

E-M@il für Dich

Langzeitarchivierung von E-Mails in Rosetta
am Beispiel des Projekts „E-Mail Archiv 2.0“
des Hochschularchivs der ETH Zürich

Masterarbeit im Rahmen des
MAS ALIS 2020-2022

bei Georg Bächler
am Historischen Institut
der Philosophisch-historischen Fakultät
der Universität Bern
am 25.07.2022

Claudia Briellmann
Dittingerstrasse 39
4053 Basel
Tel. +41 (0)79 576 15 01
claudia.briellmann@library.ethz.ch
Matrikelnummer: 10-054-419

Inhaltsverzeichnis

Abstract	2
1. Einleitung.....	2
1.1. Methodik.....	4
1.2. Überblick über den Forschungsstand	4
2. Das Projekt „E-Mail-Archiv 2.0“	7
2.1. Die rechtlichen Aspekte der E-Mail-Archivierung	8
3. Die Bewertung von Big Data.....	12
3.1. Bewertungsmodelle für E-Mail-Bestände	14
4. Die technischen Grundlagen der E-Mail-Archivierung	20
4.1. Auszulesende Metadaten.....	23
4.2. Archivwürdige Formate für E-Mails	25
5. Grundlagen der Erschliessung von E-Mails.....	31
6. Die praktische Umsetzung im Hochschularchiv.....	33
6.1. Die Bewertungsentscheide im Hochschularchiv	33
6.2. Metadaten-Extraktion im Hochschularchiv	34
6.3. Die Entscheidung über das Format	36
6.4. Die Frage von Anhängen und Linked Content.....	38
6.5. Die Erschliessung in CMI AIS.....	40
6.6. Der Ingest in Rosetta.....	42
7. Fazit und weiteres Vorgehen	44
8. Bibliographie.....	47
Anhang 1	54

Abstract

E-Mails sind eine Form der Korrespondenz, die aus der heutigen privaten und geschäftlichen Welt nicht mehr wegzudenken sind. Doch während die Bedeutung von E-Mails mittlerweile in den meisten Institutionen klar ist, so ist die digitale Langzeitarchivierung von E-Mails bisher in den Wenigsten fortgeschritten. Im Rahmen dieser Masterarbeit soll das Thema der E-Mail-Archivierung auf einer theoretischen und einer praxisbezogenen Ebene untersucht werden. Im Fokus stehen hierbei die Bewertung und Erschliessung von E-Mails, die sich auf Grund der grossen Datenmenge von der klassischen Bewertung, wie sie aus dem Bereich des analogen Schriftguts bekannt ist, unterscheidet. Im Bewertungsdiskurs wird neben dem klassischen Records Management der Capstone Approach behandelt, der für die Bewertung von E-Mails entwickelt wurde. Bei der Erschliessung wird der MPLP-Ansatz von Meissner und Greene als Möglichkeit der Erschliessung von Big Data analysiert. Als praktisches Beispiel wird das Projekt „E-Mail-Archiv 2.0“ des Hochschularchivs der ETH Zürich behandelt.

1. Einleitung

E-Mails sind eine Form der Korrespondenz, die aus der heutigen privaten und geschäftlichen Welt nicht mehr wegzudenken sind. Von der Artefactual Systems and Digital Preservation Coalition wird eine E-Mail folgendermassen definiert:

Email is a system of asynchronous exchange of messages between computers. An email message is a structured text document which typically includes a header and body, may contain embedded images or other objects, and may include attachments. Email messages are typically structured using the Internet Message Format (IMF) (RFC 5322, 2008) and Multipurpose Internet Mail Extensions (MIME) (RFC 2405, 1996) standards.¹

Die Task Force on Technical Approaches to Email Archives wählt einen anderen Ansatz für ihre Definition wählt:

Email is both one thing and many things. It's an individual message; it's a collective noun for all the messages in a mailbox; it's an active verb ("I'll email you later today"). But behind each of these uses, email is the system that creates, distributes, and receives messages according to the rules of a defined, extensible set of standards.²

¹ Artefactual Systems and the Digital Preservation Coalition: Preserving Email, DPC Technology Watch Report, Digital Preservation Coalition, 2021. Online: <<https://doi.org/10.7207/twgn21-08>>.

² Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives: a report from the Task Force on Technical Approaches for Email Archives, Washington, DC 2018 (CLIR publication 175), S. 22.

Die erste E-Mail wurde nach heutigem Wissensstand 1971 von einem Mitarbeiter von ARPANET (Advanced Research Projects Agency Network) namens Ray Tomlinson verschickt.³ Damals hiess diese Übertragung von Nachrichten noch nicht E-Mail und der Begriff wurde wohl von V.A. Shiva Ayyadurai, der 1978 eine Dokumentation für eine Erfindung mit dem Titel „Email“ verfasste, die sich heute im Smithsonian befindet.⁴ Es ist nie abschliessend geklärt worden, wer nun als der wahre Erfinder der E-Mail, wie wir sie heute kennen, gelten soll. Die Entstehung der E-Mail lässt sich jedoch auf die 1970er festlegen und seitdem hat diese Erfindung grosse Fahrt aufgenommen. Seit den 1990er Jahren hat die E-Mail besonders im Geschäftsverkehr an Relevanz gewonnen. Zwar wird aktuell immer noch häufig auf eine hybride Geschäftsform geachtet, ein Grossteil der geschäftsrelevanten, elektronischen Dokumente bestehen mittlerweile jedoch aus E-Mails.⁵ Robert Kretschmar bezeichnet E-Mails in seinem Artikel *Alles neu zu durchdenken? Archivische Bewertung im digitalen Zeitalter* sogar als „fast eher schon zum klassischen Schriftgut“⁶ gehörend.

Doch während die Bedeutung von E-Mails mittlerweile in den meisten Institutionen klar ist, ist die digitale Langezeitarchivierung von E-Mails bisher in den wenigsten Institutionen fortgeschritten. Christopher Prom beschreibt in seinem *DPC Technology Watch Report 2019*, dass nur sehr wenige Repositorien bereits darauf ausgelegt sind, E-Mails zu archivieren und „*an even smaller number have developed policies, implementation strategies, procedures, tools and services that systematically do so.*“⁷ Trotzdem ist die E-Mail-Archivierung ein komplexes Thema, welches immer mehr an Wichtigkeit gewinnt.⁸

³ Brühwiler, Tom: Internet-Pionier Ray Tomlinson: Der Mann, der die E-Mail erfand, Cyon, 26.02.2021, <<https://www.cyon.ch/blog/Internet-Pionier-Ray-Tomlinson-Erfinder-E-Mail>>.

⁴ Kolawole, Emi: Smithsonian acquires documents from inventor of 'EMAIL' program, in: The Washington Post, 17.02.2012. Online: <https://www.washingtonpost.com/national/on-innovations/va-shivaayyadurai-inventor-of-e-mail-honored-by-smithsonian/2012/02/17/glQA8gQhKR_story.html>.

⁵ Sturm, Patrick: Die E-Mail: Ein Kommunikationsmedium des frühen 21. Jahrhunderts quellenkundlich betrachtet, in: Berwinkel, Holger; Kretschmar, Robert; Uhde, Karsten u. a. (Hg.): *Moderne Aktenkunde*, Marburg 2016 (Veröffentlichungen der Archivschule Marburg, Hochschule für Archivwissenschaft), S. 110; Schwarz, Karin: E-Mail-Archivierung, in: Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.3., 2010, S. 550; Benauer, Maria: E-Mails, ihr Wert und ihre Bewertung, in: VÖA - Verband Österreichischer Archivarinnen und Archivare (Hg.): *Scrinium* 74, 2020, S. 87; Prom, Christopher: *Preserving Email* (2nd Edition), DPC Technology Watch Report, Digital Preservation Coalition, 05.2019, S. 4. Online: <<https://doi.org/10.7207/twr19-01>>.

⁶ Kretschmar, Robert: *Alles neu zu durchdenken? Archivische Bewertung im digitalen Zeitalter*, in: *Archivpflege in Westfalen und Lippe* 80, 2014, S. 9.

⁷ Prom: *Preserving Email* (2nd Edition), 2019, S. 5.

⁸ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives*, 2018, S. 1.

1.1. Methodik

Im Rahmen dieser Masterarbeit soll das Thema der E-Mail-Archivierung auf einer theoretischen und einer praxisbezogenen Ebene untersucht werden. Im Fokus steht hierbei die Bewertung von E-Mails, die sich auf Grund der grossen Datenmenge von der klassischen Bewertung, wie sie aus dem Bereich des analogen Schriftguts bekannt ist, unterscheidet. Es werden nicht nur die Möglichkeiten der Bewertung von Big Data analysiert, sondern auch die in der Fachliteratur genannten spezifischen Methoden zur Bewertung von E-Mail-Konten aufgezeigt. Dabei wird ein Blick auf Institutionen geworfen, welche bereits erfolgreich E-Mails archivieren. Ein weiterer Punkt, der betrachtet wird, ist die Erschliessung ebenjener E-Mail-Konten. Auch in diesem Bereich spielt der Umfang der übernommenen Daten eine sehr grosse Rolle, wenn es darum geht, welche Möglichkeiten sich bieten, die nicht zu viele Ressourcen für eine sehr lange Zeit binden. Der Punkt der Vermittlung dieser archivierten E-Mails soll in der Arbeit nicht gesondert behandelt werden. Er wird bei den jeweiligen Kapiteln jedoch einfließen, sofern Entscheidungen auf Grund der möglichen zukünftigen Vermittlung getroffen werden (z.B. bei der Darstellung in Rosetta). Da durch Einsichtsgesuche auch frühere Einsichten möglich sind, fliessen grundlegende Überlegungen, wie die E-Mails im Bedarfsfall zugänglich gemacht werden können, bereits jetzt in die Entscheidungen der Langzeitarchivierung ein und werden an den entsprechenden Stellen erwähnt.

Schliesslich sollen diese theoretischen Ansätze und Vergleiche mit anderen Institutionen praktisch an der Umsetzung des Projekts „E-Mail-Archiv 2.0“ im Hochschularchiv der ETH Zürich aufgezeigt werden. Dabei ist zu bemerken, dass in diesem letzten Teil grosse Rücksicht auf die gegebenen Umstände genommen werden muss. Es werden die Ressourcen, Tools und Applikationen, welche der ETH zur Verfügung stehen in Betracht gezogen und auf Grund dieser Entscheidungen gefällt. Hätten allfällig neue Tools gekauft werden müssen, so wäre dies nur zu rechtfertigen gewesen, wenn das gleiche oder ein ähnlich gutes Ergebnis nicht auch mit den vorhandenen Tools zu erreichen gewesen.

Dabei ist zu beachten, dass es sich um eine archivische Arbeit handelt, weshalb zwar durchaus auf technische Fragestellungen eingegangen wird, diese jedoch nicht auf Detailebene beschrieben werden können und deshalb hier die Archivarbeit im Vordergrund steht. Grundlegende Kenntnisse über Grundlagen der Archivierung und vor allem der digitalen Langzeitarchivierung werden für diese Arbeit vorausgesetzt und entsprechende Fragen nicht in den jeweiligen Kapiteln erläutert.

1.2. Überblick über den Forschungsstand

Damit diese Masterarbeit im aktuellen Forschungsstand verortet werden kann, muss zuerst ein kurzer Blick auf die aktuelle und bisherige Forschung geworfen werden.

Mit Gewinn an Wichtigkeit der E-Mail in der Geschäftswelt in den 1990er Jahren, begannen sich bereits damals die ersten Forschenden damit zu beschäftigen, wie E-

Mails korrekt gemanaged werden sollten. David Bearman war mit seinem Artikel *Managing electronic mail*⁹ ein Vorreiter des elektronischen Records Managements, wie man es heute kennt, und Magret Hedstrom veröffentlichte 1993 den Artikel *Descriptive Practices for Electronic Records*¹⁰, in dem sie einen Blick auf die Erschliessung von grossen Datenmengen, wie sie zu erwarten waren, vorwegnahm. 1997 veröffentlichte Nicole Periat in der Zeitschrift *Archives* aus Québec den Beitrag *Politique de gestion du courrier électronique: des mesures à prendre*¹¹ in dem es spezifisch um das Management von E-Mails geht.

In den frühen 2000er Jahren kamen dann bereits die ersten Projekte zur E-Mail-Archivierung auf. Hier sind Projekte wie das eDAVID Projekt¹² in Belgien oder das InterPares 3 Projekt¹³, welches von 2007 bis 2009 in Italien durchgeführt wurde, zu nennen. Auch gab es am King's College in London das InSPECT Projekt¹⁴, das sich ebenfalls zwischen 2007 und 2009 mit der Ermittlung von signifikanten Eigenschaften¹⁵ von E-Mails auseinander setzte. 2002 wurde zudem die Digital Preservation Coalition gegründet, die sich aus Forschern aus dem Vereinigten Königreich und Irland zusammensetzte und sehr viel zu digitaler Archivierung und der E-Mail-Archivierung publizierte. Beispielhaft lassen sich hier die zwei Reports von Christopher Prom zum Thema E-Mail-Archivierung nennen, *Preserving Email*¹⁶ von 2011 und die überarbeitete Version von 2019¹⁷. 2006 publizierte zudem Maureen Pennock vom *Digital Curation Center* einen Bericht mit dem Titel *Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages*.¹⁸

2016 starteten die Andrew W. Mellon Foundation in New York und die Digital Preservation Coalition in Glasgow den Versuch, mit der neu gegründeten Task Force on Technical Approaches for Email Archives, die Ergebnisse dieser vielen unterschiedlichen Projekte zusammenzuführen. Das Ergebnis ist das aktuell wohl umfassendste Dokument, welches zu E-Mail-Archivierung existiert: *The future of email archives: a report from the Task Force on Technical Approaches for Email*

⁹ Bearman, David: Managing electronic mail, in: Archives & Manuscripts 22 (1), 1994. Online: <<https://publications.archivists.org.au/index.php/asa/article/view/8385>>.

¹⁰ Hedstrom, Margaret: Descriptive Practices for Electronic Records: Deciding What Is Essential and Imagining What Is Possible, in: Archivaria 36, 1993.

¹¹ Periat, Nicole: Politique de gestion du courrier électronique: des mesures à prendre, in: Archives 29 (1), 1997, S. 3–56.

¹² Boudrez, Filip; Van den Eynde, Sofie: DAVID: Archiving Email, Leuven - Antwerp 2002. Online: <<http://www.expertisecentrumdavid.be/davidproject/teksten/Rapporten/Report4.pdf>>.

¹³ The InterPARES 3 Project, Team Italy: General Study 05 – Keeping and Preserving E-mail, 2009. Online: <http://www.interpares.org/ip3/display_file.cfm?doc=ip3_italy_gs05a_final_report.pdf>.

¹⁴ King's College London: InSPECT: Investigating the Significant Properties of Electronic Content Over Time, Projektwebseite, <<https://significantproperties.kdl.kcl.ac.uk/>>.

¹⁵ Definition aus: The Library of Congress: PREMIS Data Dictionary for Preservation Metadata, Version 3.0, 2015. Online: <<https://www.loc.gov/standards/premis/>>: Characteristics of a particular object subjectively determined to be important to maintain through preservation actions.

¹⁶ Prom, Christopher J.: Preserving Email, DPC Technology Watch Report, Digital Preservation Coalition, 2011. Online: <<https://doi.org/10.7207/twr11-01>>.

¹⁷ Prom: Preserving Email (2nd Edition), 2019.

¹⁸ Pennock, Maureen: Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages, DCC Digital Curation Manual, 2006. Online: <<https://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf>>.

Archives.¹⁹ 2021 folgte dann von der DPC in Zusammenarbeit mit den Artefactual Systems ein Report zu *Preserving Email*.²⁰

Parallel zu den Versuchen, den grossen Themenkomplex der E-Mail-Archivierung zusammenzufassen und Guides zum besseren Management von E-Mails und ihrer Archivierung zu erstellen, gab es mehrere Publikationen zu spezifischen Themen betreffend der E-Mail-Archivierung.

Anne Burnel mit ihrem Artikel *Si le courriel est l'avenir de la lettre... Enjeux juridiques, technologiques et patrimoniaux d'archivage du courrier électronique*²¹, Daniel Noonans *Email: An Appraisal Approach* sowie James Lappin *The defensible deletion of government email* sind gute Beispiele für die Auseinandersetzung mit der Bewertung von E-Mails. Publikationen wie die von Vellino und Alberst über *Assisting the appraisal of e-mail records with automatic classification*²² oder Vinh-Doyles *Appraising email (using forensics): techniques and challenges*²³ brachten Mitte der 2010er Jahre die Wichtigkeit von technischer Hilfe bei der Bewertung von grossen E-Mail-Beständen in die Forschungsdiskussion ein.

Die Erschliessung von E-Mail-Beständen wurde im Forschungsdiskurs bisher eher weniger behandelt, auch wenn zum Beispiel im 2019 veröffentlichten Artikel *Appraising, processing, and providing access to email in contemporary literary archives*²⁴ sowohl auf die Bewertung, die Erschliessung wie auch auf die Vermittlung eingegangen wird. Aber auch in den grossen Publikationen zur E-Mail-Archivierung finden diese Punkte Eingang.

Im deutschsprachigen Raum begann der Diskurs zur E-Mail-Archivierung 2012 mit dem Projekt des Archivs der sozialen Demokratie (AdsD) der Friedrich-Ebert-Stiftung.²⁵ Maria Benauer schreibt über die Beiträge zur E-Mail-Archivierung, dass sie meist „auf kleinen Praxisstudien basieren“ und mehrheitlich „punktuelle Problemanalysen“ sind, die „konkrete technische Löschungsvorschläge für die digitale Langzeitarchivierung von E-Mails vorstellen“, aber im internationalen

¹⁹ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives*, 2018.

²⁰ Artefactual Systems and the Digital Preservation Coalition: *Preserving Email*, 2021.

²¹ Burnel, Anne: *Si le courriel est l'avenir de la lettre... Enjeux juridiques, technologiques et patrimoniaux de l'archivage du courrier électronique*, in: *Gazette des archives* (214), 2009, S. 79–85.

²² Vellino, André; Alberts, Inge: *Assisting the appraisal of e-mail records with automatic classification*, in: *Records Management Journal* 26 (3), 2016, S. 293–313. Online: <<https://doi.org/10.1108/RMJ-02-2016-0006>>.

²³ Vinh-Doyle, William P.: *Appraising email (using digital forensics): techniques and challenges*, in: *Archives and Manuscripts* 45 (1), 02.01.2017, S. 18–30. Online: <<https://doi.org/10.1080/01576895.2016.1270838>>.

²⁴ Schneider, J.; Adams, C.; DeBauche, S. u. a.: *Appraising, processing, and providing access to email in contemporary literary archives*, in: *Archives and Manuscripts* 47 (3), 2019, S. 305–326. Online: <<https://doi.org/10.1080/01576895.2019.1622138>>.

²⁵ Zuchet, Mike: *Pilotprojekt zur Langzeitarchivierung digitaler E-Mail-Korrespondenzen des Bundesvorstandes der Vereinten Dienstleistungsgewerkschaft ver.di.*, in: *Verein Deutscher Archivare: Fachgruppe Archivare an Archiven der Parlamente, der Politischen Parteien, Stiftungen und Verbände: Mitteilungen der Fachgruppe* 6 36, 2012, S. 165–170.

Vergleich „zeitlich und konzeptionell“ hinkt.²⁶ Trotzdem sollen hier einige Publikationen hervorgehoben werden, wie das soeben zitierte Werk von Maria Benauer von 2021, *E-Mails, ihr Wert und ihre Bewertung*, oder Corinna Knoblochs *Überlegungen zur Übernahme und Archivierung von E-Mail-Konten*²⁷ aus Österreich. Als Beispiele für Deutschland sind Patrick Sturms *Die E-Mail: Ein Kommunikationsmedium des frühen 21. Jahrhunderts quellenkundlich betrachtet*²⁸ oder Kristina Starkloffs *Überlieferung von E-Mail-Konten als genuin digitale Unterlagen. Archivwürdigkeit, Übernahmemethodik und Einblicke in die Entwicklung eines Werkzeugs*²⁹, beide von 2016 zu nennen. In der Schweiz wiederum ist die Forschung zur E-Mail-Archivierung bisher mehrheitlich durch Abschlussarbeiten vertreten. So zum Beispiel die nicht publizierte Masterarbeit von Armin Eberle *Die vorarchivische Selektion und archivische Bewertung von E-Mails*³⁰ oder die publizierte Zertifikatsarbeit von Beda Kupper *E-Mail-Archivierung*³¹ aus dem Jahr 2008.

2. Das Projekt „E-Mail-Archiv 2.0“

Beim Projekt „E-Mail-Archiv 2.0“ handelt es sich um ein von den zentralen Informatikdiensten der ETH Zürich (ID) geleitetes Projekt, das in Zusammenarbeit mit dem Hochschularchiv der ETH Zürich (HSA) sowie der Gruppe Forschungsdatenmanagement und Datenerhalt (FDD) der ETH-Bibliothek durchgeführt wird. Es handelt sich um ein mehrstufiges Projekt, wobei in dieser Masterarbeit lediglich auf die Bereiche der archivischen Bewertung, Übernahme und Erschliessung der E-Mails, sowie die Langzeitarchivierung im ETH Data Archive (Langzeitarchiv der ETH Zürich) eingegangen wird.

Die Ausgangslage für dieses Projekt ist, dass bisher innerhalb der ETH Zürich alle Objekte in Microsoft Outlook (E-Mails, Kalender, Termine oder Notizen), welche innerhalb von 30 Tagen nicht aus dem eigenen E-Mail-Postfach gelöscht wurden, automatisch in einem E-Mail-Archiv, dem sogenannten Vault, zwischengespeichert

²⁶ Benauer: *E-Mails, ihr Wert und ihre Bewertung*, 2020, S. 88.

