beigefügt, die dafür verwendet werden können.⁴³

Formatauswahl. Denn sie wissen was sie tun, Referat am nestor-Praktikertag 2021 «Formate – Erkennen, Validieren und Behandeln», 9.6.2021.

⁴¹ Das OAIS unterscheidet zwischen «Submission Information Packages» (SIP) – das, was die Archive übernehmen –, «Archival Information Packages» (AIP) – die nach Anreicherung und Weiterverarbeitung im Archiv in einer langfristig aufbewahrbaren Form vorliegen – und «Dissemination Information Packages» (DIP) – das, was Archivbenutzenden, je nach rechtlicher Bedürfnisse, zielgruppenorientiert zugänglich gemacht wird. Vgl. Neuroth: Nestor Handbuch, 2010, S. 82.

⁴² Birn; Naumann: Bewertung schwach strukturierter Unterlagen, 2019, S. 13f.

⁴³ Eine aktuelle Übersicht über Tools für den Bereich der digitalen Langzeitarchivierung findet sich auch in der Nestor-Toolbox unter <https://wiki.dnb.de/display/NESTOR/Toolbox> (Zugriff 2.7.2021).

3.2. Der Nachlass von Heinz Stefan Herzka

2007 übergab Heinz Stefan Herzka – Kinder- und Jugendpsychiater sowie Professor an der Universität Zürich, 1935 in Wien geboren – dem AfZ seinen Vorlass. Dieser enthält biografische Unterlagen und Korrespondenz, Materialien zu seiner Forschungs- und Lehrtätigkeit sowie eine grosse Anzahl Publikationen. Schon damals vereinbarten der Donator und das AfZ, dass weitere Materialien zu einem späteren Zeitpunkt folgen würden. 2008 und 2010 übergab Herzka dem Archiv zusätzliche Unterlagen zu seinen Tätigkeiten. 2015 stand eine Nachlieferung von etwa 3.5 Laufmetern analogem Material an. Als die zuständige Person dies bei ihm abholte, übergab Herzka zusätzlich eine Harddisc. Die überraschte Kollegin nahm diese entgegen ohne zu wissen, was der Speicher genau enthielt. Die Sicherung der elektronischen Daten ergab, dass es sich um etwa 700 GB Daten handelte.

Für die Zuständigen im AfZ war klar, dass diese grosse Datenmenge aufgrund der für die Archivierung nötigen Nachfolgearbeiten nicht komplett archiviert werden konnte. Der Donator wurde also mit dem Anliegen kontaktiert, das Archiv bei der Bewertung zu unterstützen, indem er mithilfe zu eruieren, welches Material zentral für seinen Vorlass sei. Herzka nahm die Harddisc nochmals mit und erreichte durch zahlreiche Streichungen und die Eliminierung grosser Doppel eine Reduktion der Datenmenge auf 400 GB. 2016 übergab der Donator dem Archiv die bereinigte Harddisc. Bei dieser Gelegenheit erfolgte eine sorgfältige Abklärung über den Inhalt der digitalen Dateisammlung. Es ging dabei insbesondere um die Frage nach dem Verhältnis des analogen und des digitalen Materials, also welche Datenversion zu welchem Zeitpunkt die vollständig(er)e sei und wo sie sich überschneiden. Auch das Vorgehen bei weiteren digitalen Ablieferungen war Thema der Besprechung. Die zuständige Mitarbeiterin vereinbarte mit dem Donator, dass er bereits abgelieferte und neu angelegte Daten konsequent trenne. So sollten bei einer nächsten Ablieferung nur die neuen Daten übernommen werden können. 2019 meldete sich Herzka beim AfZ mit dem Hinweis, er habe die abgelieferte Datenmenge in der Zwischenzeit nochmals reduziert. Die bereinigte (nunmehr rund 175 GB umfassende) Harddisc wurde unter der Vereinbarung übernommen, dass diese neu die massgebende Version bilde und die früher abgelieferten Daten ersetze. Auch wurde festgelegt, dass ein Abgleich mit dem analogen Teil des Bestands zu aufwendig wäre. Also entschied man – sich der dadurch entstehenden teilweise doppelten Überlieferung bewusst – auf einen Abgleich zwischen analogen und digitalen Daten zu verzichten. Zudem bot das AfZ Herzka eine Records-Management-Beratung an. Die Durchführung dieser war nochmals eine Gelegenheit, detailliert über die abgelieferten Unterlagen zu sprechen. Es ging dabei unter anderem um die Frage nach Datensicherungen, die relevanten Ablageorte der Daten, den Aufbau der Ablagestruktur, die Logik und Benennung der Ordner und Dokumente, Metadaten und die Frage nach der Archivierung von E-Mails. Zudem wurde geklärt, was mit den digitalen Daten geschehe, wenn er «sie selbst sie nicht mehr nutzen» könne. Teil der Abmachung war es, dass der Donator die Vorkehrungen dafür treffe, dass auch die neuen Daten dem Archiv abgeliefert würden. Herzka verstarb am 14.2.2021. Bislang wurden dem AfZ keine weiteren Unterlagen übergeben.

Die Frage nach dem Umgang mit der Hybridität von Beständen muss bei jeder Übernahme individuell beurteilt werden. Entschloss man sich beim Nachlass Herzka aus Überlegungen des Aufwands gegen einen Abgleich, beschreiben Kramski und von Bülow in einem 2001 erschienenen Artikel vom detaillierten Abgleich der analogen und digitalen Daten im Falle der Übernahme eines Schriftstellernachlasses.⁴⁴

3.3. Nachlass Herzka im Workflow der KOST

Die ausführliche Beschreibung der Übernahme des Nachlasses von Herzka macht deutlich, dass die Akzession dem im Musterworkflow vorgesehenen (als idealtypisch zu verstehenden) Ablauf von Prozess- und Bearbeitungsschritten nicht folgte.⁴⁵ So gelangten die digitalen Unterlagen ohne jegliche Beratung oder Sichtung ins Archiv. Das Material wurde übernommen und gesichert. Ohne die Notwendigkeit einer tieferen Analyse wurde deutlich, dass die Datenmenge die Möglichkeiten der integralen Archivierung überstieg. Indem der Donator kontaktiert und um eine Unterstützung bei der Reduktion der Datenmenge gebeten wurde, sollte der Prozessschritt der Vorbereitung des Aktenangebots nachgeholt werden. Bei der Übergabe der reduzierten Ablieferung wurde in einem ausführlichen Gespräch eine archivfachliche und eine technische Ersteinschätzung vorgenommen. Der daraufhin erfolgte Übernahmeentscheid wurde drei Jahre später vom Donator überraschend revidiert, indem er die Ablieferung in einer wiederum überarbeiteten Form dem Archiv übergab. Zu diesem Zeitpunkt bot man Herzka eine Records-Management-Beratung an, die gleichzeitig zur Klärung von Fragen der archivfachlichen Sichtung genutzt wurde. Da der Donator kurz darauf verstarb ist fraglich, ob er überhaupt noch weitere digitale Unterlagen erstellte. Man kann also davon ausgehen, dass die ersten drei Phasen des Musterworkflows – die Beratung, die Sichtung und die Übernahme – als abgeschlossen gelten können. Während der analoge Bestandteil des Nachlasses erschlossen ist, blieb der digitale Anteil aufgrund von mangelnden Strategien und Tools bislang unbearbeitet.

