

Arbeitskreis Wissenschaftliche Archive in Berlin

Protokoll zum Workshop E-Mail-Archivierung

Zeit und Ort: 30.11.2016, Archiv der International Mathematical Union (IMU)

Thema: E-Mail-Archivierung

Teilnehmende: 23 Personen, s. Teilnehmerliste im Anhang

1. Begrüßung

Eröffnet wurde der Workshop durch Birgit Seeliger (Leitung IMU-Archiv). Nach einer kurzen Vorstellung des Archivs und der Workshop-Teilnehmer, ließ Frau Seeliger ein Gruppen-Foto der Teilnehmer erstellen. Dieses Foto kann von jeder beteiligten Einrichtung und für den Archiv-Blog www.berlinerarchive.de frei verwendet werden. Die Teilnehmer gaben ihre Einwilligung.

2. Einführung ins Thema

Frau Starkloff (Archiv MPG) beschreibt die Struktur von E-Mails bestehend aus dem Header (diverse Informationen können für die Forschung genutzt werden, da sie anders als beim analogen Brief vom Rest der E-Mail abgetrennt beforscht werden können), dem Body (enthält i. d. R. die Information/Nachricht) und den Attachements (Anlagen; können diverse Formate und Inhalte aufweisen). Insgesamt ist die E-Mail selten ein in sich geschlossenes Dokument. Häufig gleicht sie aufgrund der ebenfalls enthaltenen vorangegangenen Korrespondenzen eher einem Rotulus als einem in sich gesondert betrachteten Schreiben. Trotz zahlreicher strukturierter Metadaten (vornehmlich im Header etc.) handelt es sich meist um unstrukturierte Dateiablagen. Innerhalb der Clientumgebung weisen E-Mails, je nach Client, unterschiedliche Formate auf. Demgegenüber haben sie auf der Serverebene ein einheitliches Format (überwiegend MIME).

Frau Seeliger stellt Aspekte einer OAIS-konformen E-Mail-Archivierung und deren Unterschiede bzw. Parallelen zu analogen Archivalien vor. Dabei ist die Beziehung zwischen Archivar und Nutzer von zentraler Bedeutung. Da digitale Medien drohen, sehr bald und schnell verloren zu gehen, bleibt für ihre Bewertung und Übernahme oftmals wenig Zeit. Kriterien der Authentizität und Integrität zerfließen mehr und mehr, wenn zugleich Funktionalität und Zugang zu den digitalen Archivalien gewährleistet sein sollen. Datentransformation (bei der Übernahme: Kopie, Migration in Langzeitformat) und Refreshing (erneute Migration der Daten in neue Langzeitformate) verändern die digitalen „Originale“. Diese Arbeit mit den Daten im Langzeitarchiv und ihre Speicherung verursachen Kosten und benötigen Manpower, die vom Archivträger mitgetragen werden müssen. Die E-Mail als formatfreier Datenstrom ist ein Sonderfall des digitalen Archivguts. Aufgrund der Vermischung von privater und dienstlicher Nutzung, müssen hier insbesondere Persönlichkeitsrechte des Account-Inhabers beachtet werden.

Sich an die Einführung anschließende Nachfragen und Bemerkungen:

Es gibt beim Archiv der IMU zwei Möglichkeiten zur Übergabe von E-Mails¹:

¹ Genaueres zum Archivierungsverfahren folgt im Kapitel 4.

1. Kopie des eigenen Accounts in einen temporären Account auf IMU-Server (Account-Inhaber behält seine E-Mails, kann Ordner anlegen, E-Mails löschen, Sperrfristen für einzelne Ordner separat vergeben).
2. Mailinglisten (werden vom Server automatisch archiviert).

Mit Hilfe der eindeutigen Mail-ID kann die Speicherung von redundanten E-Mails verhindert werden. Da die benötigten Speicherkapazitäten für E-Mails jedoch kein entscheidendes Kriterium für oder gegen eine Archivierung sind, können Dopplungen durchaus in Kauf genommen werden. Hingegen können E-Mail-Anhänge weitaus mehr Speicherplatz benötigen als die E-Mail selbst (i. d. R. PDF-Dateien, Textdokumente, Fotos, Audio- und Videodateien). Priorität hat aber in erster Linie die möglichst vollständige Übernahme der E-Mails, auch wenn zurzeit noch nicht alle Hilfsmittel für z. B. Migration und Konvertierung entwickelt sind. Es gilt auch hier, wie bei der Archivierung von digitalen Unterlagen im Allgemeinen, den ersten Schritt so bald wie möglich zu unternehmen und von Fehlern zu lernen.