²⁷ Knobloch, Corinna: *Überlegungen zur Übernahme und Archivierung von E-Mail-Konten*, in: *Digitale Archivierung. Innovationen – Strategien – Netzwerke*. Tagungsband zur 19. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ (Mitteilungen des Österreichischen Staatsarchivs 59/2016), Wien 2016, S. 221–231.

²⁸ Sturm: *Die E-Mail: Ein Kommunikationsmedium des frühen 21. Jahrhunderts quellenkundlich betrachtet*, 2016.

²⁹ Starkloff, Kristina: *Überlieferung von E-Mail-Konten als genuin digitale Unterlagen. Archivwürdigkeit, Übernahmemethodik und Einblicke in die Entwicklung eines Werkzeugs*, in: Puchta, Michael; Naumann, Kai (Hg.): *Kreative digitale Ablagen und die Archive: Ergebnisse eines Workshops des KLA-Ausschusses Digitale Archive am 22./23.11.2016 in der Generaldirektion der Staatlichen Archive Bayerns*, München 2017 (Sonderveröffentlichungen der Staatlichen Archive Bayerns), S. 39–43.

³⁰ Eberle, Armin: *Die vorarchivische Selektion und archivische Bewertung von E-Mails*. Masterarbeit im Rahmen des Masterstudiengangs Archiv- und Informationswissenschaft 2006-2008 an den Universitäten Bern und Lausanne, 2008.

³¹ Kupper, Beda: *E-Mail-Archivierung*, in: *Actualité archivistique suisse : travaux du certificat en archivistique et sciences de l'information = Archivwissenschaft Schweiz aktuell: Arbeiten aus dem Zertifikat in Archiv- und Informationswissenschaften*, Baden 2008, S. 88–117.

wurden.³² Dabei handelt es sich nicht um eine zertifizierte digitale Langzeitarchivierungslösung, sondern um das, was Karin Schwarz als ein „E-Mail-Archivierungssystem“ beschreibt. Dieser Begriff bezeichnet ein System, welches eigens für die Aufbewahrung von E-Mails entwickelt wurde und in welchem die E-Mails automatisch zwischengespeichert werden. Solche Systeme sind normalerweise auf eine mittelfristige, aber nicht dauerhafte Speicherung ausgelegt. In der Regel erfolgt vor der Speicherung in dieser Art von Zwischenarchiv keine archivistische Bewertung.³³

Aktuell befinden sich im E-Mail-Vault der ETH Zürich die E-Mail-Postfächer derjenigen Mitarbeitenden mit einer @ethz.ch-Adresse, die ihre E-Mails nicht vor Ablauf der Frist von 30 Tagen gelöscht haben. Im Rahmen des Projekts „E-Mail-Archiv 2.0“ Projekts sollen nun diese E-Mail-Konten bewertet und diejenigen als archivrelevant eingestuft werden durch das HSA und die Gruppe FDD langzeitarchiviert werden. Im Rahmen des Projekts soll von Seiten des Hochschularchivs definiert werden, welche E-Mail-Konten archiviert werden müssen, um die geschäftsrelevanten Vorgänge an der ETH Zürich zu dokumentieren. In enger Zusammenarbeit mit der Gruppe FDD und den ID müssen die Möglichkeiten des Exports und des Transfers ins Langzeitarchiv geklärt und die Arbeitsabläufe definiert werden.

Im Projekt geht es einerseits um eine retrospektive Bewertung der bereits im Vault gespeicherten Outlook-Objekte sowie um die prospektive Festlegung des weiteren Vorgehens. Es soll geklärt werden, wie in Zukunft welche E-Mails in welchem Rhythmus ans HSA abgeliefert werden und langzeitarchiviert werden sollen. Der im Projekt „E-Mail-Archiv 2.0“ definierte Prozess für die E-Mail-Archivierung an der ETH Zürich soll im Idealfall als Anwendungsbeispiel für die Langzeitarchivierung von E-Mails dienen können, das auch auf andere E-Mail-Bestände, zum Beispiel aus Privatnachlässen, angewendet werden kann.

2.1. Die rechtlichen Aspekte der E-Mail-Archivierung

Wenn es um die Archivierung von E-Mails geht, ist es wichtig, auch die rechtlichen Aspekte kurz zu beleuchten. Die Aufgabe des Hochschularchivs der ETH Zürich ist klar definiert. In Art. 1 des Reglements für das Archiv der ETH Zürich steht:

Das Archiv der ETH Zürich nimmt die Aufgaben eines öffentlichen Archivs gemäss dem BGA für die gesamte ETH Zürich sowie den ETH-Rat wahr. Es leistet damit einen Beitrag zur Rechtssicherheit sowie zur kontinuierlichen und rationellen Verwaltungsführung und schafft insbesondere Voraussetzungen für die historische und sozialwissenschaftliche Forschung. Damit dient es sowohl den Verwaltungen der ETH Zürich und des ETH-Rates als auch der Lehre und

³² ETH Zürich: Standard Dienstleistungsvereinbarung (Service based SLA) für „Mail und Groupware“, 12.10.2021.

³³ Schwarz: E-Mail-Archivierung, 2010, S. 558–559.

Forschung. Darüber hinaus sorgt es auch dafür, dass die Interessen der Öffentlichkeit im Rahmen der gesetzlichen Bestimmungen wahrgenommen werden können. Das Archiv wirkt zudem aktiv an der Erforschung und Vermittlung der Geschichte der Hochschule mit.³⁴

In diesen Bereich fallen auch diejenigen E-Mails aus dem Vault, die das Verwaltungshandeln widerspiegeln und für die historische und sozialwissenschaftliche Forschung von Bedeutung sind. Maria Benauer hielt im Zusammenhang mit der E-Mail-Archivierung in Österreich fest, dass „Einrichtungen und Stellen, bei deren Aufgabenerledigung E-Mails anfallen, archivrechtlich dazu angehalten sind, diese dem jeweils zuständigen Archiv anzubieten.“³⁵ Entsprechend kann auch mit der Anbietepflicht von E-Mails an der ETH Zürich argumentiert werden. Hinzu kommt, dass die Archivierung von E-Mails der historischen Forschung bisher ungeahnte Möglichkeiten der Auswertung eröffnet.³⁶

Aus archivischer Sicht sind E-Mails datenschutzrechtlich³⁷ nicht anders zu behandeln als analoge Korrespondenz und Geschäftsunterlagen, die in früheren Jahren ins HSA gekommen sind. So vergleicht auch Corinna Knobloch die Funktionen von E-Mails mit denjenigen eines Schriftstücks sowie gut geführte E-Mail-Konten mit klassischen Aktenablagen.³⁸ Ein weiteres Beispiel aus dem Ausland ist das Landesarchiv Baden-Württemberg, das, wie Robert Kretzschmar beschreibt, E-Mail-Accounts archiviert, weil das Archiv darauf reagieren musste, „*dass in den Behörden relevante E-Mails nicht selbstverständlich in den Geschäftsgang und zu den Akten gegeben werden*“.³⁹

Dabei ist festzuhalten, dass es hier ausschliesslich um Verwaltungsschriftgut geht, das im Kontext der Arbeitsprozesse von Mitarbeitenden an der ETH Zürich entstanden ist und somit der Anbietepflicht unterliegt. Die generelle Anbietepflicht für Verwaltungsschriftgut findet sich im Falle der ETH Zürich im Bundesgesetz über die Archivierung (BGA)⁴⁰, dann aber auch im oben erwähnten Reglement für das Archiv der ETH Zürich. In der Benutzungsordnung für Informations- und Kommunikationstechnologie an der ETH Zürich (BOT)⁴¹ ist in Art. 8^{bis} 4 festgehalten, dass archivwürdige E-Mails nach einer 10-jährigen Aufbewahrungsfrist vor der

³⁴ ETH Zürich: Reglement für das Archiv der ETH Zürich, 01.04.2015. Online: <<https://rechtssammlung.sp.ethz.ch/Dokumente/420.1.pdf>> Art. 1.

³⁵ Benauer: E-Mails, ihr Wert und ihre Bewertung, 2020, S. 93.

³⁶ Ebd.; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 9; Prom: Preserving Email (2nd Edition), 2019, S. 9; Pennock: Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages, 2006, S. 14.

³⁷ Bundesgesetz über den Datenschutz (DSG), 01.03.2019. Online: <https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/de>.

³⁸ Knobloch: Überlegungen zur Übernahme, 2016, S. 222; Schwarz: E-Mail-Archivierung, 2010, S. 551.

³⁹ Kretzschmar: Alles neu zu durchdenken?, 2014, S. 9.

⁴⁰ Bundesgesetz über die Archivierung (Archivierungsgesetz, BGA), 01.05.2013. Online: <<https://www.fedlex.admin.ch/eli/cc/1999/354/de>>.

⁴¹ ETH Zürich: Benutzungsordnung für Informations- und Kommunikationstechnologie an der ETH Zürich (BOT) und Anhang, 01.05.2022. Online: <<https://rechtssammlung.sp.ethz.ch/Dokumente/203.21.pdf>>.

Löschung dem Hochschularchiv angeboten werden müssen. Hier sind zudem die diejenigen Funktionsstufen aufgeführt, deren E-Mail-Accounts als archivwürdig eingestuft wurden. Die aktuelle Version der BOT ist nach Start des Projekts „E-Mail-Archiv 2.0“ in Kraft getreten und berücksichtigt bereits getroffene Bewertungsentscheide.⁴² Auf der öffentlichen Website der IT-Services der ETH Zürich ist darüber hinaus ebenfalls ersichtlich, dass Outlook-Elemente unter gewissen Bedingungen im Vault gespeichert werden.⁴³

Diejenigen Personen, die heute noch an der ETH tätig sind und deren E-Mails ins HSA kommen sollen, werden zusätzlich im Vorfeld individuell darüber informiert. Es soll vermieden werden, dass es zu einem ähnlichen Fall kommt wie im Landesarchiv Baden-Württemberg: Das Landesarchiv hatte den E-Mail-Account des früheren Ministerpräsidenten Stefan Mappus aus einem Server-Backup übernommen. Daraufhin verklagte Mappus das Landesarchiv und pochte auf die Löschung seiner E-Mails. Das Landesarchiv argumentierte, dass die E-Mails archiviert worden waren, um retrospektiv das Verwaltungshandeln nachzuvollziehen. Das Verwaltungsgericht und der Volksgerichtshof entschieden, dass die E-Mails gelöscht werden mussten. Die Begründung stützte sich jedoch nicht darauf, dass die E-Mails nicht archivwürdig gewesen seien. In diesem speziellen Fall begründeten die Gerichte ihren Entscheid damit, dass es sich um E-Mails aus einem Server-Backup handelte, die niemals zur Archivierung vorgesehen waren. Die Gerichte folgten Mappus' Argument, dass die E-Mails sein Privateigentum seien, nicht.⁴⁴

Die bei der Archivierung der E-Mails an der ETH Zürich anfallenden Personendaten in den Metadaten zu den E-Mail-Accounts sind an sich nicht besonders schützenswert. Es handelt sich dabei um Personen, deren Funktionsstufen in der BOT klar ersichtlich sind. Zudem ist öffentlich bekannt, wer an der ETH Zürich zu welchem Zeitpunkt diese Funktionen innehatte. So wie zu erwarten ist, dass sich Handakten aus der Zeit eines Rektors oder einer Rektorin im Hochschularchiv befinden, so ist auch zu erwarten, dass aus derselben Amtszeit E-Mails archiviert werden, welche das Verwaltungshandeln widerspiegeln. In Kapitel 6.1 wird darauf eingegangen, wie die E-Mail-Accounts dementsprechend bewertet wurden.

Eine weitere wichtige Frage ist, wie schützenswert der Inhalt dieser E-Mails ist und ob sie neben geschäftsrelevanten auch rein private Informationen enthalten. In einer Studie von Robert Capra, Julia Khanova und Sarah Ramdeen aus dem Jahr 2013 über den Gebrauch von Arbeits-E-Mails und privaten E-Mails gaben 88% der Befragten an, zumindest gelegentlich ihre Arbeits-Accounts auch für private Zwecke zu benutzen.⁴⁵ Auch Beda Kupper weist darauf hin, dass auch der Einsatz von

⁴² In den früheren Versionen der BOT fehlte diese Information noch, es wurde aber bereits in diesen Vorgängerversionen im Anhang darauf hingewiesen, dass E-Mails bzw. Daten dem Hochschularchiv angeboten werden können. Vgl. ETH Zürich: Benutzungsordnung für Informations- und Kommunikationstechnologie an der ETH Zürich (BOT) und Anhang, 01.06.2021.

⁴³ <https://ethz.ch/staffnet/de/it-services/katalog/email-kalender/mailarchiv.html>

⁴⁴ Kretzschmar: Alles neu zu durchdenken?, 2014, S. 14.

⁴⁵ Capra, Robert; Khanova, Julia; Ramdeen, Sarah: Work and personal e-mail use by university employees: PIM practices across domain boundaries, in: Journal of the American Society for

Funktions-Accounts keine klare Trennung zwischen privaten und geschäftlichen E-Mails ermöglicht.⁴⁶ Zusätzlich ist anzumerken, dass es gemäss BOT nicht verboten ist, den Geschäfts-Account für persönliche E-Mails zu verwenden. Es ist entsprechend davon auszugehen, dass auch in den E-Mails aus dem Vault persönliche E-Mails vorhanden sind. Aufgrund der grossen Datenmenge ist im Vault keine detaillierte oder systematische Suche nach besonders schützenswerten Personendaten möglich. Laut Brossi und Winkler ist es zwar ideal, wenn Mitarbeitende private E-Mails kennzeichnen – zum Beispiel durch einen Ordner „Privat“ - eine Institution kann jedoch davon ausgehen, dass E-Mails in einem Geschäfts-Account auch geschäftsrelevant sind, selbst wenn keine solch klare Kennzeichnung vorliegt.⁴⁷

Ein Teil derjenigen Personen, deren E-Mails als archivwürdig eingestuft wurden, besaßen an der ETH Zürich zwei oder noch mehr E-Mail-Accounts, die teilweise als privat betitelt waren. Diese privaten E-Mail-Accounts werden dementsprechend nicht in HSA übernommen. Für die Zukunft wurde in der aktuellen BOT, als Konsequenz des Projekts „E-Mail Archiv 2.0“, klar festgehalten, dass private E-Mails im Geschäfts-Account entweder gelöscht oder in einen bereitgestellten Ordner „Private do not archive“ verschoben werden müssen.⁴⁸ Dieser Ordner wird nicht im Vault zwischengespeichert und kann danach auch nicht langfristig archiviert werden. Somit kann bei der restlichen Korrespondenz davon ausgegangen werden, dass sie zumindest ab 2022 geschäftsrelevant ist und somit bei einem ausgewählten Personenkreis auch archivwürdig ist.

Trotzdem muss bei allen Accounts davon ausgegangen werden, dass die Möglichkeit besteht, dass sich auch nach BGA sogenannte „besonders schützenswerte Personendaten“ darin befinden.⁴⁹ Diese können selbstverständlich nicht nur in privater Korrespondenz, sondern auch in geschäftsrelevanten E-Mails vorkommen. Die Artefactual Systems and the Digital Preservation Coalition beschreibt dies sehr gut:

The body of an email is unstructured text which may contain personally identifiable information (PII), defined broadly as ‘any information relating to an identified or identifiable natural person’ (GDPR, Definitions, 2016). Email attachments may also contain PII or other sensitive personal data. [...] PII and

Information Science and Technology 64 (5), 2013, S. 1039. Online: <https://doi.org/10.1002/asi.22815>.

⁴⁶ Kupper: E-Mail-Archivierung, 2008, S. 94.

⁴⁷ Brossi, Pietro; Winkler, Maria: E-Mail Archivierung: Leitfaden zur Umsetzung und rechtliche Grundlagen für die Schweiz, Deutschland und Österreich, Rheinfelden 2008, S. 14.

⁴⁸ Vgl. zudem ETH Zürich: Standard Dienstleistungsvereinbarung (Service based SLA) für „Mail und Groupware“, 19.05.2022. Online: <https://ethz.ch/content/dam/ethz/associates/services/Service/IT-Services/files/sla/sla-mail-groupware.pdf>.

⁴⁹ Nach dem Bundesgesetz über den Datenschutz handelt es sich bei besonders schützenswerten Personendaten um Daten über: 1. die religiösen, weltanschaulichen, politischen oder gewerkschaftlichen Ansichten oder Tätigkeiten, 2. die Gesundheit, die Intimsphäre oder die Rassenzugehörigkeit, 3. Massnahmen der sozialen Hilfe, 4. administrative oder strafrechtliche Verfolgungen und Sanktionen. Vgl. hierzu Bundesgesetz über den Datenschutz (DSG), 2019.

sensitive personal data can also appear accidentally because the user contravenes organizational policies governing email use, because there are no such policies, or because the user employs an email account for both business and personal purposes.⁵⁰

Alle E-Mail-Konten werden deshalb im HSA sicherheitshalber mit einer Schutzfrist von 50 Jahren belegt.⁵¹ Vor Ablauf dieser Schutzfrist können die E-Mails nicht ohne vorgängiges Einsichtsgesuch und eine manuelle Prüfung durch die Archivmitarbeitenden eingesehen werden können. Einzelne Sachakten (ohne schützenswerte Personendaten) können so unter Umständen auch früher eingesehen werden.⁵²

E-Mails aus Privatnachlässen werden hier nicht diskutiert, da sie nicht der gesetzlichen Anbietepflicht unterliegen. Es ist den Schenkenden überlassen, ob sie dem HSA ihre E-Mails anbieten möchten oder nicht. Bei einer Übernahme ist darauf zu achten, dass eine archivische Bewertung der E-Mails stattfindet und die Benutzungsbedingungen im Schenkungsvertrag festgehalten sind.

3. Die Bewertung von Big Data

Im digitalen Zeitalter, in dem immer mehr Daten entstehen und die Kosten für Speicher vernachlässigbar gering sind, stellt sich die Frage, wie relevant die Bewertung vor der Archivierung noch ist. In der klassischen Bewertung handelt es sich nach Menne-Haritz um die „Identifizierung der archivwürdigen Teile des Schriftguts, das von einer Behörde nach der Aussonderung angeboten wird.“ Mit dem Ziel der „Feststellung des Sekundärwertes von Verwaltungsunterlagen als Voraussetzung für die dauerhafte Aufbewahrung“.⁵³ Früher ging es darum, möglichst viel zu archivieren, denn die Bestände waren kleiner und die Informationen sollten erhalten bleiben. Wenn man dies auf digitale Bestände anwendet, so sind diese viel umfangreicher als die traditionellen analogen Bestände und die Bewertung an sich ist ein sehr aufwändiger Prozess. Geoffrey Yeo stellte deshalb 2019 die bewusst provokante Frage: „*Can we – should we – try to keep everything?*“⁵⁴

Die Bewertung an sich unterlag schon immer Kritik: Viele Bewertungskriterien sind arbiträr und die Bewertung an sich kann unvorhersehbar, inkonsistent oder voreingenommen durchgeführt werden. Trotzdem wurde in der Archivwelt bisher kein Konsens in der Frage gefunden, ob stattdessen alle Daten für immer aufbewahrt

⁵⁰ Artefactual Systems and the Digital Preservation Coalition: Preserving Email, 2021, S. 3.

⁵¹ Laut BGA ist eine Schutzfrist von 50 Jahren die höchste zu vergebende Schutzfrist.

⁵² Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 9–19, 53.

⁵³ Menne-Haritz, Angelika: Schlüsselbegriffe der Archivterminologie: Lehrmaterialien für das Fach Archivwissenschaft, Marburg 2011 (Veröffentlichungen der Archivschule Marburg, Institut für Archivwissenschaft Nr. 20), S. 56.

⁵⁴ Yeo, Geoffrey: Can we keep everything? The future of appraisal in a world of digital profusion, in: Brown, Caroline (Hg.): Archival Futures, London 2019, S. 45.

werden sollen. Die Ordnung und Aussonderung der Daten ist immer noch ein Grundstein für die Auffindbarkeit und Nutzbarkeit von Inhalten bzw. von Information. Maria Benauer argumentiert, dass auch im digitalen Kontext grundsätzlich die gleichen Bewertungsmodelle angewendet werden können, wie man es im analogen Bereich getan hat und lediglich in der Praxis Anpassungen vornehmen muss.⁵⁵ Auch Terry Cook sagt deutlich: „*even if archivists could keep everything, they should not do so*“.⁵⁶ Er befasste mit der Bewertung Case Files, die ebenfalls so umfangreich sein können, dass Alternativen zu einer klassischen Bewertung gefunden werden müssen. Die wohl gängigsten Methoden mit solch umfangreichen Beständen umzugehen, ist entweder alles zu behalten oder alles zu vernichten, Samples anzufertigen oder nur die von den Archivarinnen oder Archivaren als wichtig eingestuft Dokumente zu überliefern. Alternativ kann man bei gleichförmigen Akten auch auf ein Zufalls-Sample zurückgreifen, wie es Neumayer und Rauber⁵⁷ sowie Buchholz⁵⁸ vorschlagen.⁵⁹

Auch von der technischen Seite her birgt die Archivierung von solchen Big Data auch immer ein gewisses Risiko, die langfristige Archivierung gewährleisten zu können. Um die Integrität der Daten zu sichern, müssen auch regelmässig Ressourcen in die Prüfung und Sicherung dieser Daten investiert werden. Dabei ist noch nicht eingerechnet, dass je mehr Daten vorhanden sind, auch mehr Daten migriert werden müssten, sofern ein Datei-Format nicht weiter unterstützt wird.⁶⁰

Andererseits muss bedacht werden, dass sich im Zeitalter von Big Data auch viel mehr Recherchemöglichkeiten auftun, wenn mehr Daten aufbewahrt und zugänglich gemacht werden. Die Datenintegrität könnte gewahrt werden und es würde nirgendwo mehr der Kontext fehlen, in dem die Daten entstanden sind. Es scheint keine Notwendigkeit mehr zu geben, Daten zu vernichten und es zeigt sich immer mehr ein Trend, der dahingeht, mehr Material aufzubewahren. David Holdsworth schrieb dazu in seinem Artikel *Strategies for digital preservation: „So If it costs very little to keep digital data, we might resist the temptation to discard those items of little*

⁵⁵ Benauer: E-Mails, ihr Wert und ihre Bewertung, 2020, S. 94–95.