⁴⁴ Kramski; Heinz Werner; von Bülow, Ulrich: «Es füllt sich der Speicher mit köstlicher Habe» – Erfahrungen mit digitalen Archivmaterialien im Deutschen Literaturarchiv Marbach, in: Robertson-von Trotha, Caroline Y.; Hauser, Robert (Hg.), Neues Erbe. Aspekte, Perspektiven und Konsequenzen der digitalen Überlieferung, Karlsruhe 2001, S. 154.

⁴⁵ Es ist davon auszugehen, dass dies auf die allermeisten Übernahmen privater Nachlässe zutrifft.

4. Test der Tools

Die im dritten Kapitel innerhalb des Musterworkflows der KOST-Arbeitsgruppe «Dateiablage» verortete technische Analyse wird nachfolgend anhand von Archifiltre, DROID und TreeSize erprobt.

4.1. Archifiltre

Archifiltre wurde in einem von der «fabrique numérique des ministères sociaux» finanzierten Projekt entwickelt. Beteiligt waren Archivarinnen und Archivare des französischen Sozialministeriums und IT-Fachleute. Die erste Version der kostenlosen Open-Source-Software wurde 2018 veröffentlicht.⁴⁶

4.1.1. Funktionalitäten von Archifiltre

Archifiltre ist ein Werkzeug zur Visualisierung digitaler Dateiablagen. Es ermöglicht die Analyse der Bestandsstruktur sowie der vorhandenen Dateien. Ebenfalls erlaubt es, diese zu prüfen, zu sortieren und zu bearbeiten.⁴⁷ 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 (vgl. Abbildung 1). Fährt man mit der Maus ins Balkendiagramm und verharrt 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.



Abb. 1: Farbskala nach Dateitypen⁴⁸

⁴⁶ Software Archifiltre: «Qui sommes-nous?», vgl. <https://archifiltre.fabrique.social.gouv.fr/qui-sommes-nous> (Zugriff 13.7.2021).

⁴⁷ GitHub: Wiki Archifiltre, vgl. <https://github.com/SocialGouv/archifiltre/wiki/Wiki-Archifiltre> (Zugriff 13.7.2021).

⁴⁸ Ebd.

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.

4.1.2. Analyse des Nachlasses Herzka mit Archifiltre

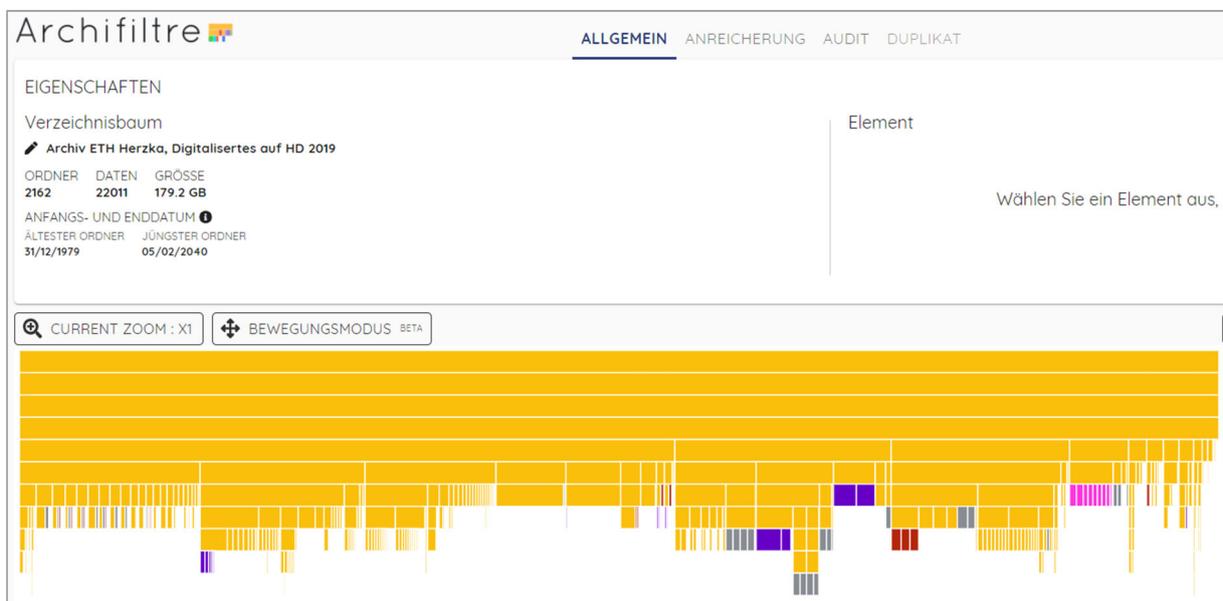


Abb. 2: Benutzeroberfläche mit Übersicht über die digitale Dateiablage

Dass die digitale Dateiablage des Nachlasses Herzka eine tiefe Strukturierung besitzt, wird bei der Visualisierung derselben mithilfe von Archifiltre auf den ersten Blick deutlich. Der in Abbildung 2 (in der allgemeinen Benutzeroberfläche von Archifiltre) als Balkendiagramm dargestellte Bestand erscheint grossmehrheitlich in gelber Farbe, was die grosse Anzahl von Ordnern und Unterordnern zeigt. Das Feld «Eigenschaften» gibt einen Überblick über die Grösse der zu analysierenden Ablage (179.2 GB), die Anzahl vorhandener Dateien (22'011) und vorhandener Ordner (2'162). Ebenfalls angezeigt werden die Daten des ältesten Ordners (31.12.1979) und des jüngsten (5.2.2040).⁴⁹

⁴⁹ Es wäre interessant zu eruieren, wie diese Angaben zustande kommen. Das in der Zukunft liegende Datum des jüngsten Ordners ist offensichtlich falsch. Dasjenige des ältesten Ordners muss aufgrund der Angabe seiner Erstellung an einem 31. Dezember hinterfragt werden.



Abb. 3: Auswahl eines Unterelements innerhalb der Dateiablage (Bildausschnitt)

Die Darstellung der Dateiablage als Balkendiagramm wird von oben nach unten gelesen, wie in Abbildung 3 ersichtlich ist. Mit dem Klick auf ein Unterelement verändert sich die Ansicht. Im oberen Teil finden sich die Metadaten des ausgewählten Unterelements. Bei der Abbildung handelt es sich um ein PDF-Dokument mit dem Titel «1519_I_1-Teil Herzka» und einer Grösse von 1.3 GB. Innerhalb der Grafik wird der Pfad hin zur gewählten Einheit hervorgehoben. Rechts davon sieht man die Namen der Ordner und Unterordner des Pfades. Unterhalb der Grafik ist angegeben, wie viel Prozent der Gesamtgrösse des Bestands diese Untereinheit ausmacht (0.7 %). Rechts unten sieht man, wo der gezoomte Bereich sich innerhalb der gesamten Dateiablage befindet.