⁵⁶ Cook, Terry: «Many Are Called, but Few Are Chosen»: Appraisal Guidelines for Sampling and Selecting Case Files, in: *Archivaria* 32, 1991, S. 33. Online: <<https://archivaria.ca/index.php/archivaria/article/view/11759>>.

⁵⁷ Neumayer, Robert; Rauber, Andreas: *Why Appraisal is not ‘Utterly’ Useless and Why It’s Not the Way to Go Either. A Provocative Position Paper (PPP)*, Wien 2007.

⁵⁸ Buchholz, Matthias: *Archivische Überlieferungsbildung im Spiegel von Bewertungsdiskussion und Repräsentativität*, Köln 2011 (Archivhefte / Landschaftsverband Rheinland, Archivberatungsstelle Rheinland 35).

⁵⁹ Yeo: *Can we keep everything?*, 2019, S. 45–49; Neumayer; Rauber: *Why Appraisal is not ‘Utterly’ Useless and Why It’s Not the Way to Go Either. A Provocative Position Paper (PPP)*, 2007, S. 1; Cook: «Many Are Called, but Few Are Chosen»: *Appraisal Guidelines for Sampling and Selecting Case Files*, 1991, S. 31–33; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives*, 2018, S. 38; Cox, Richard J: *Appraisal and the future of archives in the digital era*, in: Hill, Jennie (Hg.): *The Future of Archives and Recordkeeping*, 2010, S. 220. Online: <<https://doi.org/10.29085/9781856048675.012>>; Maitland, Jo: *Best Practices: Email Archiving*, Cambridge, Cambridge 2008, S. 6.

⁶⁰ Funk, Stefan E.: *Einführung Digitale Erhaltungsstrategien*, in: Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.3., 2010, S. 161.

*interest to us, but which later researchers might find valuable.*⁶¹ Ebenfalls zu beachten gilt es, dass es mittlerweile genügend Tools gibt, welche zumindest eine semiautomatische Verarbeitung von Daten ermöglicht. Gewisse Metadaten können automatisch extrahiert werden und es benötigt deutlich weniger manuelle Arbeit, um Bestände zu erschliessen. Dadurch, dass die Archivarinnen und Archivare schlichtweg nicht mehr in der Lage sind, solche Big Data manuell zu bearbeiten kann davon ausgegangen werden, dass die Tools zur Verarbeitung in Zukunft immer besser werden.⁶² Trotzdem müssen Archivare und Archivarinnen weiterhin an diesem Prozess beteiligt werden. So betont Susanne Belovari, wenn es um die Bewertung geht:

This cannot be left to IT and digital preservation staff who are unaware of record creators, appraisal approaches, institutional collection policies, and the whole universe of contextual and subject expertise that archivists bring to selecting materials for permanent retention.⁶³

Anstatt daran zu arbeiten, komplett vom Konzept der Bewertung weg zu kommen, ist es wohl realistischer, dass sich lediglich die Prozesse verschieben und dem neuen Umfeld anpassen. So wird in Zukunft sicherlich mehr ein Blick darauf geworfen werden müssen, welche Daten überhaupt entstehen und so abgelegt werden, dass sie in einem weiteren Schritt ohne grosse Zusatzbewertung archiviert werden können. Dadurch werden auch die Aufgaben des Record Management immer wichtiger werden. Gleichzeitig muss aber auch bedacht werden, dass mit dem steten technologischen Wandel auch die standardisierten Prozesse immer neu überdacht und angepasst werden müssen.⁶⁴

Final muss aber festgehalten werden, dass selbst wenn versucht werden würde, *alles* aufzubewahren, dieser Versuch zum Scheitern verurteilt wäre. Daten werden immer verloren gehen, weil sie nicht gepflegt wurden, weil sie unachtsam gelöscht wurden oder schlichtweg nicht gespeichert worden sind.⁶⁵

3.1. Bewertungsmodelle für E-Mail-Bestände

In ihrem Paper *The Future of Email Archives* beschreibt die Task Force on Technical Approaches for Email Archive die Problematik der Archivierung von E-Mails. Diese gewinnen immer mehr an Bedeutung als historische Quellen und doch gibt es noch

⁶¹ Holdsworth, David: Strategies for digital preservation, in: Deegan, Marilyn; Tanner, Simon (Hg.): Digital Preservation, 2013, S. 57. Online: <<https://doi.org/10.29085/9781856049863.005>>.

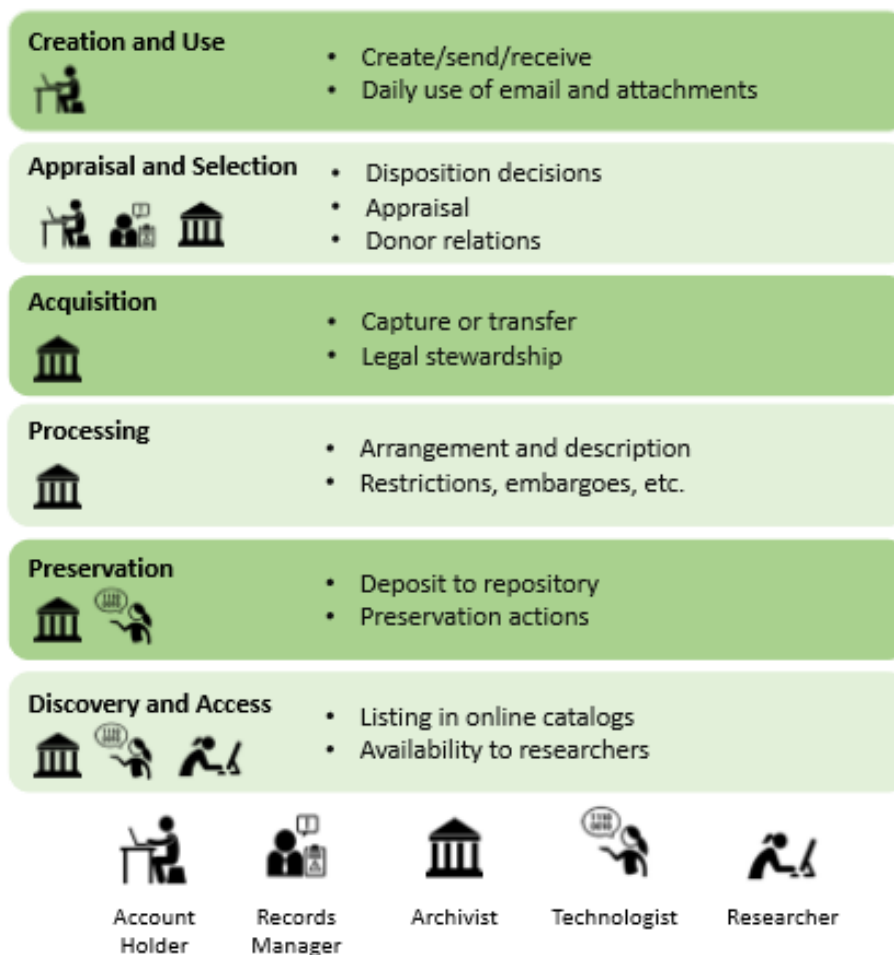
⁶² Yeo: Can we keep everything?, 2019, S. 48–51; Cox: Appraisal and the future of archives, 2010, S. 213–219.

⁶³ Belovari, Susanne: Expedited Digital Appraisal for Regular Archivists: An MPLP-Type Appraisal Workflow for Hybrid Collections, in: Journal of Archival Organization 16 (4), 2019, S. 215. Online: <<https://doi.org/10.1080/15332748.2019.1682793>>.

⁶⁴ Yeo: Can we keep everything?, 2019, S. 56; Cox: Appraisal and the future of archives, 2010, S. 217–230; Kretzschmar: Alles neu zu durchdenken?, 2014, S. 13.

⁶⁵ Yeo: Can we keep everything?, 2019, S. 54–56; Benauer: E-Mails, ihr Wert und ihre Bewertung, 2020, S. 94.

keine allgemeine Lösung, wie man E-Mails richtig bewertet, übernimmt, erschliesst, langzeitarchiviert und wieder zugänglich macht. In diesem Kapitel geht es um die Bewertung der E-Mails, ein Stadium im Lebenszyklus der E-Mails nach der aktiven Nutzung und vor der definitiven Übernahme der Daten durch das Archiv.⁶⁶ Der ganze Lebenszyklus von E-Mails wird von der Taskforce on Technical Approaches to Email Archives in ihrem Artikel *The future of email archives* dargestellt:



Grafik 1: Der Lebenszyklus von E-Mails⁶⁷

Im Idealfall sind Archivarinnen und Archivare, wenn es um die Übernahme von E-Mails geht, bereits vor der finalen Bewertung involviert. So kann bereits während der aktiven Nutzung von E-Mail-Konten darauf geachtet werden, dass z.B. logische Ordnerstrukturen gebildet oder E-Mails richtig benannt werden. Dadurch sind die E-Mail-Konten später leichter zu bewerten und archivieren. Sofern dies nicht geschehen ist, muss ein Weg gefunden werden, die relevanten E-Mails zu archivieren. Denn es ist, wie bei den meisten anderen behördlichen Unterlagen,

⁶⁶ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives*, 2018, S. 13–15.

⁶⁷ Ebd., S. 15.

davon auszugehen, dass ein Teil der E-Mail-Konten archivwürdig ist, ein anderer Teil jedoch nicht.⁶⁸

Eine Möglichkeit der Bewertung, die in den USA früh genutzt wurde, war es, jede E-Mail einzeln zu bewerten. Dabei wurde auf vorher festgelegte Aufzeichnungen und Leitlinien geachtet oder gar die Inhaberinnen und Inhaber der E-Mail-Konten gebeten, die E-Mails selbst zu bewerten und diese zu markieren, die übernommen werden sollten oder gar die E-Mails zu löschen, welche nicht ins Archiv sollten. Dabei können Tools, welche eben jenen Prozess erleichtern, wie ePADD der Stanford Universität und dem Universitäts-Archiv, helfen. In solchen Tools ist es nicht nur abliefernden Personen möglich, ihre Nachrichten zu sortieren, gewisse Entitäten können automatisch erkannt und zensiert werden, so dass kein Risiko besteht, dass Dinge wie Sozialversicherungsnummern in den E-Mails enthalten bleiben, die im Archiv landen. Dies ist jedoch ein sehr zeitaufwändiger Prozess und die automatische Erkennung von Entitäten funktioniert normalerweise auch nur in dem Sprachraum, in dem das Tool entwickelt wurde.⁶⁹ Tools zur automatischen Bewertung von E-Mail-Konten wie sie von Vellino und Alberts beschrieben werden könnten diese Probleme umgehen, sind aber noch nicht ausgereift genug, dass sie eine so geringe Fehlerquote aufweisen können, um sie flächendeckend einzusetzen.⁷⁰

Als Alternative zu dieser sehr zeitaufwändigen Einzelbewertung haben die National Archives and Records Administrations (NARA) 2013 einen alternativen Ansatz entwickelt: Den Capstone Approach.⁷¹ Bei diesem handelt es sich um eine Bewertungsmethode, in der ausschliesslich E-Mail-Konten von bestimmten Personen, sogenannten Capstones (auf Deutsch nach Benauer „Schlussteine“⁷²), innerhalb der Verwaltung übernommen werden. Das Hauptargument für diesen Ansatz ist, dass die meisten geschäftsrelevanten und somit als archivwürdig eingestuften E-Mails von diesen Personen gesendet oder empfangen werden. Dadurch, dass nicht jede E-Mail den gleichen Wert besitzt, kann Zeit und Arbeit gespart werden, indem nicht jede E-Mail eigenständig bewertet werden muss, weil stattdessen ganze E-Mail-Accounts übernommen werden. Der Capstone Approach bewegt sich im Rahmen des Macro Appraisal und bringt viele Vorteile mit sich: Zum einen bleibt so das Provenienzprinzip der Bestände erhalten und es kann eine repräsentativere Überlieferung erreicht werden, wenn komplette E-Mail-Accounts übernommen werden. Die überlieferten E-Mail-Postfächer sind zudem nach Benauer homogener und hochwertiger in der Qualität. Dies führt auch dazu, dass der

⁶⁸ Ebd., S. 3–14; Knobloch: Überlegungen zur Übernahme, 2016, S. 221.

⁶⁹ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 16; Stanford Libraries: ePADD, Projektwebseite, <<https://library.stanford.edu/projects/epadd>>; Vinh-Doyle: Appraising email (using digital forensics), 2017, S. 21–22; Maitland: Best Practices, 2008, S. 2–4.

⁷⁰ Vellino; Alberts: Assisting the appraisal of e-mail records with automatic classification, 2016.

⁷¹ National Archives and Records Administrations (NARA): Guidance on a New Approach to Managing Email Records, 2013-02, NARA Bulletin, 2013. Online: <<https://www.archives.gov/records-mgmt/bulletins/2013/2013-02.html>>.

⁷² Benauer: E-Mails, ihr Wert und ihre Bewertung, 2020, S. 105.

Evidenzwert dieser E-Mails erhalten bleibt, da die interne Ordnungsstruktur und die Metadaten der Postfächer erhalten bleibt.⁷³ Schliesslich gibt es auch einen rechtlichen Aspekt, den es zu beachten gibt, der für den Capstone Approach spricht:

Indem E-Mail-Accounts als Einheit herangezogen werden, ist eine systematische und überwachte Skartierung von E-Mails mit nur kurz- oder mittelfristigem Wert für das Records Management wesentlich einfacher realisierbar, was eine höhere Rechtssicherheit garantieren kann.⁷⁴

Eine andere Möglichkeit, E-Mails zu archivieren, ohne dabei komplette E-Mail-Konten zu übernehmen, ist es, E-Mails aus einem Records Management System (RMS) zu übernehmen. Dies ist ein Ansatz, der bereits 1994 von David Bearman in seinem Artikel *managing electronic mail* beschrieben wurde.⁷⁵ Innerhalb dieser RMS können E-Mails als Teil eines digitalen Geschäftsdossiers abgelegt und nach Abschluss geordnet ins Archiv übermittelt werden. Dadurch bleibt der Entstehungszusammenhang der E-Mails klar erhalten und kann überliefert werden. Als Beispiele sind hier die GLA Email Policy⁷⁶ zu nennen, in der Mitarbeitende in der Greater London Authority darauf hingewiesen werden, dass jede E-Mail direkt bewertet und wenn nötig in der entsprechenden Aktenablage gespeichert werden soll, sowie *La gestion et l'archivage des courriels: manuel partique*⁷⁷, in der neben praktischen Tipps zum Umgang mit E-Mails unter anderem auch Records Management, Archivierung und die Gesetzgebung für die E-Mail-Archivierung beschrieben werden. Wenn es darum geht, was archivwürdig ist und Eingang in ein RMS finden soll, ist die Schulung von Mitarbeitenden sehr wichtig. Mit dem *defensible deletion approach to email* empfiehlt James Lappin sogar, dass E-Mails, welche nicht gesetzlich oder für den normalen Arbeitsalltag gebraucht werden, regelmässig gelöscht werden sollen.⁷⁸

In der Theorie ist der Ansatz, dass alle geschäftsrelevanten E-Mails Eingang in ein RMS finden, sehr gut. Es zeigt sich aber leider, dass RM-Anleitungen nicht immer befolgt werden. Maria Benauer weist in ihrem Artikel *E-Mails, ihr Wert und ihre Bewertung* auf mehrere Studien hin, die belegen, dass RMS unzuverlässig und inkonsequent geführt werden und dadurch ein grosser Teil der geschäftsrelevanten

⁷³ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives*, 2018, S. 16–17, 54; Plante, Jeanette: "Challenges of Email as a Record Archiving Email Symposium." Presentation at Archiving Email Symposium, Department of Justice, 2015; Benauer: *E-Mails, ihr Wert und ihre Bewertung*, 2020, S. 103–105; National Archives and Records Administrations (NARA): *White Paper on The Capstone Approach and Capstone GRS*, 2015, S. 7–8. Online: <<https://www.archives.gov/files/records-mgmt/email-management/final-capstone-white-paper.pdf>>; Artefactual Systems and the Digital Preservation Coalition: *Preserving Email*, 2021, S. 7.

⁷⁴ Benauer: *E-Mails, ihr Wert und ihre Bewertung*, 2020, S. 106.

⁷⁵ Bearman: *Managing electronic mail*, 1994.

⁷⁶ Munn, David: *GLA Email Policy. Policy for managing email in the Greater London Authority*, London 2004.

⁷⁷ Bissonnette, Natalie: *Gestion des courriels : stratégies, technologies et bonnes pratiques*, in: *Archives* 44 (1), 2012, S. 77–113.

⁷⁸ Lappin, James; Jackson, Tom; Matthews, Graham u. a.: *The defensible deletion of government email*, in: *Records Management Journal* 29 (1/2), 2019, S. 49–51. Online: <<https://doi.org/10.1108/RMJ-09-2018-0036>>.

E-Mails keinen Eingang ins System findet. Andere E-Mails oder geschäftsrelevante Dateien werden falsch oder mehrfach abgelegt oder ausgedruckt und dann mit den Papierakten ins Archiv gegeben. Ausserdem wird selbst bei einer korrekten Ablage im RMS die Entscheidung, welche E-Mails am Schluss archiviert werden, nicht von den Archivaren und Archivarinnen getroffen, sondern von den Personen, welche die E-Mails ablegen. Diese Personen bewerten die Wichtigkeit der E-Mails normalerweise vor ihrem eigenen Hintergrund und nehmen nur wenig, wenn überhaupt, Rücksicht auf den Gesamtkontext. Daniel Guber weist zudem darauf hin, dass die „*schwer kontrollierbare Flut an E-Mails in der Geschäftstätigkeit*“⁷⁹ zu Problemen beim Records Management führen kann. Ein weiterer negativer Punkt ist, dass durch die Ablage der E-Mails im RMS der Entstehungskontext verloren geht und aus archivarischer Sicht eine Bestandsbildung nach dem Provenienzprinzip nicht mehr, oder nur sehr schwer, möglich ist.⁸⁰

Wie die NARA jedoch in ihrem Whitepaper zum Capstone Approach festhalten, schliessen sich diese beiden Methoden nicht gegenseitig aus:

It is important to note that the implementation of a Capstone approach for the management of agency email does not negate other recordkeeping responsibilities. As outlined in the Capstone Bulletin, agencies may still have a need to cross-file certain email with other records, such as case files or subject files. Agencies will remain responsible for ensuring that all other records, including program and project files, are still scheduled and managed appropriately. NARA will continue to receive other permanent record series from federal agencies outside the Capstone approach.⁸¹

Der Capstone Approach soll also mehrheitlich dazu dienen, diejenigen E-Mails zu archivieren, die keinen Eingang in ein RMS gefunden haben und soll nicht die korrekte digitale Aktenführung obsolet werden lassen.⁸²

Schliesslich gibt es noch einen dritten Ansatz, der vom Nationaal Archief in den Niederlanden eingesetzt wird: Die Hot-Spot-Methode. Bei dieser werden nicht Capstones definiert, sondern Ereignisse und Themen, die einen grossen Einfluss auf die entsprechende Organisation haben. Auf ihrer Website hat das Nationaal Archief ihre Hot-Spot-Kriterien festgehalten, die für andere Archive und Institutionen variieren können. Es handelt sich dabei um Themen, die eine gewisse Art von Unruhe

⁷⁹ Gubler, Daniel: E-Mail-Archivierung, die brennenden Punkte, in: Arbido: Records Management in Verwaltung und Privatwirtschaft – ein neues Aufgabenfeld? (2010/2), 2010.

⁸⁰ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 39–55; Benauer: E-Mails, ihr Wert und ihre Bewertung, 2020, S. 93–106; Prom: Preserving Email (2nd Edition), 2019, S. 9–10; Vinh-Doyle: Appraising email (using digital forensics), 2017, S. 19; Vellino; Alberts: Assisting the appraisal of e-mail records with automatic classification, 2016, S. 295–296; Bissonnette: Gestion des courriels : stratégies, technologies et bonnes pratiques, 2012, S. 87–89; Burnel: Si le courriel est l'avenir de la lettre... Enjeux juridiques, technologiques et patrimoniaux de l'archivage du courrier électronique, 2009, S. 85.

⁸¹ National Archives and Records Administrations (NARA): White Paper on The Capstone Approach and Capstone GRS, 2015, S. 8.

⁸² Benauer: E-Mails, ihr Wert und ihre Bewertung, 2020, S. 107.

verursacht haben und nicht um Themen, die eine Trend-Analyse von allgemeinen gesellschaftlichen Interessen abbilden.⁸³

Solche Hot-Spots werden auf einer Liste festgehalten und es wird zugleich definiert, ob die in diesem Zusammenhang entstandenen Dateien archivwürdig sind und wie lange sie aufbewahrt werden müssen. E-Mails werden im Nationaal Archief schliesslich aus einer Mischung gängiger Bewertungslisten, dem Capstone Approach und diesen Hot-Spot-Listen bewertet und übernommen. Alle drei können parallel angewendet werden und erlauben eine andere Interpretation des Kontextes, in dem Informationen entstehen und verarbeitet worden sind.⁸⁴

Maria Benauer spricht einen wichtigen Punkt an mit ihrer Aussage:

Für die Bewertung von E-Mails gibt es aus archivischer Sicht keine allgemein gültige Best-Practice-Strategie. Stattdessen gilt es angesichts der jeweils gegebenen Rahmenbedingungen, Vor- und Nachteile verschiedener Möglichkeiten abzuwägen, um individuell anwendbare Good-Practice-Lösungen zu erarbeiten. [...] Das einzige, was grundlegend falsch wäre, wäre nichts zu tun und damit gravierende Überlieferungslücken zu verantworten.⁸⁵

Es kann darüber debattiert werden, ob eine RMS-Lösung oder der Capstone Approach alleine oder in Verbindung eingesetzt werden müssen oder ob zusätzlich noch Hot-Spot-Listen wie in den Niederlanden erstellt werden müssen. Höchstwahrscheinlich gibt es nicht diese eine Lösung, die für alle Archive passt. Wichtig ist aber, dass klare Aufbewahrungs- und Ablagerichtlinien innerhalb der Institutionen oder Behörde festgelegt werden müssen, in denen geregelt ist, wie lange ein Geschäft oder eine E-Mail aufbewahrt werden muss. In diesem Zeitraum darf nichts gelöscht werden, was als archivwürdig eingestuft worden ist und einer Anbietepflicht unterliegt. Auch ist es höchst unwahrscheinlich, dass in Zukunft die Bewertung weiterhin manuell erfolgen wird. Vielmehr wird man sich immer mehr auf digitale Tools verlassen müssen, die die Bewertung und anderen Aufgaben innerhalb des Archivs teilweise oder vollständig übernehmen werden. Ansonsten werden solch

⁸³ Nationaal Archief Ministerie van Onderwijs, Cultuuren en Wetenschap: Hotspotlijst maken, <<https://www.nationaalarchief.nl/archiveren/kennisbank/hotspotlijst-maken>>, Stand: 14.05.2022.

⁸⁴ Nationaal Archief Ministerie van Onderwijs, Cultuuren en Wetenschap: Handreiking e-mail archiveren, <<https://www.nationaalarchief.nl/archiveren/kennisbank/handreiking-e-mail-archiveren#collapse-92082>>, Stand: 14.05.2022; Nationaal Archief Ministerie van Onderwijs, Cultuuren en Wetenschap: Selectiemethode bepalen. Informatieblad archiveren chatberichten, <<https://www.nationaalarchief.nl/archiveren/kennisbank/selectiemethode-bepalen>>, Stand: 14.05.2022; Nationaal Archief Ministerie van Onderwijs, Cultuuren en Wetenschap: Model selectielijst voor automatisch bewaarde e-mail conform de Handreiking bewaren van e-mail Rijksoverheid, Den Haag 2020. Online:

<https://www.informatiehuishouding.nl/binaries/informatiehuishouding/documenten/instrumenten/2020/05/18/model-selectielijst-voor-automatisch-bewaarde-e-mail-conform-de-handreiking-bewaren-van-e-mail-rijksoverheid/00828+Model+selectielijst+voor+automatisch+bewaarde+e-mail_V2.pdf>.

⁸⁵ Benauer: E-Mails, ihr Wert und ihre Bewertung, 2020, S. 113.

grosse Datenmengen nicht mehr, oder nur unzureichend, bearbeitet werden können.⁸⁶

4. Die technischen Grundlagen der E-Mail-Archivierung

In der Langzeitarchivierung gibt es nicht «den richtigen Ansatz». Christoph Prom beschreibt die Ausgangssituation sehr treffend in seinem Technology Watch Report:

Before beginning a preservation program, one must understand two basic technical details regarding the sending, receipt and storage of electronic mail messages:

- 1) email transmission and receipt is completely standardized around an open standard; and
- 2) processed surrounding storage and display are standardized only to a very limited extent. Each client and server implement storage and viewing in slightly different way.

The way in which these two facts play out in a local context will shape the decisions that an information professional will make when implementing preservation tools and services. We can take some comfort in the fact that even though each system stores email a little bit differently, there are multiple pathways to convert email from one form or another. In principle, this means that after selecting a preservation target format, staff can design effective workflows to acquire, process, preserve and provide access to email.⁸⁷

In der Langzeitarchivierung von E-Mails, bzw. allgemein ursprünglich bei digitalen Unterlagen wird normalerweise ein Repositorium bzw. ein Langzeitarchivierungssystem benötigt und die Möglichkeiten, wie archiviert wird, sind sehr vielseitig. Die Task Force on Technical Approaches to Email archives weist darauf hin, dass zwar im Idealfall bei der Frage der Erhaltung von E-Mails, deren ganzer Lebenszyklus von der Entstehung und Nutzung eines E-Mail-Accounts bis zur Vermittlung der archivierten E-Mails beachtet wird, in den meisten bisherigen Projekten aber die finale Langzeitarchivierung im Repositorium im Vordergrund steht.⁸⁸ Diese Repositorien sind sehr nützlich dabei, Daten mittels standardisierter Abläufe zu importieren. Sie können zudem automatisch Prüfsummen generieren und Fixity Checks durchzuführen, regelmässige Backups erstellen, File-Formate identifizieren und in beliebigen Abständen die importierten Daten validieren. Ein

⁸⁶ Artefactual Systems and the Digital Preservation Coalition: Preserving Email, 2021, S. 3; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 17, 39; Vinh-Doyle: Appraising email (using digital forensics), 2017, S. 21; Benauer: E-Mails, ihr Wert und ihre Bewertung, 2020, S. 110; Lappin u. a.: The defensible deletion of government email, 2019, S. 47.

⁸⁷ Prom: Preserving Email (2nd Edition), 2019, S. 15.

⁸⁸ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 58–59.

grosser Teil der Arbeit kann somit automatisiert werden. Dabei ist jedoch noch nicht festgelegt, wie die Daten aus dem Quellsystem exportiert werden und auf welchem Level die Archivierung stattfinden wird.

Die Task Force unterscheidet zwischen den vier aktuell gängigsten Export-Möglichkeiten: Beim direkten Export von E-Mails wird in Kooperation mit IT-Personal direkt aus dem Quellsystem exportiert und dabei bereits auf die wichtigsten Punkte, wie z.B. nötige Metadaten oder archivtaugliche Formate, geachtet. Diese Art von Export kommt häufiger in Institutionen vor, welche Richtlinien für die Archivierung von E-Mails haben, denn bei Privatpersonen, welche ihre E-Mails ins Archiv geben wollen und diese dafür selbstständig und ohne technischen Support abgespeichert haben. Beim webbasierten Export werden die E-Mails aus einer Web-Applikation heraus direkt in ein Repositorium überführt. Diese Methode ist anfälliger für Formatierungs- und Konvertierungsfehler und den Verlust von Metadaten, erlaubt aber eine grobe Bewertung durch die Archivarin oder den Archivar vor dem Export. Bei Client-Basierten Exports handelt es sich um den Export eines kompletten E-Mail-Kontos direkt aus dem vom User genutzten Programm. Das sogenannte Disk-Imaging findet statt, wenn Archivarinnen oder Archivare nicht Zugriff auf die Original-E-Mails haben, sondern lediglich ein Speichermedium erhalten, auf welchem die E-Mails bereits abgelegt sind. Die E-Mails können von diesem Speichermedium exportiert werden, da es sich dabei um eine zweite Speicherung handelt, muss mit den Formaten, Metadaten etc. gearbeitet werden, die vorgefunden werden.⁸⁹

Weiter beschreibt die Task Force drei Ansätze, wie diese E-Mails archiviert werden können: Bitspeicher, Migration und Emulation. Die Sicherstellung des Bitstreams, die redundante Speicherung und Erstellung von Prüfsummen für Integritätsprüfungen ist zwingend, um Risiken wie die unbeabsichtigte Löschung oder Veränderung von Dateien sowie Datenkorruption durch Bit Rot zu minimieren und stellt die Basis der digitalen Langzeitarchivierung dar. Dabei werden andere Aspekte wie die korrekte Darstellung der Dateien oder deren Kontext aussen vorgelassen. Sobald der Bitstream sichergestellt ist, können Gedanken zur Langzeitarchivierung gemacht werden, bei der die Speicherung und die mögliche Nutzung im Vordergrund stehen. Es bietet sich an, dass die Daten vor dem Import in ein Repositorium vorgängig noch einmal bearbeitet werden. Für die E-Mail-Archivierung ist es nicht anzuraten, nur den Bitstream sicherzustellen, da eine E-Mail neben wenigen strukturierten Daten, dem Header, auch sehr viele unstrukturierte Daten enthält, wie den Body, Anhänge oder verlinkte Inhalte. Damit diese zu einem späteren Zeitpunkt nutzbar gemacht werden können, bieten sich Verfahren wie die Emulation oder die Migration an: Bei der Emulation wird die komplette Entstehungs-Umgebung simuliert, so dass durch ein Programm auf das ursprüngliche Umfeld (Betriebssystem) mit der originalen Software zugegriffen werden kann. Dies ist eine Methode, die je nach E-Mail-Bestand sinnvoll sein kann, wenn die Arbeitsumgebung wichtig für das Verständnis

⁸⁹ Ebd., S. 44; Prom: Preserving Email (2nd Edition), 2019, S. 22–24; Funk, Stefan E.: Migration, in: Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.3., 2010, S. 169–170.

der archivierten E-Mails ist. Da die Emulation ein sehr aufwändiges und komplexes Unterfangen ist, werden E-Mails in der Regel migriert. Bei der Migration werden digitale Daten konvertiert, entweder von nicht archivwürdigen Formaten in ein archivwürdiges Format oder aber von mehreren Formaten in eine übersichtlichere Menge von Standardformaten. Dabei gibt es sehr viele unterschiedliche Empfehlungen zu archivtauglichen Formaten und für welche Dateien diese genutzt werden können und sollen. Als Beispiele sind hier die Berichte der Library of Congress⁹⁰, der NARA⁹¹ oder der Bibliothèque nationale de France⁹² zu nennen, wie im schweizerischen Kontext das Bundesarchiv⁹³ und die Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST)⁹⁴, die solche Standardformate gesammelt und beschrieben haben.⁹⁵

Migration birgt aber neben den Vorteilen immer ein gewisses Risiko, dass bei der Konvertierung Daten verloren gehen oder korrumpiert werden, was im schlimmsten Fall die Folge haben kann, dass die migrierte Datei nicht lesbar ist. Eine manuelle Qualitätskontrolle sollte am Ende einer Migration unternommen werden, um solchen Datenverlust zu verhindern. Es gilt entsprechend, die Vor- und Nachteile des Migrationsansatzes abzuwägen und es empfiehlt sich ohnehin, bei einer Migration auch das Original-Format aufzubewahren. Die Migration ist sehr beliebt, wenn es um die E-Mail-Archivierung geht. Wenn mit gängigen Tools zur E-Mail-Archivierung gearbeitet wird, sind diese meist auf spezifische Formate für die Archivierung festgelegt, in den meisten Fällen auf MBOX für ganze Postfächer und EML für einzelne E-Mails.⁹⁶

Ein weiterer wichtiger Punkt bei der E-Mail-Archivierung ist die Authentizität von E-Mails, ohne die keine authentische Überlieferung von E-Mails möglich ist. Im Idealfall werden nicht nur signifikante Metadaten festgehalten, sondern auch wichtige Arbeitsschritte, wie die Herkunft, Kontextinformationen, Konvertierungsprozesse und Fixity Checks dokumentiert. Die Auswahl, welche Informationen festgehalten werden, schliesst auch immer ein, dass es andere Informationen gegeben hätte, die nicht festgehalten werden. Es muss also eine Abwägung stattfinden, welche Informationen

⁹⁰ Library of Congress: Recommended Format Specifications 2021-2022, 2021. Online: <<https://www.loc.gov/preservation/resources/rfs/RFS%202021-2022.pdf>>.

⁹¹ National Archives and Records Administrations (NARA): Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images, o. D. Online: <<https://www.archives.gov/files/preservation/technical/guidelines.pdf>>.

⁹² Bibliothèque nationale de France: Formats de données pour la préservation à long terme : la politique de la BnF. Version initiale pour appel à commentaires, 2021. Online: <https://www.bnf.fr/sites/default/files/2021-04/politiqueFormatsDePreservationBNF_20210408.pdf>.

⁹³ Schweizerisches Bundesarchiv: Standards für die Archivierung digitaler Unterlagen, 04.2020.

Online:

<https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte_und_weisungen/archivtaugliche_dateiformate.1.pdf.download.pdf/archivtaugliche_dateiformate.pdf>.

⁹⁴ Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST): Katalog archivischer Dateiformate (KaD), Version 6.2, 12.2021, <https://kost-ceco.ch/cms/kad_main_de.html>.

⁹⁵ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 57–60; Prom: Preserving Email (2nd Edition), 2019, S. 11, 18–19.

⁹⁶ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 59.

für die Nachnutzung und Nachvollziehbarkeit notwendig sind.⁹⁷ Die Task Force on Technical Approaches for Email Archive fasst es aber gut zusammen, wenn sie schreibt:

For email collections, authenticity might be formally demonstrated through a series of documented processes and events. For example, if a business email was created and maintained in the course of normal business using an approved email application, if it contains the official email address and key elements of the message (e.g., header information, attachments, signature blocks), and if it is preserved as part of recordkeeping processes under the terms of a records retention schedule and in a system that prevents unauthorized modification, it is a reasonable assumption that the message is authentic.⁹⁸

4.1. Auszulesende Metadaten

Einer der grössten Vorteile von Born Digitals ist, dass sie bereits Metadaten enthalten, die bei analogen Medien zuerst manuell erfasst werden müssen. Metadaten können qualitative Kriterien wie Authentizität, Nachnutzbarkeit oder Integrität nachweisen. Zu einem guten Datenmanagement gehört es, dass gewisse vordefinierte Metadaten für jeden Datensatz vorhanden sein müssen, damit die Daten aufgefunden und genutzt werden können. Metadaten werden normalerweise in Sets zusammengefasst und nach Typologien geordnet. Im angelsächsischen Raum werden drei Typen von Metadaten unterschieden: Deskriptive, Strukturelle und Administrative.⁹⁹

Die automatische Extraktion von Metadaten kann den Prozess der Metadaten-Erfassung deutlich vereinfachen. Die ausgelesenen Metadaten können wiederum dabei helfen, sich bei Suchen in grossen Beständen wie E-Mail-Postfächern besser zurechtzufinden. Die automatisch extrahierten Metadaten dienen in diesem Fall als eine Alternative zu den traditionellen Findmitteln, da sie mehrere Möglichkeiten bieten, eine bestimmte E-Mail zu finden. Die automatische Metadatenextraktion wurde in diversen Tools für die E-Mail-Verarbeitung, z.B. ePADD, integriert. Es muss dabei jedoch beachtet werden, dass nur so viele Metadaten ausgelesen werden

⁹⁷ Benauer: E-Mails, ihr Wert und ihre Bewertung, 2020, S. 96–97; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 47; Waugh, Dorothy; Russey Roke, Elizabeth; Farr, Erika: Flexible processing and diverse collections: a tiered approach to delivering born digital archives, in: Archives and Records 37 (1), 02.01.2016, S. 8. Online: <<https://doi.org/10.1080/23257962.2016.1139493>>; Pennock: Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages, 2006, S. 9.

⁹⁸ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 46.

⁹⁹ Ebd., S. 8; Morisod, Pascal: Données et métadonnées, aux sources de l'information numérique, in: Arbido: Metadaten – Datenqualität (2017/3), 2017; Quiquerez, Fanny; Minana, Floriane: La gestion des données de la recherche: enjeux et perspectives, in: Arbido: Metadaten – Datenqualität (2017/3), 2017.

können, wie vorhanden sind. Wenn also zum Beispiel nichtssagende Titel für E-Mails gewählt wurden, kann bei der automatischen Metadaten-Auslese auch nicht ein besserer Titel gesetzt werden.¹⁰⁰

Eine der wichtigsten Fragen, die sich vor der Metadatenextraktion stellt, ist, *welche* Metadaten ausgelesen und gespeichert werden sollen. Dies kann nicht eindeutig beantwortet werden, sondern muss auf die jeweiligen Bedürfnisse abgestimmt werden. Einige Metadaten werden automatisch beim Ingest in ein Repositorium kreiert, wie die Angabe zum identifizierten Dateiformat, das Datum des Ingests, allfällige Bearbeitungen und Versionen. Diese Informationen werden im System festgehalten und müssen nicht noch zusätzlich aus den Dateien extrahiert werden. Metadaten, die häufig noch automatisch mit Hilfe eines Tools ausgelesen werden, sind strukturelle Metadaten, kontextuelle Metadaten und inhaltliche Metadaten. Dadurch ist es zum Beispiel möglich zu sehen welcher Kollektion die Dateien angehören und in welchem Kontext sie zueinander stehen, welche Titel die Dateien tragen, von wem sie erschaffen wurden und wer Zugriff auf diese hatte. Andere Metadaten, wie Angaben zur Bewertung, dem Vorgehen bei der Archivierung oder Schutzfristen, müssen jedoch immer noch händisch eingegeben werden oder zumindest manuell eingegeben und auf mehrere Datensätze angewendet werden.¹⁰¹ In vielen Fällen wird zusätzlich für jede Datei ein einzigartiger Digital Object Identifier (DOI) vergeben.¹⁰²

Besonders viele relevante Metadaten finden sich im Header einer jeden E-Mail. Dabei handelt es sich um die Angaben, die in E-Mails normalerweise direkt visualisiert werden, wie Absender, Empfänger, weitere Empfänger (einschliesslich CC aber nicht BCC), Betreff und Sendedatum. Corinna Knobloch sieht einen deutlichen Mehrgewinn darin, diese Metadaten zu sichern, stellt den Nutzen für die Zukunft von erweiterten Header-Informationen wie dem Wissen um IP-Adressen jedoch in Frage.¹⁰³

¹⁰⁰ Bunn, Jenny: Born digital archive cataloguing and description, DPC Technology Watch Report, 2021, S. 1. Online: <<https://doi.org/10.7207/twgn21-05>>; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 21.

¹⁰¹ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 49; Shepherd, Elizabeth; Yeo, Geoffrey: Managing Records: a handbook of principles and practice, 2013, S. 130, 138–139. Online: <<https://doi.org/10.29085/9781856049788>>, Stand: 14.05.2022.

¹⁰² Dadurch, dass durch die Vergabe von einem DOI direkt auf ein digitales Objekt zugegriffen werden kann, ist es in bestimmten Fällen nicht ratsam, einen solchen Identifier zu vergeben. Bei der digitalen Archivierung von Datensätzen, welche für einen längeren Zeitraum unter Schutzfrist stehen und nur nach vorgängigem Einsichtsgesuch gesichtet werden können, werden deshalb im ETH Data Archive keine DOIs vergeben. Deshalb werden auch im Projekt „E-Mail-Archiv 2.0“ keine DOIs vergeben.

¹⁰³ Knobloch: Überlegungen zur Übernahme, 2016, S. 225; Schwarz: E-Mail-Archivierung, 2010, S. 556–557.

4.2. Archivwürdige Formate für E-Mails

Wenn es um die Frage geht, in welchem Datei-Format E-Mails archiviert werden sollen gibt es ebenfalls kein objektives Richtig und Falsch, sondern nur Entscheidungen, die auf Grund der eigenen Bedürfnisse gefällt werden. Es gilt hauptsächlich, die Risiken und Vorteile, zum Beispiel bei der Migration in ein archivtaugliches Format, gegeneinander abzuwägen. Dabei hat die Übernahme von Ablieferungen aus einem System den grossen Vorteil gegenüber Ablieferungen von Privatpersonen, dass höchstwahrscheinlich weniger Formate vorhanden sind und diese eine gewisse Einheitlichkeit vorweisen – basierend auf der Aufbewahrungspolitik der Institution.¹⁰⁴

Häufig werden E-Mails zuerst in einem proprietären Format erstellt und für die Langzeitarchivierung in ein nicht-proprietäres Format wie MBOX, PST oder EML umgewandelt. Diese Formate enthalten alle ein eigenes Set von Metadaten, die unterschiedlich relevant sein können und nicht nur vom Format, sondern auch vom System abhängen, aus welchem die Dateien extrahiert wurden.¹⁰⁵

Viele Institutionen archivieren ihre E-Mails in den nicht proprietären Formaten MBOX und EML und konvertieren, wenn E-Mails in proprietären Formaten vorliegen. ePADD importiert E-Mails nur im MBOX Format¹⁰⁶, die Archives Nationales archivieren in EML sowie MBOX oder PST für ganze Postfächer¹⁰⁷ und die Task Force on Technical Approaches for Email Archive spricht sich für eine Archivierung in diesen Formaten aus.¹⁰⁸ Die Artefactual Systems and the Digital Preservation Coalition hat eine komplette Übersicht über Formate erstellt, welche für die Speicherung von E-Mails genutzt werden können, und in welcher EML und MBOX als archivwürdig eingestuft werden¹⁰⁹:

¹⁰⁴ Artefactual Systems and the Digital Preservation Coalition: Preserving Email, 2021, S. 3; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 58.

¹⁰⁵ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 27, 45; Shepherd; Yeo: Managing Records, 2013, S. 120.

¹⁰⁶ Stanford Libraries: ePADD.

¹⁰⁷ Sin Blima-Barru, Martine: Archives nationales – Archiver les courriers électroniques, Séminaire annuel des Archives de France (Dijon, 20-22 novembre 2019), Dijon 2019, S. 1.

¹⁰⁸ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 59.

¹⁰⁹ Artefactual Systems and the Digital Preservation Coalition: Preserving Email, 2021, S. 3–6.