Abb. 4: «Audit» des Bestands

Abbildung 4 zeigt die Ebene «Audit». Auf der linken Seite werden die Eckdaten angegeben: die Anzahl Dateien und die Tiefe der Bestandsgliederung (16 Ebenen). Auf der rechten Seite werden die vorhandenen Dateitypen anhand der Farbskala⁵⁰ dargestellt. Die Maus liegt im türkisen Farbbereich, wodurch die Details zu diesem angezeigt werden – nämlich, dass die Ablage 599 E-Mail-Dateien enthält, was einer Gesamtgrösse von 98 MB entspricht.

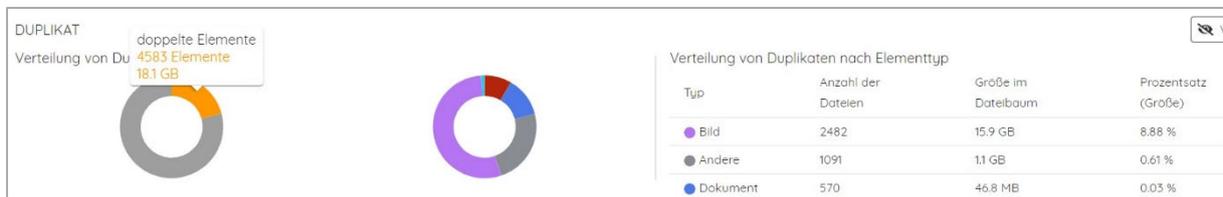


Abb. 5: Übersichtsdarstellung über Dateiduplikate

Abbildung 5 zeigt die Ansicht der Dublettensuche innerhalb der analysierten digitalen Dateiablage. Archifltre vergleicht die Dateien anhand ihres Hashwerts. Das Kreisdiagramm links gibt Auskunft über das Verhältnis von Dubletten (rund 4'580 Elemente, was rund 18 GB des Gesamtvolumens entspricht) zu Dateien und Ordnern, die nicht doppelt vorkommen (rund 17'420 Elemente, was rund 161 GB entspricht). Das Kreisdiagramm rechts zeigt die Verteilung der Dubletten nach Dateityp. Auf der rechten Seite sind diese Informationen in Tabellenformat dargestellt. Es wird ersichtlich, dass es sich bei den Dubletten zu einem hohen Anteil um doppelte Bilder handelt (rund 2'480).

Name	Typ	Größe	Anzahl der Dateien	Letzte Änderung	Pfad	Zu löschen
Adressen Steinerschule klasse 2014.pdf	pdf	228.9 kB	2	20/06/2014		Zu löschen
Adressen Steinerschule klasse 2014.pdf	pdf	228.9 kB		20/06/2014	/Archiv ETH Herzka, Digitalisiertes auf HD 2019/ADRESSE N/Adressen Steinerschule klasse 2014.pdf	Zu löschen
Adressen Steinerschule klasse 2014.pdf	pdf	228.9 kB		20/06/2014	/Archiv ETH Herzka, Digitalisiertes auf HD 2019/Planung & Kalender 2007 bis 2015_/2014 Planen/2014-08-28 Steinerschultreffen/Adressen Steinerschule klasse 2014.pdf	Zu löschen

Abb. 6: Detailansicht der Dateiduplikate

Abbildung 6 zeigt die Detailansicht der Dublettensuche. Das herausgegriffene Beispiel umfasst drei Zeilen. Die oberste zeigt, welches Element als Doppel identifiziert wurde. Die beiden darunterliegenden Zeilen sind die eigentlichen Doppel, wobei die verschiedenen Ablageorte am Pfad erkennbar sind.

⁵⁰ Vgl. Abb. 1.

4.2. DROID

DROID steht als Abkürzung für «Digital Record Object Identification». Es wurde vom Department für Digital Preservation des National Archives entwickelt – einer britischen Behörde, zuständig unter anderem für das Nationalarchiv des Vereinten Königreichs. Die erste Version wurde 2005 veröffentlicht.⁵¹ Die kostenlose und quellenoffene Software arbeitet mit einer PRONOM-Datenbank.⁵² Darin wird den Formaten eine sogenannte PUID (PRONOM Unique Identifier) vergeben. Das Tool ist imstande, etwa 1'800 Dateiformate zu identifizieren.⁵³

4.2.1. Funktionalitäten von DROID

Die wichtigste Funktion von DROID ist die Formatidentifizierung innerhalb einer digitalen Dateiablage. Zudem ist das Tool imstande eine grosse Anzahl an 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. Damit das Excel-Dokument verstehbar wird, sind gewisse Einstellungen bei DROID und im 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 im Excel. Die Angaben umfassen unter anderem die Dateiendung, die Grösse, das Datum der letzten Bearbeitung, das Format und die oben erwähnte PUID. Exportiert man die Daten in ein Excel-Dokument werden zusätzliche Angaben sichtbar, etwa der URI und der Pfad.

⁵¹ The National Archives: DROID user guide, vgl. <https://www.nationalarchives.gov.uk/documents/information-management/droid-user-guide.pdf> (Zugriff 13.7.2021).

⁵² The National Archives: The technical registry PRONOM, vgl. <https://www.nationalarchives.gov.uk/aboutapps/PRONOM/tools.htm> (Zugriff 13.7.2021).

⁵³ Lindlar, Micky: Dateiformatidentifizierung. Was, warum, wie, womit, ..., Referat am nestor-Praktikertag 2021 «Formate – Erkennen, Validieren und Behandeln», 9.6.2021, Folien 14, 17.

4.2.2. Analyse des Nachlasses Herzka mit DROID

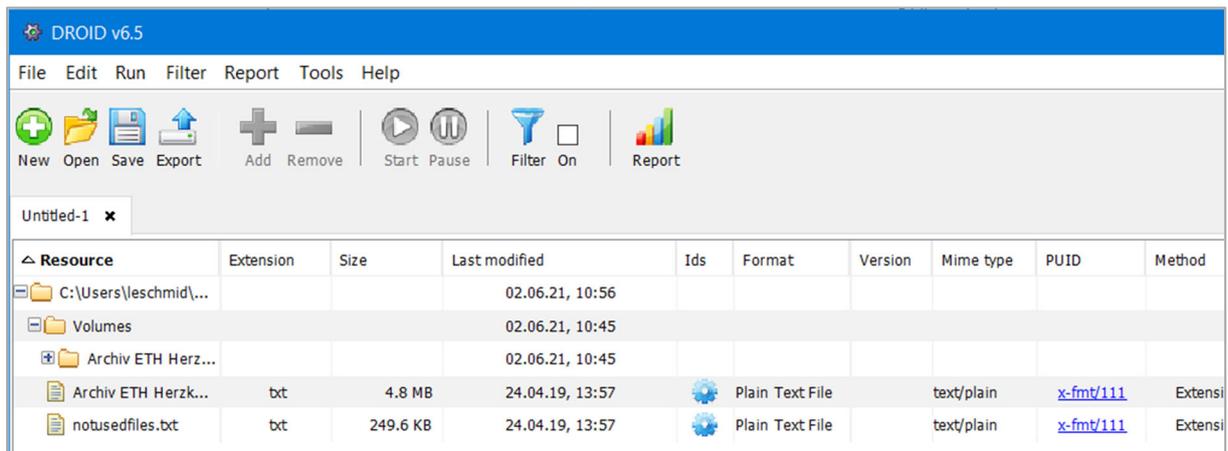


Abb. 7: Benutzeroberfläche mit Übersicht über Bestandsstruktur

Hat man den zu analysierenden Bestand ins Tool importiert, erscheint die in Abbildung 7 dargestellte Ansicht. Sie ist geprägt vom Symbol für Ordner und demjenigen für Dateien. Die Dateien sind mit Metadaten versehen. Es sind dies die Dateiendung, die Grösse, das Datum der letzten Bearbeitung, die ID, das Format, die Formatversion, der sogenannte Mimetype (eine grobe Einteilung in Datei- und Medientypen), die PUID, die Identifizierungsmethode und – auf der obigen Abbildung nicht sichtbar – der Hashwert der Datei. Zu den Ordnern sind keine Angaben verzeichnet.