File format	Extensions	Brief Summary
Internet Message Format	.eml	<p>EML is a text-based format designed to conform to the original RFC email specification (RFC822, 1982) and its successors. EML files are typically single email messages with attachments either embedded as encoded base64 content in the file or referenced from the file. The messages (but typically not the attachments) can be opened and read in any text editor. EML can be exported from and rendered by many email programs, including Microsoft Outlook, Zimbra, Eudora, Mozilla Thunderbird, Apple Mail, and Gmail. There is also good tool support for converting email exports in other formats to EML. A useful list of tools used in capturing, converting, and preserving emails as EML and other formats is provided by the Task Force on Technical Approaches to Email Archives (Email Archiving Tools, 2018).</p> <p>EML is based on RFC standards, but there is no specification for the format itself, and the ability to manage threads, attachments, and other features relating to the internal structure of email accounts is limited (EA-PDF Working Group, p.26). Because EML is text-based and ubiquitous it may well be renderable for the foreseeable future. Some major email preservation projects, such as Harvard University's EASi platform, migrate other formats to EML for preservation.</p> <p>EML messages can be extracted from email accounts and stored as Maildir, a directory format structured to represent the delivery status of the messages. Maildir can be exported from and loaded into a number of email applications; however, it lacks detailed specifications and has some file system compatibility issues which may require remediation before the files can be read by certain email programs (Wikipedia, Maildir, 2020). The open-source OfflineIMAP tool can be used to extract email accounts remotely</p>

		<p>and store them as MailDir.</p> <p>Library and Archives Canada and the US National Archives and Records Administration both consider EML files to be preferred transfer formats for email (LAC, 2015; NARA, 2020).</p>
MBOX	.mbox	<p>Like EML, MBOX is a text-based format designed to implement RFC email standards. Unlike EML, however, a single MBOX file can contain an aggregation of messages, typically the contents of a folder. MBOX is well supported by numerous email clients; a user can export and store an account as a series of MBOX files, each representing a folder of messages with attachments embedded in them as base64 encoded text. These files can be imported into other email clients or read in EPADD, an open-source tool designed by Stanford University to support ‘appraisal, processing, preservation, discovery, and delivery of historical email archives’ (Stanford Libraries, 2021). University of North Carolina’s RATOM (Review, Appraisal, and Triage of Mail) project has also developed tools for working with MBOX files.</p> <p>A single MBOX file can encapsulate an entire email folder, which results in more resilient relationships between email messages and threads and makes it easy to load the folder into an email client or a tool like EPADD. However, due to the large volume of emails that can reside in a single folder, MBOX files can often become excessively large and unwieldy; it also means that, in some cases, the corruption of a single message can prevent the entire file from loading or opening (EA-PDF Working Group, p. 26; DPC and Prom, 2019, p.17). MBOX as a preservation format also suffers from the fact that it is a family of formats with four variants (MBOXO, MBOXRD, MBOXCL and MBOXCL2) that are not entirely compatible with one another</p>

		<p>(Library of Congress (LC), 2016). Despite these drawbacks, MBOX files show considerable promise as both a preservation and access format, given the format's ubiquity, tool support, ability to represent entire email folders, and underlying text-based structure.</p> <p>Library and Archives Canada and the United States Archives and Records Administration consider MBOX to be a preferred email transfer format (LAC, 2015; NARA, 2020).</p>
<p>Microsoft Outlook Personal Folders</p>	<p>.pst .olm</p>	<p>PST is a proprietary format developed by Microsoft to support email download, storage, and offline use. It has two versions, ANSI (the earlier of the two) and Unicode, which are 'differentiated primarily by software implementation versions, character sets, maximum file size constraints and bit values' (LC, 2013). A PST file can be an entire email account or a portion of one, and can include not only messages but also contacts, calendar items, tasks, and other products of Outlook and related tools.</p> <p>Microsoft maintains a freely available specification for PST (Microsoft, [MSPST], 2020). The ubiquity of Microsoft Outlook and related tools makes PST a relatively common format. Library and Archives Canada lists PST as an acceptable 'transfer format', with the preferred format being MBOX and EML, while the National Archives and Records Administration of the United States lists all three as preferred formats (LAC, 2015; NARA, 2020). The RATOM project's toolset works with PST files in addition to MBOX.</p> <p>However, be aware that in Mac implementations of Microsoft Outlook the export format is called Outlook for Mac (OLM). OLM does not have an openly documented specification like PST and is subject to unidentified changes (Microsoft, 2015). There are a number of free and paid</p>

		OLM to PST converters such as Softaken and Mailvita, but testing and quality assurance is advised before selecting a tool.
Portable Document Format	.pdf	<p>When single email messages or threads are converted to PDF or PDF/A, they lose much of their metadata and attachments can become inaccessible (CLIR, 2018, pp.82-83). However, a dedicated task force on email preservation recently proposed a specification ‘to set the stage for software developers to create email capture and representation systems leveraging PDF to represent the core metadata, content, attributes, and context’ of email accounts’ (EA-PDF Working Group, 2021, p.7). The authors argue that MBOX and EML ‘are actually somewhat amorphous families of formats defined more by client implementations than by authoritative specifications’ and that ‘PDF provides a means to represent these implementations in a normalized packaging model, regardless of the underlying source.’ (EA-PDF Working Group, 2021, p.8).</p> <p>PDF versions of email messages are highly accessible due to widespread software support for the format, and the efforts of the EA-PDF Working Group demonstrate how PDF could be enhanced to serve as both a preservation and access format for entire email accounts rather than individual messages.</p>
Extensible Markup Language	.xml	XML has been explored as a preservation format by a number of digital preservation projects, including PeDALS, CERP, and DArcMail. DArcMail is the most recent of these, and the successor to CERP. DArcMail, which stands for Digital Archives for Email, is run by the Smithsonian Institution Archives, and is the name of both the project and a python-based opensource tool available on the project’s homepage. DArcMail converts MBOX files to XML and includes a graphical

		user interface that 'supports appraisal, processing, and providing access to email accounts or large groups of related email messages.' (Smithsonian Institution Archives, 2017, p. 3). The XML file can represent an entire email account, retaining internal folder structure; header and other metadata; and messages and their attachments. If the Smithsonian Institution Archives receives email in nonMBOX formats, it converts them to MBOX and then uses DArcMail to convert the MBOX files to XML (Smithsonian Institution Archives, 2017, pp.4-5).
Microsoft Outlook Item	.msg	An MSG file is a single email or other Microsoft Outlook Item which can be downloaded by Outlook end users. Unlike EML or MBOX, MSG files cannot be read in text editors; however, Microsoft maintains a published specification for the format (Microsoft, [MS-OXCMSG], 2020). MSG is not likely to be used for transferring entire email accounts to preserving institutions, but files in that format may appear in transfers of diverse bodies of digital objects from network drives, personal computers, or electronic recordkeeping systems. There has been little discussion in the digital preservation community of MSG as a preservation format, although the open specification and ongoing Microsoft support for MSG suggest that it will be readable for the foreseeable future. When capturing email accounts, archivists should avoid MSG in favour of MBOX, EML, or PST.

Weniger beliebt bei der Langezeitarchivierung sind proprietäre Formate wie PST und MSG. E-Mails können zwar normalerweise in diesen Formaten exportiert werden, sind aber tendenziell weniger geeignet für die Archivierung, da sie nicht als Text-Datei gelesen werden können und nur von Outlook, der proprietären Software von Microsoft gelesen werden kann.¹¹⁰ Es handelt sich aber trotzdem um sehr

¹¹⁰ Beim Argument der Proprietät ist jedoch anzumerken, dass Microsoft seine Format-Spezifikationen öffentlich zugänglich gemacht hat. Dies bedeutet, dass auf Grundlage derer auch Tools entwickelt werden könnten, welche diese Formate trotzdem auslesen können.

weitverbreitete Formate und es ist nicht davon auszugehen, dass sie in den nächsten Jahren verschwinden.¹¹¹

Es gibt jedoch auch viele Institutionen, die auf eine Archivierung in diesen empfohlenen Formaten verzichten und stattdessen E-Mails in das PDF-, beziehungsweise das für die Langzeitarchivierung noch besser geeignetere PDF/A-Format konvertieren.¹¹² Dabei ist zu unterscheiden, ob E-Mails bereits im PDF-Format aus dem Quellsystem exportiert werden oder nach dem Export in dieses Format konvertiert werden. Wenn eine E-Mail im PDF- oder PDF/A-Format exportiert wird, ist es durchaus wahrscheinlich, dass Metadaten verloren gehen, die in einem anderen Format erhalten bleiben würden. Dafür eignet sich PDF/A besonders aus der Perspektive der Benutzenden, da der Inhalt visuell mehrheitlich gleich bleibt.¹¹³ Vor allem im deutschsprachigen Raum konvertieren mehrere Archive ihre E-Mail Bestände ins PDF/A-Format. Als Beispiele sind das Archiv der sozialen Demokratie (AdsD) der Friedrich-Ebert-Stiftung¹¹⁴, das Österreichische Staatsarchiv¹¹⁵ oder das Landesarchiv Baden-Württemberg¹¹⁶ zu nennen.

5. Grundlagen der Erschliessung von E-Mails

Angelika Menne-Haritz definiert Erschliessung folgendermassen:

Sie besteht aus den beiden Tätigkeiten der Verzeichnung und Ordnung von Archivgut als Einheit bei der Vorbereitung zur Bereitstellung der Bestände. Dazu gehören die Analyse der Entstehungszusammenhänge und der Strukturtypen, die Titelbildung für die einzelnen Verzeichnungseinheiten, der Aufbau der Gliederung und die Eingliederung der einzelnen Verzeichnungseinheiten in die Gesamtstruktur sowie die Darstellung der Geschichte des Bestandes und seiner Entstehungsbehörde in den Zusatzinformationen etwa für die Findbucheinleitung. Ergebnis der Erschliessung ist der auswertbare Bestand, repräsentiert im Findbuch.¹¹⁷

¹¹¹ Murray, Kate: Shaking the Email Format Family Tree, The Signal, 04.04.2014, <<https://blogs.loc.gov/thesignal/2014/04/shaking-the-email-format-family-tree/>>, Stand: 29.05.2022; Artefactual Systems and the Digital Preservation Coalition: Preserving Email, 2021, S. 6.

¹¹² Im Folgenden wird jeweils der Begriff „PDF“ bzw. „PDF-Format“ verwendet. Damit ist immer das PDF/A-Format gemeint, welches in der digitalen Langzeitarchivierung gängig ist.

¹¹³ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 82–83; Fanning, Betsy: Preservation with PDF/A (2nd Edition), DPC Technology Watch Report, Digital Preservation Coalition, 2017, S. 12. Online: <<https://doi.org/10.7207/twr17-01>>.

¹¹⁴ Zuchet: Pilotprojekt zur Langzeitarchivierung digitaler E-Mail-Korrespondenzen des Bundesvorstandes der Vereinten Dienstleistungsgewerkschaft ver.di., 2012, S. 165–168.

¹¹⁵ Knobloch: Überlegungen zur Übernahme, 2016, S. 226.

¹¹⁶ Landesarchiv Baden-Württemberg: PDF/A als Anbieters- und Archivierungsformat. Richtlinie des Landesarchivs Baden-Württemberg, 05.06.2013. Online: <https://www.landesarchiv-bw.de/sixcms/media.php/120/PDF_richtlinie_mit_erlaeuterungen.pdf>.

¹¹⁷ Menne-Haritz: Schlüsselbegriffe der Archivterminologie, 2011, S. 64.

Diese Definition bezieht sich auf die Erschliessung von analogem Schriftgut und ist somit nicht ohne Anpassungen auf die digitale Erschliessung anzuwenden. Im Bereich der Erschliessung von digitalen Daten, in diesem Fall von grossen E-Mail-Beständen, stellen sich teilweise andere Herausforderungen, die Bestände zugänglich zu machen. Die Frage der Ordnung und der Beschreibung von digitalen Daten kam bereits in den 1990er Jahren auf. Während bei analogen Unterlagen die Ordnung in der traditionellen Erschliessung so gut wie möglich beibehalten wird, stellt sich bei elektronischen Ablieferungen häufig die Frage, ob die ursprüngliche Ordnung identifiziert und erhalten werden kann. Jane Zhang und Jenny Bunn sind ganz klar der Auffassung, dass auch in der Erschliessung von digitalen Unterlagen, wenn möglich, versucht werden sollte, die ursprüngliche Struktur, die Ordnung oder das Fehlen einer solchen, wiederzugeben. So kann im Idealfall der Entstehungskontext der Unterlagen nachvollzogen werden. Dieser Prozess wird etwas schwieriger, sofern einer Ablieferung und der darauffolgenden Erschliessung kein gut funktionierendes Records Management System zu Grunde liegt.¹¹⁸

Wenn man auf die Erschliessung von E-Mails schaut, ergibt sich zusätzlich die Schwierigkeit, dass bereits im Originalmedium eines Postfachs E-Mails unterschiedlich gefiltert und dargestellt werden können. Die Ordnung kann anhand der einzelnen Metadaten variabel herausgespielt werden und entsprechend können auch die Daten für die Erschliessung unterschiedlich genutzt werden. Ordnungssysteme innerhalb der Postfächer können vorhanden sein und dabei helfen, den Kontext der Daten im Archiv wiederzugeben. Diese Ordnung entsteht aber meist ohne die Hilfe von Archivarinnen oder Archivaren und ist mehrheitlich deshalb nachvollziehbar, weil man weiss, wie man aus den Metadaten von E-Mails deren Struktur ableiten kann. Es muss aber meist einfach damit gearbeitet werden, was vorgefunden wird. Hinzu kommt, dass bei solch grossen Datenmengen die Erschliessung auf Einzelstückebene kaum möglich oder sinnvoll ist. Bei der Erschliessung von E-Mails kommt zusätzlich der Umstand hinzu, dass innerhalb von E-Mail-Postfächern bereits mit Volltext gesucht werden kann und die Auffindbarkeit somit besser gegeben ist als bei einer detaillierten Erschliessung.¹¹⁹ In mehreren Paper von Spezialisten für E-Mail Archivierung, wie Jenny Bunn's *DPC Technology Watch Report*¹²⁰ oder dem Report *The future of email archives* der Task Force on Technical Approaches to Email Archives¹²¹ wird deshalb eine möglichst niederschwellige Art der Erschliessung für E-Mails vorgeschlagen: Der „More Product, Less Process“ Ansatz von Dennis Meissner und Mark Greene, den die

¹¹⁸ Zhang, Jane: Original Order in Digital Archives, in: *Archivaria* 74, 2012, S. 167–187; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives*, 2018, S. 7; Bunn: *Born digital archive cataloguing and description*, 2021, S. 23.

¹¹⁹ Zhang: *Original Order in Digital Archives*, 2012, S. 185–186; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives*, 2018, S. 7–8; Bunn: *Born digital archive cataloguing and description*, 2021, S. 26.

¹²⁰ Bunn: *Born digital archive cataloguing and description*, 2021.

¹²¹ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives*, 2018.

beiden 2005 zum ersten Mal publizierten.¹²² Ein Ansatz, der sich besonders für die Erschliessung von umfangreichen digitalen Beständen eignet, die einerseits bereits viele Metadaten enthalten, die nicht zuerst manuell erfasst werden müssen.¹²³

6. Die praktische Umsetzung im Hochschularchiv

Die Langzeitarchivierung von E-Mails an der ETH Zürich ist ein Prozess, in den mehrere Gruppen und Teams involviert sind. Während das Hochschularchiv und die Gruppe Forschungsdatenmanagement und Datenerhalt (FDD) sehr eng zusammenarbeiten, wenn es um die digitale Langzeitarchivierung geht, hat jedes Team bzw. jede Gruppe trotzdem ihre eigene Perspektive und entsprechendes Expertenwissen. Viele Tests, welche in den eher technischen Bereich des Datenerhalts fallen, wurden entsprechend von FDD durchgeführt, wo die Mitarbeitenden entsprechende Erfahrung haben. Dadurch, dass die Arbeitsabläufe für andere Ablieferungen von digitalem Material jedoch bereits definiert wurden und zum Beispiel ein Einheitsmapping für Metadatenfelder aus CMI und Rosetta besteht, oder auch bereits funktionstüchtige Ingest-Workflows mit CSV-Metadatendateien bestehen, konnte auf dieses bestehende Wissen zurückgegriffen werden. Die jeweiligen Entscheidungen wurden vom Hochschularchiv gefällt, aber immer in Absprache mit FDD oder anderen zuständigen Stellen, insbesondere wenn es um die technische Machbarkeit ging.

6.1. Die Bewertungsentscheide im Hochschularchiv

In Bezug auf die Bewertungsentscheidung des Hochschularchivs wurden alle oben genannten Punkte in Betracht gezogen und mit der Schulleitung der ETH sowie dem ETH-Rat besprochen. Wie für die Ausarbeitung von normalen Bewertungskriterien, welche im Idealfall mit den abliefernden Stellen besprochen werden, war es ein Anliegen, auch die Bewertung der E-Mail-Archivierung mit den betroffenen Stellen abzusprechen. Dabei stellte sich die Frage, ob die E-Mails nur aus einem RMS ins Hochschularchiv gelangen sollten, gar nicht. Einerseits natürlich, weil sich das Projekt „E-Mail Archiv 2.0“ auf ganze E-Mail-Konten bezieht und nicht auf allfällige Geschäftsverwaltungssysteme, andererseits weil an der ETH bisher noch kein allgemeingültiges RMS eingeführt wurde, aus welchem die E-Mails hätten übernommen werden können.

Es wurde entsprechend entschieden, nach dem Capstone Approach bestimmte Schlüsselfiguren zu definieren, welche innerhalb ETH Funktionen einnehmen bzw. einnahmen, sodass mit der Archivierung ihrer E-Mails der grösste Teil des

¹²² Greene, Mark; Meissner, Dennis: More Product, Less Process: Revamping Traditional Archival Processing, in: *The American Archivist* 68 (2), 2005, S. 208–263. Online: <<https://doi.org/10.17723/aarc.68.2.c741823776k65863>>.

¹²³ Belovari: *Expedited Digital Appraisal for Regular Archivists*, 2019, S. 197–198.

Verwaltungshandelns abgebildet werden kann. Dabei handelt es sich im genauen um alle Mitglieder der Schulleitung der ETH Zürich sowie den Präsidenten oder die Präsidentin und die Geschäftsführerin oder den Geschäftsführer (bzw. früher Generalsekretär und Generalsekretärin) des ETH-Rats. Von diesen Personen sollen retrospektiv alle E-Mails (mit Ausnahme der in Kapitel 2.1. erwähnten privaten E-Mail-Accounts) aus dem Vault ins Hochschularchiv übernommen werden. Gleichzeitig werden ihre E-Mail-Konten prospektiv als archivwürdig festgehalten und sollen in Zukunft in regelmässigen Abständen abgeliefert werden.

Zusätzlich zu den E-Mail-Konten dieser bestimmten Personen wurden auch die entsprechenden Funktions-Accounts als archivwürdig eingestuft. Bei diesen handelt es sich entweder um Shared-Mailboxen, die auch von Mitarbeitenden dieser als Capstones eingestuften Personen in deren Auftrag genutzt werden können oder um E-Mail-Accounts der Hochschulkommunikation zu einer bestimmten Funktion, die hauptsächlich für die Kommunikation zuständig sind.

Sollte ETH-weit ein RMS eingeführt werden und dieses korrekt geführt werden, kann der Entscheid, welche E-Mail-Konten archiviert werden überprüft und der Capstone Approach noch einmal überdacht werden. Bis dahin sollen aber alle diese E-Mail-Postfächer aus dem Vault ins digitale Langzeitarchiv überführt werden.

6.2. Metadaten-Extraktion im Hochschularchiv

Für die Extraktion von Metadaten wird im Hochschularchiv sowie in der Gruppe Forschungsdatenmanagement und Datenerhalt wird mit dem Apache Tika Toolkit gearbeitet.¹²⁴ Rosetta hat die üblichen Dublin Core Felder, so dass das Ziel war, eine Zuordnung der ausgelesenen Metadaten mit entsprechenden DC-Feldern festzuhalten. Um festzustellen, welche Metadaten relevant sind, musste in einem ersten Schritt festgelegt werden, welche Metadaten aus einer E-Mail, die aus dem Vault ins Hochschularchiv kam, herausgelesen werden können. Dabei zeigte sich schnell, dass es einen grossen Unterschied gab, welche Metadaten vorhanden waren, wenn eine E-Mail im EML oder im MSG Format aus dem Vault extrahiert wurde. Am auffälligsten ist das Fehlen aller Raw:Header-Angaben in den EML-Dateien. Diese beinhalten zwar teilweise die von Knobloch als für die zukünftige Nutzung fragwürdigen Kategorien betitelten Angaben wie Netzwerk ID oder Sender IP, aber auch den Hinweis, ob die E-Mail ein Attachment beinhaltet, was als sehr nützlich eingestuft wurde.

Es wurde entsprechend mit einem Test-Set von E-Mails aus mehreren E-Mail-Postfächern in den Formaten EML und MSG getestet, welche Metadaten mit Tika extrahiert werden konnten. Die Auswertung eines Test-Batches findet im Anhang 1, in der ersichtlich ist, welche Metadaten aus welchem Format extrahiert werden

¹²⁴ <https://tika.apache.org/>

können, bzw. in wie vielen Fällen die Metadaten nicht extrahiert wurden. Dabei zeigt sich deutlich, dass im MSG-Format wesentlich mehr Metadaten enthalten sind. Die von uns als wichtig eingestuft Informationen sind diejenigen über den Absender oder die Absenderin der E-Mail, den Empfänger oder die Empfängerin (sowohl die Adressatin und der Adressat sowie auch CC und BCC), das Sendedatum, der Titel der E-Mail sowie die Kontextinformation, ob die E-Mail einen Anhang enthält.

Auf Grund dieser Auswertung wurden die entsprechenden Felder festgelegt, welche ausgelesen und für den Ingest in Rosetta aufbereitet werden sollen. Zusammengefasst werden folgende Metadaten vor dem Ingest mit einem von FDD-Mitarbeitenden entwickelten Tool mithilfe von Tika extrahiert und in eine CSV-Metadatenfile überführt, aus welcher Rosetta während des Ingest die Metadaten in die entsprechenden Dublin Core-Metadatenfelder speichert. Die mehrfach aufgeführten Dublin Core-Metadatenfelder werden in Rosetta vervielfältigt und mehrfach dargestellt:

Metadaten-Feld	Mapping Rosetta	Inhalt
Message:From-Name	FILE – Creator (DC)	Absender Name oder Adresse
Message:From-Email	FILE – Creator (DC)	Absender Adresse
Message-To	FILE - Contributor (DC)	Adressat
Message-Recipient-Address	FILE - Contributor (DC)	Adressat Adresse
Message-Cc	FILE - Contributor (DC)	Adressat CC
Message-Bcc	FILE - Contributor (DC)	Adressat BCC
dc:title	FILE - Description (DC) ¹²⁵	E-Mail Subjekt
dcterms:created	FILE - Date (DC)	Erstellungs-/Sendedatum
Message:Raw-Header:X-MS-Has-Attach	FILE - Description (DC)	Anhang
meta:mapi-message-class	FILE - Type (DC)	E-Mail-Klassifikation

Dabei ist festzuhalten, dass hier nur die Metadaten aufgeführt sind, welche direkt aus den E-Mails ausgelesen werden und auf der File-Ebene in Rosetta dargestellt werden. Auf der IE-Ebene, werden die Metadaten manuell eingegeben und sind entsprechend hier nicht aufgeführt. Die Ergänzung „FILE“ zu den standardisierten

¹²⁵ Wir haben uns bewusst dazu entschieden, den E-Mail-Betreff nicht als Titel-Feld in Rosetta anzeigen zu lassen. Dadurch, dass Rosetta den ursprünglichen Dateinamen bei der Definition eines Titel-Felds nicht mehr anzeigt und stattdessen die Angabe aus dem Dublin Core-Metadatenfeld übernimmt, würde die Information des Dateinamens so verloren gehen. Deshalb wird das Metadatenfeld dc:title in Rosetta nicht als Titel, sondern als Description eingespielt.

Dublin Core-Metadatenfelder verweist somit auf diese Ebene. Die File-Metadaten werden automatisch in die CSV-Datei eingespielt, in welcher bereits die IE-Metadaten erfasst worden sind. Über diese werden auch Kollektionen für die jeweiligen E-Mail-Postfächer angelegt.

6.3. Die Entscheidung über das Format

Wie bereits im vorherigen Kapitel erwähnt, ist es grundsätzlich möglich, die E-Mails aus dem Vault sowohl im EML als auch im MSG Format zu exportieren. In einem ersten Schritt erfolgte ein Export der E-Mails auf Grund der gängigen Empfehlung im EML-Format. Es zeigte sich jedoch schnell, dass in diesem Format und mit den von FDD genutzten Tools einige der für das HSA als wichtig bewertete Metadaten nicht ausgelesen werden können oder fehlen, die in den gleichen Dateien im MSG-Format exportiert, auswertbar bzw. vorhanden sind. Hinzu kommt, dass auch bei den Metadaten, welche in beiden Formaten ausgelesen werden konnten, im EML-Format vom HSA als wichtig eingestufte Informationen fehlen, wie zum Beispiel Angaben bei den Kontakten oder Zeiträume von Terminen. Es wurde also die Entscheidung getroffen, dass der Export und Transfer ins Hochschularchiv im MSG-Format stattfinden soll. Da es sich bei MSG jedoch um ein proprietäres und nur bedingt empfohlenes Archivierungsformat handelt, stand die Frage der Migration im Raum. Da es wenig sinnvoll erschien, E-Mails im MSG-Format zu exportieren, um danach das Risiko von Datenkorruption und -verlust bei einer Migration ins EML-Format einzugehen, wurde entschieden, die originalen MSG-Dateien zu archivieren, diese aber in das für die Langzeitarchivierung prädestinierte Format PDF zu konvertieren. Bei dieser Entscheidung wurde vor allem die Arbeit der EA-PDF Working Group in Betracht gezogen, die eine Spezifikation zur Archivierung von E-Mail im PDF-Format veröffentlicht hat.¹²⁶ Dabei werden keine EA-PDF-Pakete gebildet, wie sie von der Working Group beschrieben werden, sondern lediglich die Dateien ins PDF-Format konvertiert. Dabei werden die üblichen Normen und Standards für die digitale Langzeitarchivierung, die von FDD angewendet werden, eingehalten. PDF hat in diesem Bereich den grossen Vorteil, dass es ein Standard-Format für die Langzeitarchivierung ist und von fast allen Betriebssystemen gelesen werden kann. Die Working Group bezeichnet die Archivierung im PDF-Format als „*a straightforward, ubiquitous, and highly secure way to access and view archived message.*“¹²⁷ PDF bietet zudem den Vorteil, dass Risiken und Lücken, welche bei den gängigen E-Mail-Archivierungsformaten vorhanden sind, geschlossen werden. Wichtig ist dabei lediglich, dass die Metadaten, die in den klassischen E-Mail-Formaten vorhanden sind, nicht ignoriert werden. Bei einer Konvertierung nach PDF müssen diese Metadaten natürlich überführt und festgehalten werden. Der Punkt, dass meist mehr Erfahrungen mit der Archivierung in diesem Format gemacht

¹²⁶ EA-PDF Working Group: A specification for using PDF to package and represent email, Board of Trustees of the University of Illinois, Illinois 2021. Online: <<http://hdl.handle.net/2142/109251>>.

¹²⁷ Ebd., S. 7.

wurden, kann im speziellen Fall der ETH-Bibliothek ausser Acht gelassen werden, da regelmässig auch spezielle Datenformate von z.B. Forschungsdaten in Rosetta archiviert werden. Lediglich bei der Darstellung von breiten Tabellen kann es bei der Umwandlung in PDF auf Grund der Formatbeschränkung zu Umbrüchen und somit unschönen Darstellungsfehlern kommen.¹²⁸

Neben den Dateien in PDF hat sich das Hochschularchiv zusätzlich dazu entschieden, weiterhin alle MSG-Dateien sowie die PST-Datei aufzubewahren. Dadurch ist eine zusätzliche Sicherheit geschaffen, indem neben den in ein langzeitarchivtaugliches Format überführten Daten auch die Originaldatei aufbewahrt wird.¹²⁹ Es ist nicht davon auszugehen, dass das PST- bzw. das MSG-Format bald abgelöst wird und es ist weit genug verbreitet, dass für diese Fälle schnell eine Exit-Strategie erarbeitet werden könnte, zumal die Spezifikation sowohl für das PST- als auch das MSG-Format von Microsoft offengelegt wurde.¹³⁰

Um die Dateien möglichst effizient in das gewünschte PDF-Format zu überführen, muss mit einem PDF-Konverter gearbeitet werden, der die Dateien im Batch konvertieren kann. Von der ETH-Bibliothek wird dazu der 3-Heights Document Converter¹³¹ genutzt. Zur Genauigkeit der Konvertierung mit diesem Tool gibt es sowohl von der ETH-Bibliothek selbst¹³² sowie der KOST¹³³ Studien, welche es als gut einstufen. In der Studie von Mohamed El-Saad und Roland Suri zeigte sich, dass bei der Konvertierung von E-Mails durch 3-Heights teilweise Hyperlinks verloren gingen. Ansonsten konnte auch hier ein gutes Ergebnis bei der Konvertierung erzielt werden. Hinzu kommt, dass wenn der Zeitpunkt der Resultate der Studie der ETH-Bibliothek mit der Studie der KOST verglichen werden, die Resultate über die Zeit besser geworden sind, was für eine Verbesserung des Tools mit den Jahren

¹²⁸ Ebd., S. 4–8; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): Requirements for Archiving Email using PDF (Draft for review and comment), Washington, DC 2020, S. 6. Online: <<https://docs.google.com/document/d/1JNij3M2UK2XUdsnRg4j9j33XPakj5-JSvijwpqIwbRo/edit#heading=h.2u0gt39hh5d>>.

¹²⁹ In diesem Prozess wurde bewusst darauf verzichtet, mit Unterscheidungen zwischen „Preservation Master“, „Modified Master“ und „Original Master“ zu arbeiten. Dies liegt einerseits daran, weil diese Unterscheidung für die Darstellung in Rosetta vor allem bei der späteren Nutzbarkeit der Daten wichtig wird und diese Daten erst nach Ablauf der Schutzfrist für Benutzende einsehbar werden. Andererseits ist eine Strukturierung nach Dateiformaten auch ohne diese Unterscheidung möglich und bietet weniger Verwechslungspotenzial und eine einfachere Übersicht, durch klar benannte Ordnerstrukturen.

¹³⁰ Microsoft: [MS-PST]: Outlook Personal Folders (.pst) File Format, Format-Spezifikation, 2022, <https://docs.microsoft.com/en-us/openspecs/office_file_formats/ms-pst/141923d5-15ab-4ef1-a524-6dce75aae546>, Stand: 12.07.2022; Microsoft: [MS-OXMSG]: Outlook Item (.msg) File Format, Format-Spezifikation, 2022, <https://docs.microsoft.com/en-us/openspecs/exchange_server_protocols/ms-oxmsg/b046868c-9fbf-41ae-9ffb-8de2bd4eec82>, Stand: 12.07.2022.

¹³¹ <https://www.pdf-tools.com/de/produkte/document-converter/>

¹³² El-Saad, Mohamed; Suri, Roland E.: Lost in migration: document quality for batch conversion to PDF/A, in: Library Hi Tech 39(2), 2021, S. 337–351. Online: <<https://doi.org/10.3929/ETHZ-B-000269199>>.

¹³³ Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST): PDF/A: Produktreview batchtauglicher PDF/A-Konverter, Version 1.1, 2016. Online: <https://kost-ceco.ch/cms/dl/d09b7330dbd2188e3e3b02f96389e21e/konvert2pdfa_2016_v1.1_DE.pdf?target=1>.

spricht.¹³⁴ In Tests des FDD im Hinblick auf die E-Mail-Archivierung anhand eines Test-Batches mit 3-Heights konnten nur in 0.2% der Fälle E-Mails nicht konvertiert werden. Dadurch, dass neben der konvertierten PDF-Datei auch die ursprüngliche MSG-Datei langzeitarchiviert wird und die E-Mails im Zweifelsfall auch noch in der PST-Datei aufzufinden sein müssten ist dieses Ergebnis für unsere Zwecke ausreichend. Die Konvertierung mit 3-Heights eines Test-Batches von 4'963 E-Mails dauerte insgesamt 6.11 Stunden, was 13.38 konvertierten E-Mails pro Minute entspricht.

Ein grosser Vorteil ist, dass bei 3-Heights bei der Konvertierung nicht nur Metadaten ausgelesen und in der finalen PDF-Datei festgehalten werden können. Ebenfalls ist es wichtig anzumerken, dass E-Mail-Anhänge, welche in PDF konvertierbar sind, bei 3-Heights in diesem Fall direkt an das PDF der entsprechenden E-Mail angehängt werden. So kann der direkte Bezug der beiden Dateien gewährleistet werden und die Metadaten aus der ursprünglichen Datei bleiben erhalten.¹³⁵

Damit arbeitet das Hochschularchiv im Grundsatz sehr ähnlich wie das Archiv der sozialen Demokratie (AdsD) der Friedrich-Ebert-Stiftung, dessen Arbeitsablauf Mike Zuchet in seinem Artikel „Pilotprojekt zur Langzeitarchivierung digitaler E-Mail-Korrespondenzen des Bundesvorstandes der Vereinigten Dienstleistungsgewerkschaft ver.di“ beschreibt.¹³⁶

6.4. Die Frage von Anhängen und Linked Content

Ein wichtiger Punkt, der bereits vorgängig kurzzeitig erwähnt wurde, ist der Umgang mit Anhängen und Linked Content, wenn es um die E-Mail-Archivierung geht. Es kann, wenn man von E-Mail-Archivierung spricht, nicht nur das eigentliche E-Mail gemeint sind, sondern auch Anhänge oder verlinkte Inhalte.¹³⁷ Dabei stellen sich unterschiedliche Herausforderungen, wenn man diese beiden Aspekte bei der Langzeitarchivierung berücksichtigen will:

Auch bei der Archivierung von E-Mail-Anhängen stellen sich die gleichen Fragen, wie bei der E-Mail-Archivierung allgemein: Man muss das Format, die auszulesenden Metadaten und nicht zuletzt die Verbindung zu dem E-Mail, zu welchem der Anhang gehört, beachten. Dabei können E-Mail-Anhänge weitaus mehr Formate vorweisen,

¹³⁴ El-Saad; Suri: Lost in migration, 2021, S. 344–346.

¹³⁵ Whitepaper: 3-Heights™ Document Converter. Grundlagen und Anwendung, 2016. Online: <<https://www.pdf-tools.com/public/downloads/whitepapers/Whitepaper-DocumentConverter-DE.pdf>>.

¹³⁶ Zuchet: Pilotprojekt zur Langzeitarchivierung digitaler E-Mail-Korrespondenzen des Bundesvorstandes der Vereinten Dienstleistungsgewerkschaft ver.di., 2012.

¹³⁷ Artefactual Systems and the Digital Preservation Coalition: Preserving Email, 2021, S. 2; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 11.

als es die E-Mails können und diese Formate in ein archivtaugliches Format zu überführen kann entsprechend aufwändig sein.¹³⁸

In der Fachliteratur findet sich keine allgemeingültige Lösung, wie mit E-Mail-Anhängen umgegangen werden soll. Im Fall des Projekts „E-Mail-Archiv 2.0“ wurde wieder auf die Konvertierung mit 3-Heights zurückgegriffen. In 80.7% der Konvertierungen aus einem Test-Batch konnte der E-Mail-Anhang ebenfalls in das PDF-Format konvertiert werden und wird somit zusammen mit dem E-Mail in diesem Format archiviert. In den Fällen, in welchen keine Konvertierung stattfinden konnte, handelte es sich um Dateiformate, welche nicht in PDF dargestellt werden können. Auf Grund der grossen Menge der zu übernehmenden E-Mail-Postfächer wurde entschieden, dass diese ca. 20% der Anhänge nicht gefiltert und konvertiert werden, sondern im originalen Format als Anhang an die MSG-Datei archiviert wird. Dabei handelt es sich um eine Ressourcenfrage, da es nicht möglich ist, alle Anhänge mit anderen Formaten zu extrahieren und in ein entsprechendes archivwürdiges Format zu konvertieren. Hinzu kommt, dass davon auszugehen ist, dass es auch Daten in Formaten gibt, auf die gar nicht mehr zugegriffen werden kann, weil es sich im Hochschulkontext durchaus auch um Dateien handeln kann, die aus spezifischen (Forschungs-)Applikationen kommen und entsprechend ohne die entsprechende Applikation gar nicht lesbar sind. Getestet wurde dies an Dateien, die aus dem Archivierungssystem CMI entstanden sind und nur gelesen werden können, wenn CMI auf dem eigenen Gerät installiert ist und ein Zugang besteht. Entsprechend war es auch nicht möglich, diese Dateien in ein Format zu konvertieren, das allgemein lesbar oder langzeitarchivierbar ist. Die von 3-Heights generierte Fehlermeldung, dass die entsprechenden E-Mail-Anhänge nicht konvertiert werden konnten, werden auf Grund der Nachvollziehbarkeit dokumentiert bzw. die Logdateien zusammen mit den konvertierten PDF-Dateien archiviert.

Wenn es um verlinkte Inhalte geht, vertritt die Fachwelt einen deutlicheren Standpunkt: Es ist fast unmöglich, alle verlinkten Inhalte zu archivieren. Dies hat seinen Hintergrund in der analogen Archivierung, in der auch nicht jedes erwähnte Dokument gesucht und mitarchiviert wird. Das heisst nicht, dass die verlinkten Inhalte nicht relevant sind oder gegebenenfalls sogar keine Berechtigung hätten, archiviert zu werden. Es dürfte jedoch logisch erscheinen, dass eine solche Archivierung auch nur zu einem semi-aktuellen Zeitpunkt stattfinden kann, da sich der Inhalt von verlinkten Dokumenten oder Websites sehr schnell ändern kann. Die Tools, welche es bisher zur E-Mail-Archivierung gibt, bieten noch keine Möglichkeit, solche Inhalte ebenfalls abzubilden.¹³⁹

¹³⁸ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives*, 2018, S. 32, 43, 49–50; Prom: *Preserving Email* (2nd Edition), 2019, S. 23; Knobloch: *Überlegungen zur Übernahme*, 2016, S. 227.

¹³⁹ Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives*, 2018, S. 33, 51; Prom: *Preserving Email* (2nd Edition), 2019, S. 24–25.

Auch im Falle des Projekts der ETH muss darauf verzichtet werden, verlinkte Inhalte zu archivieren. Einer der Hauptgründe neben dem technischen und zeitlichen Aufwand ist die Zeitspanne, in welcher Archivierte wird. Da es sich dabei um Dateien aus einem Zwischenarchiv, dem Vault, handelt und die E-Mails bis zum Ende der 1990er Jahre zurückgehen ist es schlichtweg unmöglich nachzuvollziehen, welche Inhalte damals abgebildet wurden – sofern die Website noch existiert. Auch stellt sich die Frage des Urheberrechts, sollten Websites oder Inhalte archiviert werden, welche nichts mit der ETH Zürich zu tun haben und lediglich in einer E-Mail für den Kontext verlinkt wurden. Einzig in den Fällen, in denen ETH-Websites in E-Mails nach 2018 verlinkt werden besteht eine gute Chance, dass dieser Inhalt wiedergegeben werden kann. Dabei wird aber auch der Inhalt nicht zusätzlich zu den E-Mails archiviert, sondern es besteht die Möglichkeit, über das Webarchiv der ETH Zürich¹⁴⁰, auf diese Inhalte zuzugreifen.

6.5. Die Erschliessung in CMI AIS

In der digitalen Langzeitarchivierung im Hochschularchiv der ETH Zürich wird auf zwei Ebenen gearbeitet, einerseits im Langzeitarchivierungstool Rosetta, wo die Daten gespeichert und langzeitarchiviert werden, andererseits im CMI Client, in welchem die dazugehörigen Metadaten verzeichnet und über den Webclient im Internet publiziert und zugänglich gemacht werden. Die Ordnung in diesen beiden Systemen soll grundsätzlich so ähnlich wie möglich bleiben, damit Verbindungen einfach geschaffen werden können, während die Funktionen der einzelnen Tools sowie die Anforderungen an diese sehr unterschiedlich sind. Die Erschliessung für das Projekt „E-Mail Archiv 2.0“ muss entsprechend von beiden Seiten betrachtet und unterschiedlich angegangen werden.

In einem ersten Schritt soll auf die Erschliessung im Archivsystem von CMI eingegangen werden. Wie bereits in Kapitel 5 beschrieben, ist bei der Erschliessung von grossen Datenmengen wie E-Mail-Archiven ein Verfahren nötig, bei dem viele Daten in wenig Zeit bearbeitet und nutzbar gemacht werden können. So beschreiben auch Meissner und Greene, die Begründer des MPLP Ansatz, diesen kurz mit dem Ziel: *„Get the most material available in a usable form in the briefest time possible.“*¹⁴¹ Dabei handelt es sich nicht um eine spezifische Erschliessungsmethode sondern um Guidelines, welche dabei helfen sollen, Backlogs aufarbeiten zu können und grosse Bestände, wie in diesem Fall mehrere E-Mail-Konten, so zu verarbeiten, dass sie Benutzenden zugänglich gemacht werden können. Dabei steht mehrheitlich

¹⁴⁰ Hochschularchiv der ETH Zürich: Webarchiv der ETH Zürich, Informationswebsite, <<https://library.ethz.ch/publizieren-und-archivieren/archivieren/hochschularchiv-der-eth-zuerich/webarchiv-der-eth-zuerich.html>>.