Wünscht man Aufschluss über die Grösse und die Anzahl Dateien des Bestands, kommt die Filter-Funktion zum Einsatz. Abbildung 8 zeigt die Filter-Auswahl nach Dateien, Abbildung 9 den aus dieser Suche resultierenden Report unter Auswahl der Option «File count and sizes».

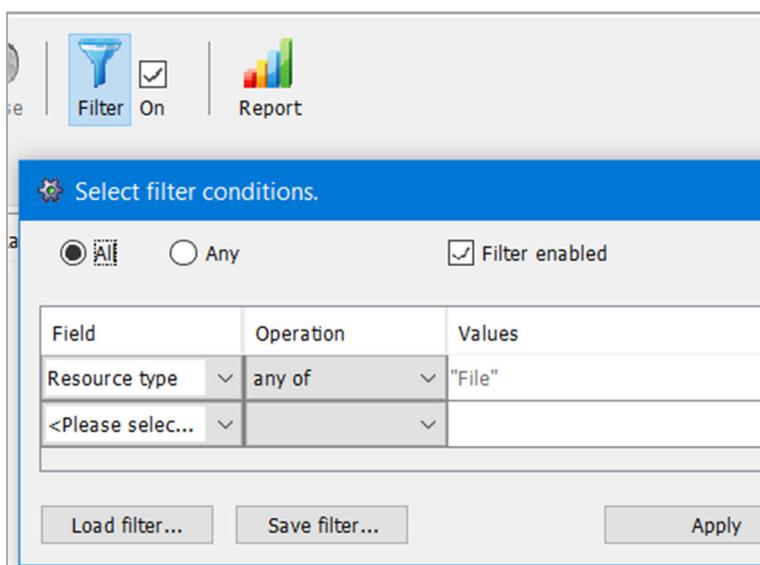


Abb. 8: Filter-Funktion zur Anzeige der Summe aller Dateien

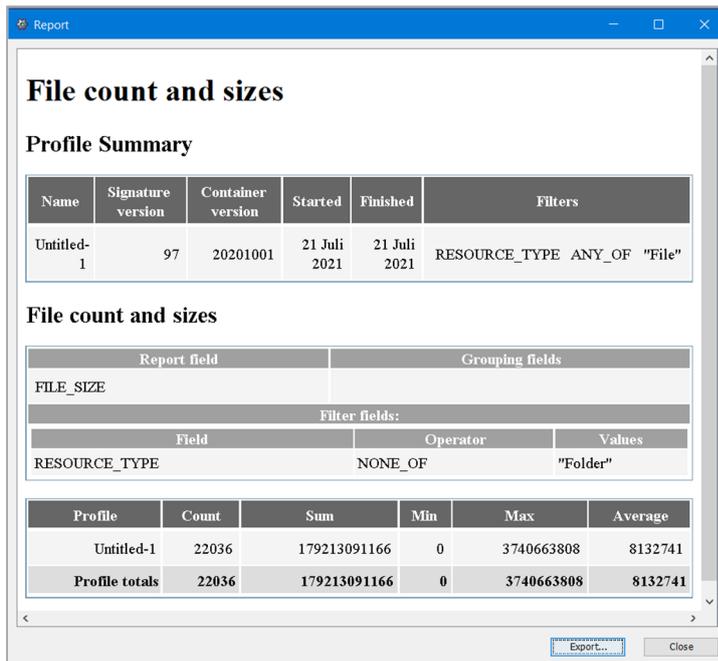


Abb. 9: «File count and sizes»-Report unter Auswahl des «File»-Filters

Die unterste Tabelle des Reports liefert Informationen zur Anzahl Dateien innerhalb der untersuchten Dateiablage (22'036) und der Gesamtgrösse dieser (179213091166 Bytes bzw. rund 179 GB). Es ist nicht möglich, mit demselben Vorgehen die Anzahl Ordner zu identifizieren. Filtert man den Bestand nach Ordnern und generiert einen Report, werden keine Treffer angezeigt.

Um doppelte Dateien zu identifizieren, muss die digitale Dateiablage in ein Excel-Dokument exportiert werden. Bei «bedingte Formatierung» im Start-Reiter findet sich die Wahlmöglichkeit «doppelte Werte» als Regel zum Hervorheben von Zeilen (Abbildung 10). Die aufgrund eines Vergleichs der Hashwerte identifizierten Dubletten können eingefärbt werden.

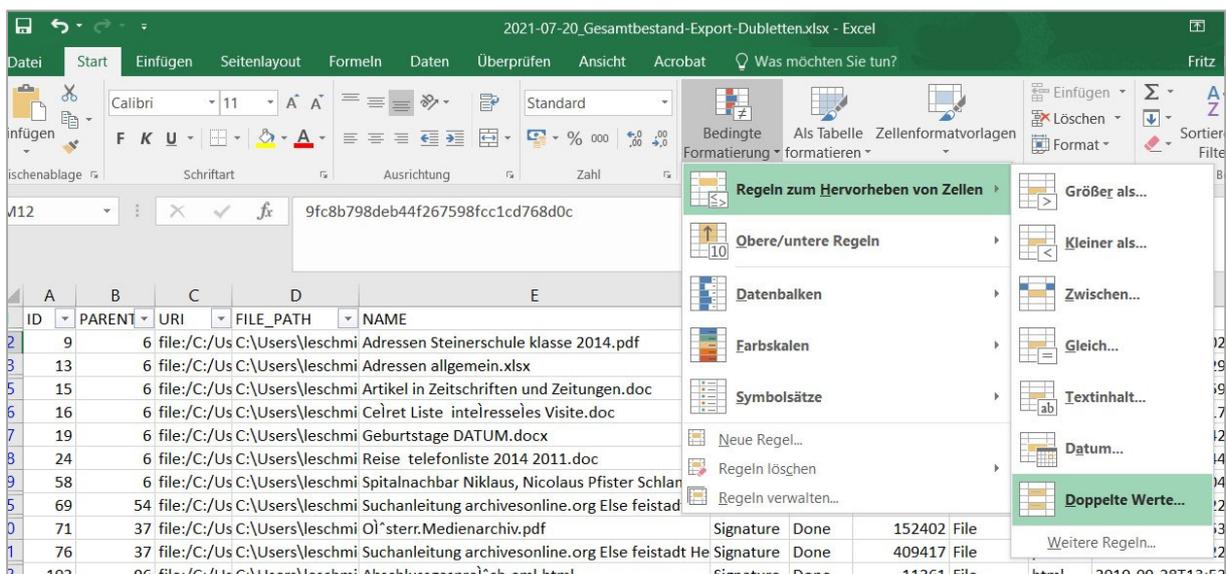


Abb. 10: Dublettensuche innerhalb des exportierten Excel-Dokuments

Filtert man die Spalte mit den MD5-Hashwerten nun nach eingefärbten Feldern, erhält man eine Liste all dieser Dateien (Abbildung 11). Insgesamt werden 7'785 Dateien als Dubletten identifiziert. Dabei ist nicht auf den ersten Blick erkennbar, welche Dateien dabei Doppel bilden. Dies müsste man anhand des Hashwerte prüfen. Um die Verteilung der Dubletten nach Dateitypen zu überblicken, filtert man am besten nach Formaten innerhalb der Kategorie Mime-type.

1	PARENT	URI	FILE_PAT	NAME	METHO	STATU	SIZE	TYPE	EXT	LAST_M	EXTENSIO	MD5_HASH
12	6	file:/C:/Us C:\Users\les Adressen Steinerschule klasse 2014.pdf			Signature	Done	228922	File	pdf	2014-06-2	false	9fc8b798deb44f267598fcc1cd768d0c
13	6	file:/C:/Us C:\Users\les Adressen allgemein.xlsx			Container	Done	199277	File	xlsx	2015-04-2	false	6018158b0557e9e053fd9e4732308b1d
15	6	file:/C:/Us C:\Users\les Artikel in Zeitschriften und Zeitungen.d			Container	Done	77824	File	doc	2010-09-0	false	657b9e270995bdfb34340dd7d7b98ec1
16	6	file:/C:/Us C:\Users\les Celret Liste interesseles Visite.doc			Container	Done	146432	File	doc	2012-08-2	false	4cbe62e462d3855cc422c68989717436
17	6	file:/C:/Us C:\Users\les Geburtstage DATUM.docx			Container	Done	137161	File	docx	2010-04-1	false	44f575dbf3867be6f7f6278d58c6c3f
18	6	file:/C:/Us C:\Users\les Reise telefonliste 2014 2011.doc			Container	Done	55808	File	doc	2014-10-1	false	089ddbba5309bdf4e812e6ca318a02d
19	6	file:/C:/Us C:\Users\les Spitalnachbar Niklaus, Nicolaus Pfister			Signature	Done	3E+06	File	pdf	2015-06-1	false	78a0abb9bec98051bfc91600b52ba7f3
25	54	file:/C:/Us C:\Users\les Suchanleitung archivesonline.org Else f			Signature	Done	409417	File	pdf	2010-11-0	false	fa617fad0900f17a8f2860641615fb3a
50	37	file:/C:/Us C:\Users\les Ol'sterr.Medienarchiv.pdf			Signature	Done	152402	File	pdf	2011-02-0	false	d1fb887c13386c22af8437b9c5a7e94
51	37	file:/C:/Us C:\Users\les Suchanleitung archivesonline.org Else f			Signature	Done	409417	File	pdf	2010-11-0	false	fa617fad0900f17a8f2860641615fb3a
83	96	file:/C:/Us C:\Users\les Abschlussgespral ch.eml.html			Signature	Done	11261	File	html	2010-09-2	false	7d4b5e2f2249bec24657d631168d72e
105	31	file:/C:/Us C:\Users\les Herzka_EGNER_Lit_liste.doc			Container	Done	45056	File	doc	2007-10-3	false	def6bbe0263ef415660980f3b3fc1753
117	5	file:/C:/Us C:\Users\les Artikel Herzka Liste.pdf			Signature	Done	123147	File	pdf	2010-09-0	false	a9b1db1dbd14b5c7ce04420f336d7d38
244	2961	file:/C:/Us C:\Users\les Autobiografie.jpg			Signature	Done	51389	File	jpg	2007-05-1	false	0b3b5effbe0fbb2b3f813ea9d8753a27
256	2961	file:/C:/Us C:\Users\les Buchverniissage & Musikfest.pdf			Signature	Done	361651	File	pdf	2007-06-2	false	c35fdbaae60780bf580f59a8eb1bc438
264	2983	file:/C:/Us C:\Users\les Autobiografie1.jpg			Signature	Done	51389	File	jpg	2007-05-1	false	0b3b5effbe0fbb2b3f813ea9d8753a27
267	2983	file:/C:/Us C:\Users\les Buchverniissage & Musikfest.pdf			Signature	Done	361651	File	pdf	2007-05-3	false	c35fdbaae60780bf580f59a8eb1bc438
273	2983	file:/C:/Us C:\Users\les Einlad 24.09.07.pdf			Signature	Done	321526	File	pdf	2007-08-1	false	fb5bc662fe7b3355d1521f0c6f42925c

Abb. 11: Ausschnitt aus der Excel-Datei zeigt identifizierte doppelte Dateien

4.3. TreeSize Professional

TreeSize ist eine lizenzpflichtige Software, die zur Analyse von digitalen Dateiablagen, dem Identifizieren und Löschen von Dubletten und unerwünschten Dateien angewandt werden kann. Das Tool wurde nicht eigens für den Archivbereich entwickelt, sondern ist ein allgemeiner, von der Firma Jam Software gegen Ende der 1990er-Jahre entwickelter «Datei- und Speicherplatzmanager».⁵⁴ Wird es zur Bewertung von Archivbeständen angewendet, gilt es Folgendes zu beachten: Das Tool greift einerseits – da es zur Optimierung von Speicherplatz und dem Löschen von nicht mehr benötigten Dateien entwickelt wurde – direkt in die Verzeichnisstruktur ein. Es sollte also immer mit einer Kopie der digitalen Ablage gearbeitet werden, nicht mit den Originaldaten. Andererseits protokolliert das Tool erfolgte Eingriffe nicht. Es ist also erforderlich, alle Bearbeitungsschritte manuell zu dokumentieren.

4.3.1. Funktionalitäten von TreeSize Professional

TreeSize 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 «Details»-Ansicht zeigt detaillierte Informationen zu allen Dateien und Ordnern des ausgewählten Elements, analog zu den Eckdaten des Pfades. Die Ansicht

⁵⁴ Jam Software: TreeSize, vgl. <https://www.jam-software.de/treesize> (Zugriff 13.7.2021).

nach «Dateitypen» bietet einen Überblick über die vorhandenen Dateitypen. Zu jedem Dateityp 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.

Die TreeSize-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.

4.3.2. Analyse des Nachlasses von Herzka mit TreeSize Professional

Name	Grösse	Belegt	Prozent (Grö...)	Dateien	Typ
Grafik-Dateien	106.1 GB	106.1 GB	63.6 %	10'268	Dateien, die Bilder, Grafik...
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 Konfiguratio...
Hilfdateien	1.8 MB	1.8 MB	0.0 %	6	Dateien für das Windows...
Datendateien	1.3 MB	1.4 MB	0.0 %	10	Dateien die verschiedenst...
Datenbankdateien	314.0 KB	316.0 KB	0.0 %	2	Dateien, die die Daten vo...
Dateien für die Softwareen...	4.0 KB	4.0 KB	0.0 %	2	Quell- und Projektdateien...