¹⁴¹ Meissner, Dennis; Greene, Mark A.: More Application while Less Appreciation: The Adopters and Antagonists of MPLP, in: Journal of Archival Organization 8 (3–4), 2010, S. 175. Online: <<https://doi.org/10.1080/15332748.2010.554069>>.

die Beschreibung der höheren Ebenen, wie Bestand und Serie im Vordergrund und nicht der Versuch, jedes einzelne Item zu verzeichnen. Während es bei gewissen Beständen sinnvoll sein kann, auf Einzelstück-Ebene zu erschliessen, so bringt es nach Meissner und Greene auch sehr viele Nachteile mit sich, da dies nicht nur sehr hohe Arbeitskosten mit sich bringt, sondern auch „*risks bending the inherent logic of archival description to the breaking point*“.¹⁴² Für Recherchezwecke sind die Beschreibungen auf Verzeichnungseinheitsstufe fast zu vernachlässigen. Davon abgesehen, dass dies bei E-Mail-Konten auf Grund der Grösse schlichtweg nicht möglich. Die Durchsuchbarkeit der digitalen Dokumente macht solch eine Erschliessung auch mehrheitlich überflüssig.¹⁴³

Im Endeffekt wird bei MPLP auf die Grundsätze vom Macro-appraisal zurückgegriffen und diese auf die Erschliessung angewandt. Ein Ansatz, den auch das Archives Nationales verwendet. Martine Sin Blima-Barru beschreibt, wenn es um E-Mail-Archivierung geht, diesen Ansatz als ideal, wobei die Problematik, ineinander verschachtelte E-Mails zu beschreiben aufgeworfen wird. Dabei ist das Ziel, die goldene Mitte zwischen wenig Aufwand und viel Ertrag zu finden.¹⁴⁴

Das Ergebnis einer Umfrage von Rachel Anchor, die sie unter Forschenden zum MPLP-Ansatz getätigt hat, zeigt, dass viele der Befragten gegenüber grundlegenden Informationen eines Bestandes genaue Detailangaben auf Item-Ebene bevorzugten. Der übergeordnete Zweck der Erschliessung der E-Mail-Konten soll es sein, dass diese genutzt werden können. Dies gilt für die Benutzung innerhalb des Archivs für Recherchezwecke sowie von Forschenden. Die Benutzung von Aussenstehenden wird wohl noch einige Jahre auf sich warten lassen. Es muss davon ausgegangen werden, dass sich innerhalb der E-Mail-Konten besonders schützenswerte Personendaten befinden. Diese unterliegen nach dem Bundesgesetz über die Archivierung einer 50-jährigen Schutzfrist. Da die ältesten E-Mails aus dem Vault Ende der 1990er-Jahre entstanden sind, werden die entsprechenden E-Mails noch einige Jahre nicht einsehbar sein. Bei diesem Bestand handelt es sich also um eine Ausnahme von Meissner und Greenes These, dass auch unbearbeitete Bestände für Recherchezwecke freigegeben werden sollten. Trotzdem sollte versucht werden, diese Daten zu gliedern und zu bearbeiten, anstatt sie über Jahre unbearbeitet liegen zu lassen. Die Auffindbarkeit und Durchsuchbarkeit sollte sowohl für den internen Gebrauch als auch für gewährte Einsichtsgesuche gewährleistet werden. Ähnlich der Beschreibung auf Schachtel-Ebene, von der Meissner und Greene sowie Anchor sprechen, wählen wir bei der Erschliessung der E-Mails auch einen eher groben

¹⁴² Ebd., S. 200.

¹⁴³ Ebd., S. 175–200; Greene; Meissner: *More Product, Less Process*, 2005, S. 208–212, 253; Anchor, Rachel: 'More product, less process': method, madness or practice?, in: *Archives and Records* 34 (2), 2013, S. 156–166. Online: <<https://doi.org/10.1080/23257962.2013.818937>>.

¹⁴⁴ Anchor: *MPLP: method, madness or practice?*, 2013, S. 165; Meissner; Greene: *More Application while Less Appreciation*, 2010, S. 216; Sin Blima-Barru: *Archives nationales – Archiver les courriers électroniques*, 2019, S. 1.

Ansatz, der von Rachel Anchor als „*pragmatic compromise in dealing with the bulk of twentieth-century collections*“¹⁴⁵ bezeichnet wird.¹⁴⁶

Dies führt konkret zu einer Erschliessung, die auf der Bestandsebene das E-Mail-Archiv beschreibt und darunter nach Serien geordnet die einzelnen Funktionsträger aufführt. Diese Serien wiederum umfassen Jahresdossiers, die kaum weitere Informationen enthalten, aber es ermöglichen die Schutzfristen genauer zu setzen und die Durchsuchbarkeit auf Jahresebene zu ermöglichen. Diese Erschliessung ist verhältnismässig wenig aufwändig und doch klar genug strukturiert, dass man sich darin zurechtfindet. Dies entspricht für die Nutzbarkeit dem, was Meissner und Greene beschreiben, wenn sie von den Aufgaben eines Referenz-Archivaren schreiben:

If a reference archivist can point a researcher to a box (or a handful of boxes) and report that “the bulk of the John Doe correspondence in this collection will be found in these boxes,” most researchers will be pleased—particularly if the alternative is to have even that level of information hidden because the collection resides in a backlog.¹⁴⁷

6.6. Der Ingest in Rosetta

Eine der ersten Fragen, die sich vor dem Ingest in Rosetta stellte, war, wie viele E-Mails eine Intellectual Entity (IE)¹⁴⁸ bilden sollen. Die Optionen und dazugehörige Vor- und Nachteile haben sowohl die Task Force on Technical Approaches to Email Archives sowie Corinna Knobloch in ihren *Überlegungen zur Übernahme und Archivierung von E-Mail-Konten* genau dargelegt: Eine IE kann entweder einer E-Mail, einem Ordner oder einem ganzen E-Mail-Konto entsprechen. Die Option, eine E-Mail als eine IE zu erschliessen bringt mehr Nachteile denn Vorteile mit sich, da es unübersichtlich und unnötig zeitaufwändig wird, weshalb diese Option gar nicht erst in Betracht gezogen wurde, als es um den Ingest in Rosetta ging. Die Option, einen Ordner zu einer IE zu machen ist etwas übersichtlicher, da beim Export aus dem Vault jedoch die Ordnerstruktur nur innerhalb der pst-Datei ersichtlich ist, war auch dies keine valide Lösung. Die letzte Option, in welcher jedes E-Mail-Konto zu einer eigenen IE wird, bringt viele Vorteile mit sich, unter anderem ist sie mit weniger Aufwand verbunden und die originale Ordnung im einzelnen Konto wird ersichtlich,

¹⁴⁵ Anchor: MPLP: method, madness or practice?, 2013, S. 156.

¹⁴⁶ Greene; Meissner: More Product, Less Process, 2005, S. 252; Meissner; Greene: More Application while Less Appreciation, 2010, S. 180–204; Anchor: MPLP: method, madness or practice?, 2013, S. 157.

¹⁴⁷ Meissner; Greene: More Application while Less Appreciation, 2010, S. 178–200; Anchor: MPLP: method, madness or practice?, 2013, S. 156.

¹⁴⁸ Definition aus: The Library of Congress: PREMIS Data Dictionary for Preservation Metadata, Version 3.0, 2015: An Intellectual Entity is a distinct intellectual or artistic creation that is considered relevant to a designated community in the context of digital preservation: for example, a particular book, map, photograph, database, or hardware or software. An Intellectual Entity can include other Intellectual Entities; for example, a web site can include a web page and a web page can include an image. An Intellectual Entity may have one or more digital or non-digital Representations.

sobald man auf dieses zugreifen kann. Im Hinblick auf die zukünftige Nutzung und die langen Schutzfristen haben wir uns dazu entschieden, die E-Mail-Konten zusätzlich in Jahresbatches in IEs umzuwandeln. Das komplette E-Mail-Postfach einer Person (bzw. das Postfach aus deren relevanten Amtsjahren) wird somit in mehreren IEs in Rosetta importiert. Damit wird zwar die Durchsuchbarkeit etwas eingeschränkt, weil im Zweifelsfall mehrere IEs angeschaut werden müssen, es ist aber besonders bei Personen mit einer langen Amtszeit möglich, einen Teil der Informationen bereits der Öffentlichkeit zugänglich zu machen, weil die Schutzfrist dieser IE bereits abgelaufen ist. Zudem wirken wir damit Problemen wie Unübersichtlichkeit in sehr grossen IEs entgegen.¹⁴⁹

Nachdem entschieden wurde, welche Metadaten aus den E-Mails exportiert werden können, galt es festzulegen, welche davon in Rosetta Eingang finden sollen. Jenny Bunn weist auf die Möglichkeiten hin, die sich durch die automatische Erkennung und Extraktion von Metadaten bieten: Die gängigen Deskriptoren von Dokumenten wie z.B. Titel, Dateiformat, Entstehungszeitraum sind bereits als solche definiert und aufbereitet und müssen nur noch ausgelesen und in das entsprechende Feld eingefügt werden. Gleichzeitig beschreibt Jane Zhang in ihrem Artikel *Original Order in Digital Archives* die klare Dualität zwischen der Beschreibung auf den hohen Ebenen, wie Serie und Bestand, und den tieferen Ebenen von Dossier und Einzelstück bzw. IE. Während auf den tieferen Ebenen die Metadaten häufig vorhanden sind, ist es auf den höheren Ebenen, in welchen Beziehungen und Hierarchien gebildet werden müssen, immer noch nötig, dass Archivarinnen und Archivare diese Aufgabe manuell ausführen. Dabei ist es wichtig, dass über die automatisierte Auslese der vorhandenen Metadaten nicht der Kontext vergessen geht und die Ordnung so nicht mehr nachzuvollziehen ist.¹⁵⁰

Die Aufgaben der Archivare und Archivarinnen haben sich also von der aufwändigen Erschliessung und damit verbundenen Recherche nach Details zu den Archivalien deutlich verschoben. Aber bereits 1993 schrieb Margaret Hedstrom zu diesem Thema:

In the electronic era, the descriptive paradigm will shift from the current practice of augmenting scarce descriptive information to one of selecting from an abundance of metadata, which could form a complete audit trail of all actions taken to create, update, and modify a record, and of all its uses. Automated systems have the capacity to capture and record far more descriptive information than was technically possible or economically feasible with manual systems. [...] As descriptive practices shift from creating descriptive information to capturing

¹⁴⁹ Knobloch: Überlegungen zur Übernahme, 2016, S. 222–224; Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): The future of email archives, 2018, S. 20.

¹⁵⁰ Zhang: Original Order in Digital Archives, 2012, S. 190–192; Bunn: Born digital archive cataloguing and description, 2021, S. 1–4.

description along with the records, archivists may discover that managing the metadata is a much greater challenge than managing the records themselves.¹⁵¹

Dies führt nach Bunn aber auch dazu, dass im heutigen digitalen Kontext, immer mehr Information benötigt wird, damit der Standard der Erschliessung und Deskription, wie man ihn neuerdings gewohnt ist, erreicht wird. Es gilt klare Standards einzuhalten. So basiert auch Rosetta auf den Dublin Core Metadaten, die wiederum innerhalb der ETH-Bibliothek klar mit Feldern aus anderen Fachapplikationen, im Fall des Hochschularchivs CMI AIS, verbunden sind, um eine logische Struktur zu erhalten.¹⁵²

Diese Unterscheidung zwischen der manuellen Erschliessung und der automatischen Extraktion von Metadaten zeigt sich auch in der Erschliessung im Projekt „E-Mail Archiv 2.0“, indem in Rosetta viele Metadaten automatisch ausgelesen werden. Welche deskriptiven Metadaten vorgängig durch Tika exportiert werden sollen und in welchen Feldern in Rosetta diese dargestellt werden, wurde bereits in Kapitel 6.2. festgehalten.

Innerhalb einer IE befinden sich also mehrere Files, die jeweils mit den nötigen Metadaten ausgestattet sind und somit auch durchsuchbar sind. Die Struktur dieser IEs wird semi-automatisch mit dem Ingest durch eine CSV-Datei geregelt, in welcher die hierarchische Struktur wie sie im Archiv üblich ist übernommen wird. Anstatt der üblichen Bestände aus CMI AIS kann in Rosetta diese Hierarchie durch Kollektionen abgebildet werden. Diese kommen hauptsächlich ohne zusätzliche Beschreibungen aus und sind vor allem durch klare Titel erkennbar. Die Struktur der einzelnen Postfächer bleibt innerhalb der pst-Dateien erhalten und ist so recherchierbar. Ebenfalls in der jeweiligen IE enthalten ist ein Report, der automatisch beim Export aus dem Vault erstellt wird. In dieser TXT-Datei ist unter anderem ersichtlich, um welches E-Mail-Konto es sich handelt, wann der Export stattgefunden hat und welchen Zeitraum der Export umfasst. Diese Datei wird zur weiteren Dokumentation zusammen mit den übrigen Daten archiviert.

7. Fazit und weiteres Vorgehen

Die Archivierung von E-Mail ist in der Archivwelt, wie bei vielen anderen born digital Unterlagen auch, noch ein verhältnismässig neues Thema. Es gibt zwar bereits sehr viele Arbeiten und Publikationen dazu, im Endeffekt ist der Umgang mit Big Data immer noch eine Neuheit und es konnte noch nicht die allgemeingültig beste Lösung für die digitale Langzeitarchivierung gefunden werden. Die Taskforce on Technical Approaches to Email Archives sowie die Digital Preservation Coalition sind sehr gute Anlaufstellen, wenn man sich mit der Archivierung von E-Mails beschäftigen will. Die Ansätze zur Bewertung sind mit dem Capstone Approach als Alternative für die

¹⁵¹ Hedstrom: Descriptive Practices for Electronic Records: Deciding What Is Essential and Imagining What Is Possible, 1993, S. 59.

¹⁵² Bunn: Born digital archive cataloging and description, 2021, S. 2–4.

klassische Archivierung von digitalen Unterlagen aus einem Records Management System weit verbreitet, sofern nicht Tools eingesetzt werden, um die Bewertung zu erleichtern. Dabei zeigt sich jedoch der Nachteil davon, dass die Forschung der E-Mail-Archivierung im angelsächsischen Raum wesentlich weiter ist, als im Deutschsprachigen. Tools, welche automatisch Entitäten erkennen, sind nur in der jeweiligen Sprachregion einsetzbar und entsprechend weniger geeignet für ein Projekt an der ETH Zürich. Auch die Entscheidung über das Format der E-Mails wurde mehrheitlich von den bestehenden Erfahrungen und Applikationen beeinflusst, da bisher bereits gute Erfahrungen mit dem 3-Heights Document Converter gemacht wurden.

Während in der Forschung darüber diskutiert wird, ob die Archivierung kompletter E-Mail-Konten ein guter Weg ist, so ist die Empfehlung betreffend die Erschliessung von E-Mails wesentlich einheitlicher. Durch die grosse Menge von E-Mails eignet sich nur ein ressourcensparender Ansatz und die MPLP-Strategie von Meissner und Greene ist dafür ideal. Die einzelnen Punkte, wie das Mapping von Rosetta und CMI sowie das genaue Vorgehen konnten auf Grund dieser Ausgangslage verhältnismässig einfach bestimmt werden.

Wenn es darum geht, ein Projekt, wie das "E-Mail-Archiv 2.0" umzusetzen, hat es viele Vorteile, wenn man in einer grossen Institution wie es die ETH Zürich ist, angesiedelt ist. Damit ein Projekt in diesem Umfang umgesetzt werden kann, muss eine gewisse Infrastruktur bereits bestehen und personelle wie auch technische Ressourcen dafür bereitstehen. Während es auch hinderlich sein kann, wenn mit den Applikationen gearbeitet werden muss, welche in der Institution vorhanden sind, kann man dafür davon ausgehen, dass die Mitarbeitenden sich entsprechend gut damit auskennen und die Funktionalitäten ideal nutzen können. Hinzu kommt, dass gewisse Entscheidungen durch die gegebene Infrastruktur abgenommen werden, wie dass im Falle des „E-Mail-Archiv 2.0“ keine Archivierung aus einem RMS stattfinden konnte, weil die ETH Zürich bisher nicht mit einem solchen System arbeitet. Dafür sind wir in der privilegierten Situation, E-Mail-Konten in einem Zwischenspeicher zu haben, die bis in die Späten 1990er Jahre hinein reichen und entsprechend interessante Quellen bieten können, wenn ihre Schutzfrist abgelaufen ist.

Es gibt jedoch auch Punkte, welche in dieser Arbeit keinen Eingang gefunden haben. Dazu gehört unter anderem der Umgang mit verschlüsselten E-Mails. Diese kommen unserer Einschätzung nach in den retrospektiv bewerteten E-Mails kaum vor. Sollte dies trotzdem der Fall sein, sind sie leider für die Archivierung verloren, da es nicht möglich ist, diese Jahre später noch zu entschlüsseln. Ein genauerer Blick wird in diesem Punkt auf die zukünftig zu übernehmenden E-Mail-Konten gerichtet, aus welchen E-Mails entschlüsselt werden müssten, bevor sie ins Langzeitarchiv gelangen. Wenn der Schlüssel an der ETH Zürich generiert wurde, sollte dies möglich sein, wenn aber der Schlüssel einer anderen Person, ausserhalb der ETH Zürich, für die Entschlüsselung nötig sein würde, ist dies nicht möglich. Auch der Punkt der Archivierung von allfälligen Spam-E-Mails wurde in dieser Arbeit nicht

behandelt. Grund dafür ist hauptsächlich, dass nur E-Mails, welche 30 bzw. neuerdings 60 Tage im E-Mail-Postfach verbleiben in den Vault kommen und davon auszugehen ist, dass Spam-E-Mails von den Mitarbeitenden der ETH Zürich direkt gelöscht werden. Diese E-Mails würden also niemals den Eingang ins Vault oder ins Langzeitarchiv finden.

Es ist zudem ein Bestreben des Hochschularchivs, dass in den nächsten Jahren ein Records Management an der ETH Zürich etabliert wird und dann wird die Schulung der Mitarbeitenden auf den korrekten Umgang und die Ablage von E-Mails noch einmal sehr wichtig werden. In diesem Zusammenhang muss dann noch einmal evaluiert werden, ob wie jetzt angedacht weiterhin ganze E-Mail-Konten übernommen werden. Diese können, wie im Kapitel 3.1 beschrieben als gute Zusatzüberlieferung zu einer Archivierung aus dem RMS dienen. Auch sollte man sich in Zukunft die Frage stellen, ob gesellschaftsaktuelle Themen, ähnlich den Hot-Spot-Listen angefertigt werden sollen, damit zum Beispiel Funktions-Accounts zu Themen, welche die Hochschulgeschichte betreffen, ebenfalls archiviert werden können. Dies ist aber nicht mehr Teil des Projekts und eine Frage, die erst nach Übernahme der E-Mail-Accounts aus dem Vault aktuell werden wird.

8. Bibliographie

- Anchor, Rachel: 'More product, less process': method, madness or practice?, in: Archives and Records 34 (2), 2013, S. 156–174. Online: <<https://doi.org/10.1080/23257962.2013.818937>>.
- Artefactual Systems and the Digital Preservation Coalition: Preserving Email, DPC Technology Watch Report, Digital Preservation Coalition, 2021. Online: <<https://doi.org/10.7207/twgn21-08>>.
- Bearman, David: Managing electronic mail, in: Archives & Manuscripts 22 (1), 1994. Online: <<https://publications.archivists.org.au/index.php/asa/article/view/8385>>.
- Belovari, Susanne: Expedited Digital Appraisal for Regular Archivists: An MPLP-Type Appraisal Workflow for Hybrid Collections, in: Journal of Archival Organization 16 (4), 2019, S. 197–219. Online: <<https://doi.org/10.1080/15332748.2019.1682793>>.
- Benauer, Maria: E-Mails, ihr Wert und ihre Bewertung, in: VÖA - Verband Österreichischer Archivarinnen und Archivare (Hg.): Scrinium 74, 2020, S. 87–115.
- Bibliothèque nationale de France: Formats de données pour la préservation à long terme : la politique de la BnF. Version initiale pour appel à commentaires, 2021. Online: <https://www.bnf.fr/sites/default/files/2021-04/politiqueFormatsDePreservationBNF_20210408.pdf>.
- Bissonnette, Natalie: Gestion des courriels : stratégies, technologies et bonnes pratiques, in: Archives 44 (1), 2012, S. 77–113.
- Boudrez, Filip; Van den Eynde, Sofie: DAVID: Archiving Email, Leuven - Antwerp 2002. Online: <<http://www.expertisecentrumdavid.be/davidproject/teksten/Rapporten/Report4.pdf>>.
- Brossi, Pietro; Winkler, Maria: E-Mail Archivierung: Leitfaden zur Umsetzung und rechtliche Grundlagen für die Schweiz, Deutschland und Österreich, Rheinfelden 2008.
- Brühwiler, Tom: Internet-Pionier Ray Tomlinson: Der Mann, der die E-Mail erfand, Cyon, 26.02.2021, <<https://www.cyon.ch/blog/Internet-Pionier-Ray-Tomlinson-Erfinder-E-Mail>>.
- Buchholz, Matthias: Archivische Überlieferungsbildung im Spiegel von Bewertungsdiskussion und Repräsentativität, Köln 2011 (Archivhefte / Landschaftsverband Rheinland, Archivberatungsstelle Rheinland 35).
- Bunn, Jenny: Born digital archive cataloguing and description, DPC Technology Watch Report, 2021. Online: <<https://doi.org/10.7207/twgn21-05>>.

- Burnel, Anne: Si le courriel est l'avenir de la lettre... Enjeux juridiques, technologiques et patrimoniaux de l'archivage du courrier électronique, in: Gazette des archives (214), 2009, S. 79–85.
- Capra, Robert; Khanova, Julia; Ramdeen, Sarah: Work and personal e-mail use by university employees: PIM practices across domain boundaries, in: Journal of the American Society for Information Science and Technology 64 (5), 2013, S. 1029–1044. Online: <<https://doi.org/10.1002/asi.22815>>.
- Cook, Terry: «Many Are Called, but Few Are Chosen»: Appraisal Guidelines for Sampling and Selecting Case Files, in: Archivaria 32, 1991. Online: <<https://archivaria.ca/index.php/archivaria/article/view/11759>>.
- Cox, Richard J: Appraisal and the future of archives in the digital era, in: Hill, Jennie (Hg.): The Future of Archives and Recordkeeping, 2010, S. 213–237. Online: <<https://doi.org/10.29085/9781856048675.012>>.
- EA-PDF Working Group: A specification for using PDF to package and represent email, Board of Trustees of the University of Illinois, Illinois 2021. Online: <<http://hdl.handle.net/2142/109251>>.
- Eberle, Armin: Die vorarchivische Selektion und archivische Bewertung von E-Mails. Masterarbeit im Rahmen des Masterstudiengangs Archiv- und Informationswissenschaft 2006-2008 an den Universitäten Bern und Lausanne, 2008.
- El-Saad, Mohamed; Suri, Roland E.: Lost in migration: document quality for batch conversion to PDF/A, in: Library Hi Tech 39(2), 2021, S. 337–351. Online: <<https://doi.org/10.3929/ETHZ-B-000269199>>.
- ETH Zürich: Standard Dienstleistungsvereinbarung (Service based SLA) für „Mail und Groupware“, 19.05.2022. Online: <<https://ethz.ch/content/dam/ethz/associates/services/Service/IT-Services/files/sla/sla-mail-groupware.pdf>>.
- ETH Zürich: Benutzungsordnung für Informations- und Kommunikationstechnologie an der ETH Zürich (BOT) und Anhang, 01.05.2022. Online: <<https://rechtssammlung.sp.ethz.ch/Dokumente/203.21.pdf>>.
- ETH Zürich: Standard Dienstleistungsvereinbarung (Service based SLA) für „Mail und Groupware“, 12.10.2021.
- ETH Zürich: Benutzungsordnung für Informations- und Kommunikationstechnologie an der ETH Zürich (BOT) und Anhang, 01.06.2021.
- ETH Zürich: Reglement für das Archiv der ETH Zürich, 01.04.2015. Online: <<https://rechtssammlung.sp.ethz.ch/Dokumente/420.1.pdf>>.
- Fanning, Betsy: Preservation with PDF/A (2nd Edition), DPC Technology Watch Report, Digital Preservation Coalition, 2017. Online: <<https://doi.org/10.7207/twr17-01>>.