Abb. 12: Benutzeroberfläche mit «Dateitypen»-Ansicht

Die in Abbildung 12 sichtbare Benutzeroberfläche ermöglicht einen ersten Überblick über die digitale Dateiablage des Nachlasses. Die Kopfzeile gibt Aufschluss über die Grösse des Bestands (166.9 GB) und zeigt die Anzahl vorhandener Dateien (22'036) sowie die Anzahl Verzeichnisse (2'161). Der Verzeichnisbaum auf der linken Seite listet die Struktur des Bestands nach Ordnergrössen auf. Der Hauptbereich bietet die Ansicht der vorhandenen Dateitypen unter Angabe von Grösse und dem Bedarf an Speicherplatz. Die einzelnen Zeilen lassen sich per Mausklick erweitern, wodurch man eine Übersicht über die vorhandenen Dateiformate erhält (Abbildung 13).

Name	Größe	Belegt	Prozent (Größe)	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 Konfiguratio...

Abb. 13: «Dateitypen»-Ansicht unter Angabe der vorhandenen Dateiformate

Die Spalte «Dateien» macht ersichtlich, dass es sich bei etwa 10'270 der insgesamt rund 22'040 Dateien um Grafik-Dateien handelt. Abbildung 13 zeigt, dass über die Hälfte der Grafiken im JPG-Format gespeichert sind. Da die TIF-Dateien jedoch eine höhere Auflösung und damit einen grösseren Speicherplatzbedarf aufweisen, machen diese prozentmässig den grösseren Anteil aus.

Die TreeSize-Dateisuche umfasst die vorgefertigten Suchmuster unter anderem nach den grössten, ältesten, temporären, doppelten Dateien oder solchen mit besonders langem Pfad (vgl. Abbildung 14). Zudem besteht die Möglichkeit weitere Suchfilter zu definieren. Abbildung 15 zeigt die Resultatansicht der Suche nach doppelten Dateien innerhalb der digitalen Dateiablage des Nachlasses Herzka nach der Vergleichsmethode der MD5-Prüfsumme. Diese ergibt, dass die Ablage 6'647 Dubletten enthält.



Abb. 14: Auswahl von Funktionen der TreeSize-«Dateisuche»

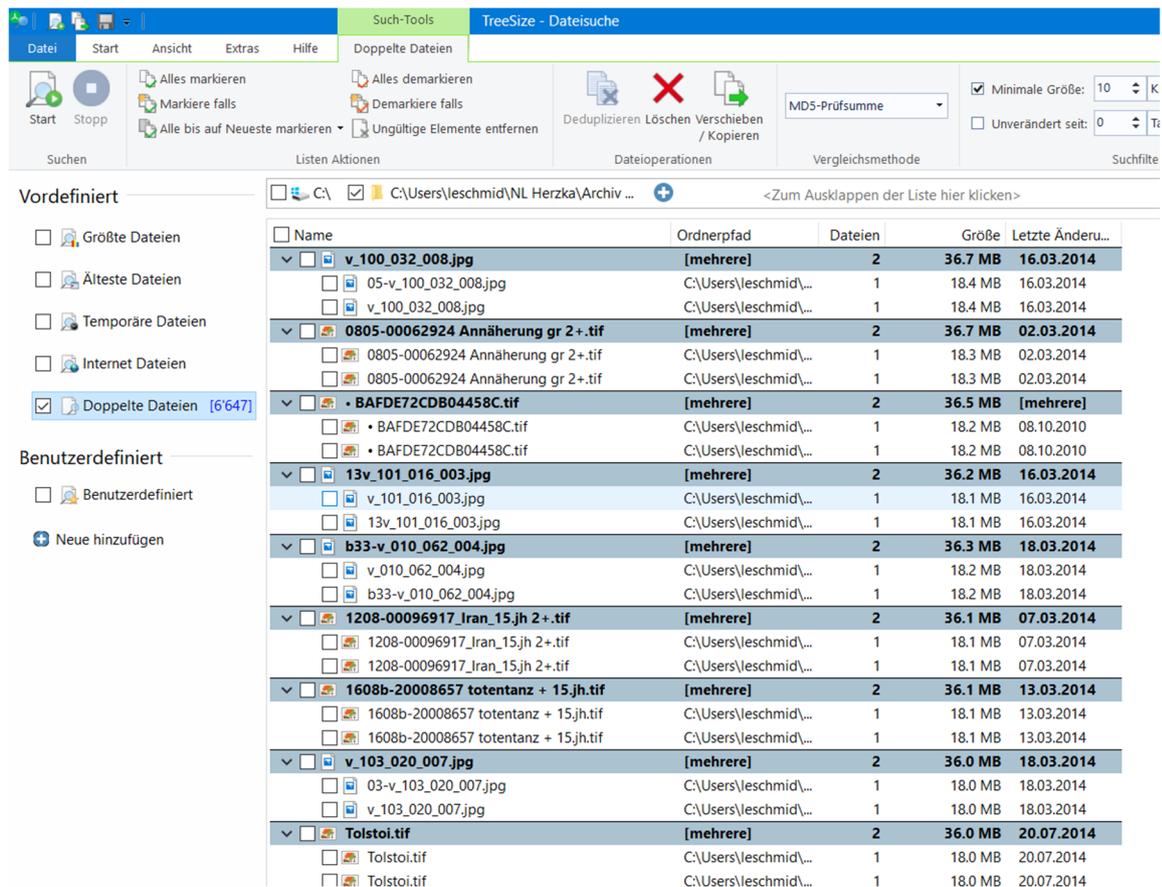


Abb. 15: Mithilfe der Funktion «Dateisuche» eruierte Doppel

5. Funktionalitäten und Gebrauchstauglichkeit der Tools

In diesem Kapitel werden die drei beschriebenen Tools bezüglich ihrer Funktionalitäten zur Umsetzung der technischen Analyse verglichen. Nachfolgend wird die Methode der System Usability Scale (SUS) zur Bestimmung der Gebrauchstauglichkeit von digitalen Tools beschrieben. Mithilfe dieser Methode werden die drei Tools ausgewertet.

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 Probleme und Herausforderungen.

5.1.1. Umfang und Mengengerüst

Alle drei Tools sind imstande 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. Vergleicht man die Ordnergrössen anhand der Tools Archifiltre und TreeSize wird ersichtlich, dass bei Archifiltre bei jedem einzelnen Ordner und bei jeder einzelnen Datei ein grösserer Umfang angegeben wird als beim entsprechenden Element bei TreeSize. Wie diese Differenz zu erklären ist, kann die Verfasserin nicht schlüssig erklären. Es gilt jedoch festzuhalten, dass der Unterschied im Gesamtumfang aufgrund der Summierung der durchgängig als etwas grösser berechneten Dateien bei Archifiltre (beziehungsweise der als etwas kleiner berechneten Dateien bei TreeSize) zustande kommt.

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 Ordner oder Datei, anwählen und sich dessen Umfang als Prozentangabe des Gesamtvolumens anzeigen lassen. Zudem ist sichtbar, wie viele Dateien desselben Formats existieren.⁵⁵ 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 der 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.

⁵⁵ 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.

5.1.2. Dateiformate und Pfadlängen

Alle drei Tools sind imstande 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 aufzuzeigen, zeigt diesen aber nicht als Wert an.

Sowohl spezielle Dateiformate als auch umfangreiche Pfadlängen bilden technische Herausforderungen beziehungsweise können zu technischen Problemen führen.

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. Wie ist diese grosse Differenz zu erklären? Auf diese Frage 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, ob überhaupt doppelte Dateien gelöscht werden sollen, dringlich. 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 Bestandserschliessung 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 auszulesen, die allesamt zu einer genaueren Bestimmung der Dateiformate beitragen. Es gilt sich die Frage zu stellen, welche Metadaten für die Erschliessung einer digitalen Dateiablage erforderlich sind.

5.2. Gebrauchstauglichkeit der Tools

Der Begriff der Usability wird als «Gebrauchstauglichkeit» ins Deutsche übersetzt und ist im ISO-Standard 9241-11 definiert. Es beschreibt «das Ausmaß, in dem ein System, ein Produkt oder eine Dienstleistung durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen».⁵⁶ Es gilt zu beachten, dass die Bewertung der Gebrauchstauglichkeit von der Nutzerin oder dem Nutzer, dem Nutzungskontext sowie von den definierten Zielen abhängig ist. Der Grund dafür liegt in der Tatsache, dass die Gebrauchstauglichkeit kein objektiv messbarer Wert darstellt.

5.2.1. System Usability Scale (SUS)

Die SUS ist ein einfacher, zehn Fragen umfassender Fragebogen. Er wurde 1986 von John Brooke entwickelt und gilt als «etablierte Methode zur quantitativen Analyse der Gebrauchstauglichkeit».⁵⁷ Die zehn Fragen – welche je nach Anwendungsgebiet angepasst werden können – sind die folgenden:

1. Ich kann mir sehr gut vorstellen das System regelmässig zu nutzen.
2. Ich empfinde das System als unnötig komplex.
3. Ich empfinde das System als einfach zu nutzen.
4. Ich denke, dass ich technischen Support brauchen würde, um das System zu nutzen.
5. Ich finde, dass die verschiedenen Funktionen des Systems gut integriert sind.
6. Ich finde, dass es im System zu viele Inkonsistenzen gibt.
7. Ich kann mir vorstellen, dass die meisten Leute das System schnell zu beherrschen lernen.
8. Ich empfinde die Bedienung als sehr umständlich.

⁵⁶ Wikipedia: Gebrauchstauglichkeit, vgl. [https://de.wikipedia.org/wiki/Gebrauchstauglichkeit_\(Produkt\)](https://de.wikipedia.org/wiki/Gebrauchstauglichkeit_(Produkt)) (Zugriff 21.7.2021).

⁵⁷ Wikipedia: System Usability Scale, vgl. https://de.wikipedia.org/wiki/System_Usability_Scale (Zugriff 21.7.2021).

9. Ich habe mich bei der Nutzung des Systems sehr sicher gefühlt.
10. Ich musste eine Menge Dinge lernen, bevor ich mit dem System arbeiten konnte.

Diese Aussagen werden auf einer Skala von 1 bis 5 bewertet, wobei die 1 vollständige Ablehnung und die 5 vollständige Zustimmung bedeutet. Die Addierung der Punkte von 0 bis 4 (4 gibt es für die bestmögliche Antwort, wobei darauf zu achten ist, ob die Aussage positiv oder negativ formuliert ist) ergibt, multipliziert mit 2.5, eine Bewertung des Tools mit einer Punktzahl zwischen 0 und 100.⁵⁸ Enthält ein Produkt eine Wertung ab 68 Punkten kann man davon ausgehen, dass die Usability gut ist. Diese steigert sich bis hin zu einer perfekten Usability bei 100 Punkten.

5.2.2. Auswertung der Tools nach der SUS

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 Anwenderinnen und Anwender verständlich und einfach zu nutzen sind.

Das Tool Archifiltre ist intuitiv erfass- und verstehbar. 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. Die Auswertung des Fragebogens ergibt für Archifiltre eine Punktzahl von 37. Damit erreicht das Tool eine Gesamtbewertung von 92.5.

DROID ist für Anwenderinnen und Anwender 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 Auswertung des SUS-Fragebogens ergab für dieses Tool eine Punktzahl von 15, was einer Gesamtbewertung von 37.5 entspricht.

Die Darstellung des Tools TreeSize ist sehr übersichtlich. Der Einsatz verschiedener Diagramme, die Farbigkeit und die Detailangaben zu einzelnen Elementen sind visuell ansprechend. Die zahlreichen Funktionalitäten sind zumeist selbsterklärend oder können ohne Hilfe einer Fachperson erlernt werden. Die Auswertung des Fragebogens attestiert dem Tool 35 Punkte. Dies entspricht einer Bewertung von 87.5.

⁵⁸ Brooke, John: SUS - A quick and dirty usability scale, s. l. 1995, vgl. https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale (Zugriff 21.7.2021), S. 4f.

5.2.3. Vergleich der Tools anhand des Kriteriums der Gebrauchstauglichkeit

Das Tool Archifiltre erhielt bei meiner Auswertung mit einer Bewertung von 92.5 von 100 eine Punktzahl, die auf eine fast perfekte Usability hindeutet. TreeSize kommt mit 87.5 von 100 Punkten ebenfalls auf eine Wertung, die auf eine sehr gute Usability hinweist. DROID dagegen erhielt 37.5 von 100 Punkten, was dem Tool nach SUS eine ungenügende Bewertung einbringt. Welche Schlussfolgerungen können daraus gezogen werden? Es ist wichtig noch einmal zu betonen, dass die Bewertung der Gebrauchstauglichkeit nach SUS 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 Nutzerinnen und Nutzern 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 fürs Archiv für Zeitgeschichte

6.1. Schlussfolgerungen

Die vorliegende Arbeit 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 hochaktuel-ler – noch relativ wenig erforschter Bereich innerhalb der digitalen Langzeitarchivierung darstellen. Publiziert wurden einzelne Werkstattberichte und Beschreibungen von Pilotprojekten. In diesen werden die Bearbeitungsschritte der Übernahme dargelegt und die verwendeten Tools beschrieben. Die KOST initiierte im vergangenen Jahr 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 der vorliegenden Zertifikatsarbeit 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 Anwenderinnen und Anwender 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 ergeben 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 Anwenderinnen und Anwendern ohne Vorkenntnisse im IT-Bereich schnell erlernt und angewendet werden können. DROID setzt (viel) 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. Ein Tool zur Verfügung zu haben das einen schnellen Überblick über digitale Dateiablagen ermöglicht, kann die Beurteilung, ob ein Bestand übernommen werden sollte und wenn ja ob integral oder teilweise, erheblich erleichtern. 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 fürs Archiv für Zeitgeschichte

Die im Rahmen der vorliegenden Zertifikatsarbeit 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. Im Rahmen dieser 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 Anwenderinnen und Anwender ohne Vorkenntnisse im IT-Bereich herausgestellt. Dies erscheint der Verfasserin relevant, da die Mehrheit der Archivarinnen und Archivare Geschichtswissenschaften studiert haben und über keine IT-Ausbildung verfügen.

Die durchgeführten Untersuchungen zeigen, dass sowohl Archifiltre als auch TreeSize bei der Übernahme digitaler Dateiablagen gewinnbringend eingesetzt werden könnten. Während mit Archifiltre ein schneller 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 im AfZ als Arbeitsinstrumente einzusetzen.