- Funk, Stefan E.: Einführung Digitale Erhaltungsstrategien, in: Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.3., 2010, S. 160–161.
- Funk, Stefan E.: Migration, in: Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.3., 2010, S. 169–174.
- Greene, Mark; Meissner, Dennis: More Product, Less Process: Revamping Traditional Archival Processing, in: *The American Archivist* 68 (2), 2005, S. 208–263. Online: <<https://doi.org/10.17723/aarc.68.2.c741823776k65863>>.
- Gubler, Daniel: E-Mail-Archivierung, die brennenden Punkte, in: *Arbido: Records Management in Verwaltung und Privatwirtschaft – ein neues Aufgabenfeld?* (2010/2), 2010.
- Hedstrom, Margaret: Descriptive Practices for Electronic Records: Deciding What Is Essential and Imagining What Is Possible, in: *Archivaria* 36, 1993.
- Hochschularchiv der ETH Zürich: Webarchiv der ETH Zürich, Informationswebsite, <<https://library.ethz.ch/publizieren-und-archivieren/archivieren/hochschularchiv-der-eth-zuerich/webarchiv-der-eth-zuerich.html>>.
- Holdsworth, David: Strategies for digital preservation, in: Deegan, Marilyn; Tanner, Simon (Hg.): *Digital Preservation*, 2013, S. 32–59. Online: <<https://doi.org/10.29085/9781856049863.005>>.
- King's College London: InSPECT: Investigating the Significant Properties of Electronic Content Over Time, Projektwebseite, <<https://significantproperties.kdl.kcl.ac.uk/>>.
- Knobloch, Corinna: Überlegungen zur Übernahme und Archivierung von E-Mail-Konten, in: *Digitale Archivierung. Innovationen – Strategien – Netzwerke. Tagungsband zur 19. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ (Mitteilungen des Österreichischen Staatsarchivs 59/2016)*, Wien 2016, S. 221–231.
- Kolawole, Emi: Smithsonian acquires documents from inventor of 'EMAIL' program, in: *The Washington Post*, 17.02.2012. Online: <https://www.washingtonpost.com/national/on-innovations/va-shivaayyadurai-inventor-of-e-mail-honored-by-smithsonian/2012/02/17/gIQA8gQhKR_story.html>.
- Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST): Katalog archivischer Dateiformate (KaD), Version 6.2, 12.2021, <https://kost-ceco.ch/cms/kad_main_de.html>.
- Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST): PDF/A: Produktreview batchtauglicher PDF/A-Konverter, Version 1.1, 2016. Online: <https://kost-ceco.ch/cms/dl/d09b7330dbd2188e3e3b02f96389e21e/konvert2pdfa_2016_v1.1_DE.pdf?target=1>.

- Kretzschmar, Robert: Alles neu zu durchdenken? Archivische Bewertung im digitalen Zeitalter, in: Archivpflege in Westfalen und Lippe 80, 2014, S. 9–15.
- Kupper, Beda: E-Mail-Archivierung, in: Actualité archivistique suisse : travaux du certificat en archivistique et sciences de l'information = Archivwissenschaft Schweiz aktuell: Arbeiten aus dem Zertifikat in Archiv- und Informationswissenschaften, Baden 2008, S. 88–117.
- Landesarchiv Baden-Württemberg: PDF/A als Anbietungs- und Archivierungsformat. Richtlinie des Landesarchivs Baden-Württemberg, 05.06.2013. Online: <https://www.landesarchiv-bw.de/sixcms/media.php/120/PDF_richtlinie_mit_erlaeuterungen.pdf>.
- Lappin, James; Jackson, Tom; Matthews, Graham u. a.: The defensible deletion of government email, in: Records Management Journal 29 (1/2), 2019, S. 42–56. Online: <<https://doi.org/10.1108/RMJ-09-2018-0036>>.
- Library of Congress: Recommended Format Specifications 2021-2022, 2021. Online: <<https://www.loc.gov/preservation/resources/rfs/RFS%202021-2022.pdf>>.
- Maitland, Jo: Best Practices: Email Archiving, Cambridge, Cambridge 2008.
- Meissner, Dennis; Greene, Mark A.: More Application while Less Appreciation: The Adopters and Antagonists of MPLP, in: Journal of Archival Organization 8 (3–4), 2010, S. 174–226. Online: <<https://doi.org/10.1080/15332748.2010.554069>>.
- Menne-Haritz, Angelika: Schlüsselbegriffe der Archivterminologie: Lehrmaterialien für das Fach Archivwissenschaft, Marburg 2011 (Veröffentlichungen der Archivschule Marburg, Institut für Archivwissenschaft Nr. 20).
- Microsoft: [MS-PST]: Outlook Personal Folders (.pst) File Format, Format-Spezifikation, 2022, <https://docs.microsoft.com/en-us/openspecs/office_file_formats/ms-pst/141923d5-15ab-4ef1-a524-6dce75aae546>, Stand: 12.07.2022.
- Microsoft: [MS-OXMSG]: Outlook Item (.msg) File Format, Format-Spezifikation, 2022, <https://docs.microsoft.com/en-us/openspecs/exchange_server_protocols/ms-oxmsg/b046868c-9fbf-41ae-9ffb-8de2bd4eec82>, Stand: 12.07.2022.
- Morisod, Pascal: Données et métadonnées, aux sources de l'information numérique, in: Arbido: Metadaten – Datenqualität (2017/3), 2017.
- Munn, David: GLA Email Policy. Policy for managing email in the Greater London Authority, London 2004.
- Murray, Kate: Shaking the Email Format Family Tree, The Signal, 04.04.2014, <<https://blogs.loc.gov/thesignal/2014/04/shaking-the-email-format-family-tree/>>, Stand: 29.05.2022.
- Nationaal Archief Ministerie van Onderwijs, Cultuuren en Wetenschap: Model selectielijst voor automatisch bewaarde e-mail conform de Handreiking

- bewaren van e-mail Rijksoverheid, Den Haag 2020. Online: <https://www.informatiehuishouding.nl/binaries/informatiehuishouding/documenten/instrumenten/2020/05/18/model-selectielijst-voor-automatisch-bewaarde-e-mail-conform-de-handreiking-bewaren-van-e-mail-rijksoverheid/00828+Model+selectielijst+voor+automatisch+bewaarde+e-mail_V2.pdf>.
- Nationaal Archief Ministerie van Onderwijs, Cultuuren en Wetenschap: Hotspotlijst maken, <<https://www.nationaalarchief.nl/archiveren/kennisbank/hotspotlijst-maken>>, Stand: 14.05.2022.
- Nationaal Archief Ministerie van Onderwijs, Cultuuren en Wetenschap: Handreiking e-mail archiveren, <<https://www.nationaalarchief.nl/archiveren/kennisbank/handreiking-e-mail-archiveren#collapse-92082>>, Stand: 14.05.2022.
- Nationaal Archief Ministerie van Onderwijs, Cultuuren en Wetenschap: Selectiemethode bepalen. Informatieblad archiveren chatberichten, <<https://www.nationaalarchief.nl/archiveren/kennisbank/selectiemethode-bepalen>>, Stand: 14.05.2022.
- National Archives and Records Administrations (NARA): White Paper on The Capstone Approach and Capstone GRS, 2015. Online: <<https://www.archives.gov/files/records-mgmt/email-management/final-capstone-white-paper.pdf>>.
- National Archives and Records Administrations (NARA): Guidance on a New Approach to Managing Email Records, 2013-02, NARA Bulletin, 2013. Online: <<https://www.archives.gov/records-mgmt/bulletins/2013/2013-02.html>>.
- National Archives and Records Administrations (NARA): Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images, o. D. Online: <<https://www.archives.gov/files/preservation/technical/guidelines.pdf>>.
- Neumayer, Robert; Rauber, Andreas: Why Appraisal is not ‘Utterly’ Useless and Why It’s Not the Way to Go Either. A Provocative Position Paper (PPP), Wien 2007.
- Pennock, Maureen: Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages, DCC Digital Curation Manual, 2006. Online: <<https://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf>>.
- Periat, Nicole: Politique de gestion du courrier électronique: des mesures à prendre, in: Archives 29 (1), 1997, S. 3–56.
- Plante, Jeanette: “Challenges of Email as a Record Archiving Email Symposium.” Presentation at Archiving Email Symposium, Department of Justice, 2015.
- Prom, Christopher: Preserving Email (2nd Edition), DPC Technology Watch Report, Digital Preservation Coalition, 05.2019. Online: <<https://doi.org/10.7207/twr19-01>>.

- Prom, Christopher J.: Preserving Email, DPC Technology Watch Report, Digital Preservation Coalition, 2011. Online: <<https://doi.org/10.7207/twr11-01>>.
- Quiquerez, Fanny; Minana, Floriane: La gestion des données de la recherche: enjeux et perspectives, in: *Arbido: Metadaten – Datenqualität* (2017/3), 2017.
- Schneider, J.; Adams, C.; DeBauche, S. u. a.: Appraising, processing, and providing access to email in contemporary literary archives, in: *Archives and Manuscripts* 47 (3), 2019, S. 305–326. Online: <<https://doi.org/10.1080/01576895.2019.1622138>>.
- Schwarz, Karin: E-Mail-Archivierung, in: *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.3.*, 2010, S. 550–563.
- Schweizerisches Bundesarchiv: Standards für die Archivierung digitaler Unterlagen, 04.2020. Online: <https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte_und_weisungen/archivtaugliche_dateiformate.1.pdf.download.pdf/archivtaugliche_dateiformate.pdf>.
- Shepherd, Elizabeth; Yeo, Geoffrey: *Managing Records: a handbook of principles and practice*, 2013. Online: <<https://doi.org/10.29085/9781856049788>>, Stand: 14.05.2022.
- Sin Blima-Barru, Martine: Archives nationales – Archiver les courriers électroniques, Séminaire annuel des Archives de France (Dijon, 20-22 novembre 2019), Dijon 2019.
- Stanford Libraries: ePADD, Projektwebseite, <<https://library.stanford.edu/projects/epadd>>.
- Starkloff, Kristina: Überlieferung von E-Mail-Konten als genuin digitale Unterlagen. Archivwürdigkeit, Übernahmemethodik und Einblicke in die Entwicklung eines Werkzeugs, in: Puchta, Michael; Naumann, Kai (Hg.): *Kreative digitale Ablagen und die Archive: Ergebnisse eines Workshops des KLA-Ausschusses Digitale Archive am 22./23.11.2016 in der Generaldirektion der Staatlichen Archive Bayerns, München 2017* (Sonderveröffentlichungen der Staatlichen Archive Bayerns), S. 39–43.
- Sturm, Patrick: Die E-Mail: Ein Kommunikationsmedium des frühen 21. Jahrhunderts quellenkundlich betrachtet, in: Berwinkel, Holger; Kretzschmar, Robert; Uhde, Karsten u. a. (Hg.): *Moderne Aktenkunde*, Marburg 2016 (Veröffentlichungen der Archivschule Marburg, Hochschule für Archivwissenschaft), S. 109–129.
- Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *Requirements for Archiving Email using PDF* (Draft for review and comment), Washington, DC 2020. Online: <<https://docs.google.com/document/d/1JNij3M2UK2XUdsnRg4j9j33XPakj5-JSvijwpqlwbRo/edit#heading=h.2u0gt39hh5d>>.
- Task Force on Technical Approaches to Email Archives; Andrew W. Mellon Foundation; Digital Preservation Coalition (Hg.): *The future of email archives:*

a report from the Task Force on Technical Approaches for Email Archives, Washington, DC 2018 (CLIR publication 175).

The InterPARES 3 Project, Team Italy: General Study 05 – Keeping and Preserving E-mail, 2009. Online: <http://www.interpares.org/ip3/display_file.cfm?doc=ip3_italy_gs05a_final_report.pdf>.

The Library of Congress: PREMIS Data Dictionary for Preservation Metadata, Version 3.0, 2015. Online: <<https://www.loc.gov/standards/premis/>>.

Vellino, André; Alberts, Inge: Assisting the appraisal of e-mail records with automatic classification, in: *Records Management Journal* 26 (3), 2016, S. 293–313. Online: <<https://doi.org/10.1108/RMJ-02-2016-0006>>.

Vinh-Doyle, William P.: Appraising email (using digital forensics): techniques and challenges, in: *Archives and Manuscripts* 45 (1), 02.01.2017, S. 18–30. Online: <<https://doi.org/10.1080/01576895.2016.1270838>>.

Waugh, Dorothy; Russey Roke, Elizabeth; Farr, Erika: Flexible processing and diverse collections: a tiered approach to delivering born digital archives, in: *Archives and Records* 37 (1), 02.01.2016, S. 3–19. Online: <<https://doi.org/10.1080/23257962.2016.1139493>>.

Yeo, Geoffrey: Can we keep everything? The future of appraisal in a world of digital profusion, in: Brown, Caroline (Hg.): *Archival Futures*, London 2019, S. 45–63.

Zhang, Jane: Original Order in Digital Archives, in: *Archivaria* 74, 2012, S. 167–193.

Zuchet, Mike: Pilotprojekt zur Langzeitarchivierung digitaler E-Mail-Korrespondenzen des Bundesvorstandes der Vereinten Dienstleistungsgewerkschaft ver.di., in: Verein Deutscher Archivare: Fachgruppe Archivare an Archiven der Parlamente, der Politischen Parteien, Stiftungen und Verbände: *Mitteilungen der Fachgruppe* 6 36, 2012, S. 165–170.

Bundesgesetz über den Datenschutz (DSG), 01.03.2019. Online: <https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/de>.

Whitepaper: 3-Heights™ Document Converter. Grundlagen und Anwendung, 2016. Online: <<https://www.pdf-tools.com/public/downloads/whitepapers/Whitepaper-DocumentConverter-DE.pdf>>.

Bundesgesetz über die Archivierung (Archivierungsgesetz, BGA), 01.05.2013. Online: <<https://www.fedlex.admin.ch/eli/cc/1999/354/de>>.

Anhang 1

Metadaten-Feld	Anzahl nicht exportierbar EML-Format	Anzahl nicht exportierbar MSG-Format
Object Type	0	0
Content-Length	12288	12287
Content-Type	0	0
Message-Bcc	12285	12283
Message-Cc	8651	8653
Message-From	3	5
Message-Recipient-Address	12288	814
Message-To	818	820
Message:BCC-Display-Name	12288	12284
Message:BCC-Email	12288	12284
Message:Raw-Header:X-MS-Exchange-Organization-Recipient-P2-Type	12288	12287
Message:CC-Display-Name	12288	8653
Message:CC-Email	12288	8654
Message:CC-Name	12288	9699
Message:From-Email	2080	5
Message:From-Name	233	5
Message:Raw-Header:Content-Transfer-Encoding	9372	5773
Message:Raw-Header:Content-Type	9372	5705
Message:Raw-Header:Date	12288	5417
Message:Raw-Header:From	12288	5420
Message:Raw-Header:Importance	11631	11756

Message:Raw-Header:MIME-Version	0	6622
Message:Raw-Header:Message-ID	0	6507
Message:Raw-Header:Priority	11771	12046
Message:Raw-Header:Received	12288	6428
Message:Raw-Header:Return-Path	12282	6526
Message:Raw-Header:Subject	12288	5539
Message:Raw-Header:Thread-Index	4	8237
Message:Raw-Header:Thread-Topic	207	8018
Message:Raw-Header:X-Priority	11769	11399
Message:Raw-Header:To	12288	5455
Multipart-Boundary	2916	12287
Multipart-Subtype	2916	12287
Message:Raw-Header:X-MimeOLE	12288	7846
Message:Raw-Header:X-OriginalArrivalTime	12288	6530
Message:Raw-Header:content-class	12288	11786
Message:To-Display-Name	12288	821
Message:To-Email	12288	822
Message:To-Name	12288	4306
X-TIKA:Parsed-By	0	0
X-TIKA:Parsed-By	0	0
X-TIKA:digest:MD5	12288	12287
X-TIKA:digest:SHA256	12288	12287
dc:creator	3	5
dc:description	12288	210

dc:subject	135	12287
dc:title	135	137
dcterms:created	0	3
dcterms:modified	12288	3
meta:mapi-from-representing-email	12288	6
meta:mapi-from-representing-name	12288	6
meta:mapi-message-class	12288	3
meta:mapi-sent-by-server-type	12288	4638
resourceName	12288	12287
Message:Raw-Header:Accept-Language	12288	12250
Message:Raw-Header:Content-Language	12288	12180
Message:Raw-Header:Received	12288	6428
Message:Raw-Header:Authentication-Results	12288	12287
Message:Raw-Header:User-Agent	12288	11930
Message:Raw-Header:DKIM-Signature	12288	12274
Message:Raw-Header:Envelope-To	12288	12287
Message:Raw-Header:X-GMX-Antispam	12288	12287
Message:Raw-Header:X-Spam-Flag	12288	11738
Message:Raw-Header:X-UI-Filterresults	12288	12287
Message:Raw-Header:X-UI-Sender-Class	12288	12287
Message:Raw-Header:X-MS-TNEF-Correlator	12288	8237
Message:Raw-Header:x-originating-ip	12288	12287
Message:Raw-Header:X-MS-Has-Attach	12288	10049
Message:Raw-Header:X-MS-Exchange-	12288	12287

Organization-AVStamp-Mailbox		
Message:Raw-Header:X-MS-Exchange-Organization-AuthAs	12288	12287
Message:Raw-Header:X-MS-Exchange-Organization-AuthMechanism	12288	12287
Message:Raw-Header:X-MS-Exchange-Organization-AuthSource	12288	12287
Message:Raw-Header:X-MS-Exchange-Organization-MessageDirectionality	12288	12287
Message:Raw-Header:X-MS-Exchange-Organization-Network-Message-Id	12288	12287
Message:Raw-Header:X-MS-Exchange-Organization-SCL	12288	12287
Message:Raw-Header:X-MS-Exchange-Processed-By-BccFoldering	12288	12287
Message:Raw-Header:X-MS-Exchange-Transport-EndToEndLatency	12288	12287
Message:Raw-Header:ARC-Authentication-Results	12288	12287
Message:Raw-Header:ARC-Message-Signature	12288	12287
Message:Raw-Header:ARC-Seal	12288	12287
Message:Raw-Header:In-Reply-To	8825	10733
Message:Raw-Header:References	9984	10794
Message:Raw-Header:X-MS-Exchange-CrossTenant-Network-Message-Id	12288	12287
Message:Raw-Header:X-MS-Exchange-CrossTenant-fromentityheader	12288	12287
Message:Raw-Header:X-MS-Exchange-CrossTenant-id	12288	12287
Message:Raw-Header:X-MS-Exchange-CrossTenant-mailboxtype	12288	12287

Message:Raw-Header:X-MS-Exchange-CrossTenant-originalarrivalttime	12288	12287
Message:Raw-Header:X-MS-Exchange-CrossTenant-userprincipalname	12288	12287
Message:Raw-Header:X-MS-Exchange-Organization-AVStamp-Mailbox	12288	12287
Message:Raw-Header:X-MS-Exchange-Transport-CrossTenantHeadersStamped	12288	12287
Message:Raw-Header:X-MailCleaner	12288	12285
Message:Raw-Header:X-MailCleaner-ID	12288	12287
Message:Raw-Header:X-MailCleaner-Information	12288	12285
Message:Raw-Header:X-MailCleaner-ReportURL	12288	12287
Message:Raw-Header:X-MailCleaner-SPF	12288	12287
Message:Raw-Header:X-MailCleaner-SpamCheck	12288	12285
Message:Raw-Header:X-NewsI	12288	12287
Message:Raw-Header:X-NiceBayes	12288	12287
Message:Raw-Header:X-Spamc	12288	12287
Message:Raw-Header:X-TM-SNTS-SMTP	12288	12287
Message:Raw-Header:received-spf	12288	12287
Message:Raw-Header:x-forefront-antispam-report	12288	12287
Message:Raw-Header:x-forefront-prvs	12288	12287
Message:Raw-Header:x-microsoft-antispam	12288	12287
Message:Raw-Header:x-microsoft-antispam-message-info	12288	12287
Message:Raw-Header:x-microsoft-antispam-prvs	12288	12287

Message:Raw-Header:x-ms-exchange-senderadcheck	12288	12287
Message:Raw-Header:x-ms-exchange-transport-forked	12288	12287
Message:Raw-Header:x-ms-office365-filtering-correlation-id	12288	12287
Message:Raw-Header:x-ms-oob-tlc-oobclassifiers	12288	12287
Message:Raw-Header:x-ms-publictraffictype	12288	12287
Message:Raw-Header:x-ms-traffictypediagnostic	12288	12287
File Original Path	0	0
File Original Name	0	0
MD5	0	0