7. Bibliografie

7.1. Literatur

- Béchar, Lorène; Fuentes Hashimoto, Lourdes; Vasseur, Édouard: Les archives électroniques. Deuxième édition, enrichie et mise à jour, Paris 2020.
- 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.
- 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.
- Corrado, Edward M.: Digital Preservation for Libraries, Archives and Museums. Second edition, Lanham 2017.
- Enge, Jürgen; Kramski, Heinz Werner: «Arme Nachlassverwalter...» Herausforderungen, Erkenntnisse und Lösungsansätze bei der Aufbereitung komplexer digitaler Dateisammlungen, in: Jörg Filthaut (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.
- Heizmann, Uwe: Strategie zur Archivierung digitaler Objekte im Kreisarchiv Reutlingen, Reutlingen 2017.
- Heizmann, Uwe: Strategie zur Archivierung digitaler Objekte im Kreisarchiv Reutlingen, in: Archivar 70 (2017), Heft 3, S. 300–307.
- 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.
- Kramski, Heinz Werner: Digitale Dokumente im Archiv, in: Lepper, Marcel; Raulff, Ulrich (Hg.), Handbuch Archiv. Geschichte, Aufgaben, Perspektiven, Stuttgart 2016, S. 178–197.

- Kramski; Heinz Werner; von Bülow, Ulrich: «Es füllt sich der Speicher mit köstlicher Habe» – Erfahrungen mit digitalen Archivmaterialien im Deutschen Literaturarchiv Marbach, in: Robertson-von Trotha, Caroline Y.; Hauser, Robert (Hg.), Neues Erbe. Aspekte, Perspektiven und Konsequenzen der digitalen Überlieferung, Karlsruhe 2001, S. 141–162.
- 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.
- 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.
- 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.
- 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.
- Neuroth, Heike: Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.3, Boizenburg 2010.
- Owens, Trevor: The Theory and Craft of Digital Preservation, Baltimore 2018.
- Pfyffer, Gaby: Personennachlässe im digitalen Zeitalter sichern. Was braucht es von Seite der Archive? Analyse und Konzept für das Archiv für Zeitgeschichte (AfZ). Masterarbeit MAS ALIS 2016–2018, Winterthur 2018.
- Purcell, Aaron D.: The Digital Archives Handbook. A Guide to Creation, Management and Preservation, Lanham 2019.
- 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. 20–38.
- 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.

Texier, Bruno: Traiter un vrac numérique – mode d’emploi, in: archimag, Nr. 342/mars 2021, Varennes-Vauzelles 2021, S. 17–18.

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.

7.2. Internetquellen

Archiv für Zeitgeschichte der ETH Zürich: Die systematische Ablage von privaten digitalen Daten, Zürich 2018, vgl. https://www.afz.ethz.ch/images/uploads/dokumente/2018-07-26_SVNE_RM_Privatpersonen.pdf (Zugriff 13.7.2021).

Brooke, John: SUS – A quick and dirty usability scale, s. l. 1995, vgl. https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale (Zugriff 21.7.2021).

Büchler, Georg: Plädoyer für archivische Demut. Referat an der 22. Jahrestagung des Arbeitskreises «Archivierung von Unterlagen aus digitalen Systemen» in Marburg, 2018, vgl. https://kost-ceco.ch/cms/dl/5997f478163a96aba8bc4e42b13e1b/AUdS2018_Beitrags_B%C3%BChler.pdf (Zugriff 29.6.2021).

GitHub: Wiki Archifiltre, vgl. <https://github.com/SocialGouv/archifiltre/wiki/Wiki-Archifiltre> (Zugriff 13.7.2021).

Jam Software: TreeSize, vgl. <https://www.jam-software.de/treesize> (Zugriff 13.7.2021).

Lenartz, Stephan: Aufbereitung und Bewertung von Fileablagen mit Python am Beispiel einer digitalen Fotosammlung, St. Gallen 2018, vgl. https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/auds-2018/praxis/24-Lenartz_Artikel_AUdS_2018.pdf (Zugriff 12.5.2021).

Lindlar, Micky: Dateiformatidentifizierung. Was, warum, wie, womit, ..., Referat am nestor-Praktikertag 2021 «Formate – Erkennen, Validieren und Behandeln», 9.6.2021, vgl. <https://www.langzeitarchivierung.de/Webs/nestor/SharedDocs/Downloads/DE/praesentationen/2021praktikertagLindlar.html?nn=182270> (Zugriff 2.7.2021).

Nestor: Toolbox, vgl. <https://wiki.dnb.de/pages/viewpage.action?pageId=134715087> (Zugriff 24.7.2021).

Röthlisberger-Jourdan, Claire: Formatauswahl. Denn sie wissen was sie tun, Referat am nestor-Praktikertag 2021 «Formate – Erkennen, Validieren und Behandeln», 9.6.2021, vgl. <https://www.langzeitarchivierung.de/Webs/nestor/SharedDocs/Downloads/DE/praesentationen/2021praktikertagRoethlisberger.html?nn=182270> (Zugriff 2.7.2021).

Software Archifiltre: «Qui sommes-nous?», vgl. <https://archifiltre.fabrique.social.gouv.fr/qui-sommes-nous> (Zugriff 13.7.2021).

The National Archives: Download DROID, vgl. <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/> (Zugriff 13.7.2021).

The National Archives: The technical registry PRONOM, vgl. <https://www.nationalarchives.gov.uk/aboutapps/PRONOM/tools.htm> (Zugriff 13.7.2021).

Wikipedia: Gebrauchstauglichkeit, vgl. [https://de.wikipedia.org/wiki/Gebrauchstauglichkeit_\(Produkt\)](https://de.wikipedia.org/wiki/Gebrauchstauglichkeit_(Produkt)) (Zugriff 21.7.2021).

Wikipedia: System Usability Scale, vgl. https://de.wikipedia.org/wiki/System_Usability_Scale (Zugriff 21.7.2021).

8. Anhang

8.1. Abbildungsverzeichnis

Screenshots aus dem Tool Archifiltre

Abb. 1: Farbskala nach Dateitypen

Abb. 2: Benutzeroberfläche mit Übersicht über die digitale Dateiablage

Abb. 3: Auswahl eines Unterelements innerhalb der Dateiablage

Abb. 4: «Audit» des Bestands

Abb. 5: Übersichtsdarstellung über Dateiduplikate

Abb. 6: Detailansicht der Dateiduplikate

Screenshots aus dem Tool DROID

Abb. 7: Benutzeroberfläche mit Übersicht über Bestandsstruktur

Abb. 8: Filter-Funktion zur Anzeige der Summe aller Dateien

Abb. 9: «File count and sizes»-Report unter Auswahl des «File»-Filters

Abb. 10: Dublettensuche innerhalb des exportierten Excel-Dokuments

Abb. 11: Ausschnitt aus der Excel-Datei zeigt identifizierte doppelte Dateien

Screenshots aus dem Tool TreeSize Professional

Abb. 12: Benutzeroberfläche mit «Dateitypen»-Ansicht

Abb. 13: «Dateitypen»-Ansicht unter Angabe der vorhandenen Dateiformate

Abb. 14: Auswahl von Funktionen der TreeSize-«Dateisuche»

Abb. 15: Mithilfe der Funktion «Dateisuche» eruierte Doppel

8.2. SUS-Auswertungen

System Usability Scale:

Das Tool **Archifltre**