Um die Notwendigkeit der E-Mail-Archivierung gegenüber dem Archivträger zu verdeutlichen bedarf es grundsätzlich keine anderen - von der analogen Archivierung abweichenden - Argumente. Die Sensibilität und Vorsicht gegenüber der Speicherung von digitalen Dokumenten ist jedoch ausgeprägter. Im Bereich der persönlichen Vorlässe müssen Archive wie bisher die Unterlagen, einschließlich der E-Mails, rechtzeitig einwerben und die Genehmigung des Vorlassers einholen. Aufgrund der sowohl privaten und dienstlichen Nutzung der E-Mail-Accounts sind die Abgebenden gefordert, ihre E-Mails dementsprechend getrennt zu strukturieren. Außerdem haben sie die Möglichkeit selbst Sperrfristen zu vergeben.² Im Bereich der amtlichen Überlieferung müssen hausinterne Regelungen in der eigenen Einrichtung getroffen werden, um auf die Abgabe und Archivierung des E-Mail-Accounts hinzuweisen und das Verfahren durch den Archivträger zu legitimieren. Eine Abgabepflicht für E-Mails gibt es laut GGO nur insoweit, dass alle wichtigen Dokumente ausgedruckt und der Papierakte beigefügt werden sollen. Dies scheitert aber häufig am Geschäftsgang der Einrichtung und es entsteht in der Behörde meist eine Art „kreative digitale Ablage“. Um diese zu strukturieren, ist die Einführung von Dokumentenmanagementsystemen in der Verwaltung ein Instrument zur Lösung des Problems.³ Bei der Einführung müssen Archive, neben dem Einsatz von Records Managern oder Registratoren in der Verwaltung, als wichtige Berater fungieren.

3. Erfahrungen und Fragen der Workshop-TeilnehmerInnen

Im Vorfeld des Workshops wurden den Teilnehmerinnen und Teilnehmern ein Fragebogen zur E-Mail-Archivierung in der eigenen Einrichtung übermittelt. Die ausführliche Auswertung der gegebenen Antworten sowie die zusammenfassende Präsentation für den Workshop finden sich im Anhang zum Protokoll.

Ergänzt wurden in der anschließenden Diskussion folgende Punkte:

- Beim **Verfahren der E-Mail-Archivierung** im Archiv der Akademie der Künste handelt es sich lediglich um einen ersten Versuch, der nur für die Nachlassbetreuung angewendet wird. Probleme bestehen noch bei der Speicherung der E-Mails auf dem Server. Es gibt die Idee, einen Aktenplan einzuführen, die E-Mails mit eindeutigen Aktenzeichen zu versehen und einem Vorgang zuzuordnen.
- Wichtig ist, dass zuerst ein **strategisches Konzept** mit Aufstellung archivwürdiger E-Mail-Accounts entwickelt werden muss.

² Bei der IMU wurden Guidelines verfasst, in denen Kategorien zur Sperrung von Dokumenten zentral festgelegt wurden.

³ In der Berliner Verwaltung soll bis 2023 durch das E-Government-Gesetz die elektronische Aktenführung flächendeckend eingeführt werden. Bisher existiert noch kein schlüssiges Konzept zur Umsetzung des Gesetzbeschlusses.

- Die Archivierung von E-Mails basiert grundsätzlich auf keiner anderen Sachlage als die Archivierung von Papierakten. Schwierigkeit ist die fehlende Akzeptanz und **Unterstützung des Archivträgers**, da Vorbehalte gegenüber der elektronischen Aktenführung und –speicherung bestehen und es sich um vergleichsweise junges Schriftgut handelt. Eine mögliche Strategie zur Erhöhung der Akzeptanz beim Abgebenden ist, das technische Verfahren der Speicherung näher zu erläutern (E-Mails liegen hinter einer sicheren Firewall und sind besser geschützt als auf dem eigentlichen E-Mail-Server).
- **Weitere Strategien zur Einführung einer E-Mail-Archivierung** können sein: Einholung einer Genehmigung der betroffenen Person zu Lebzeiten, Festlegung grundsätzlicher Regelungen innerhalb der Einrichtung durch Rundschreiben oder Richtlinien mit Verpflichtung zur Pflege eines privaten Ordners im E-Mail-Account, Vergabe von speziellen, möglicherweise auch längeren Sperrfristen einzelner E-Mail-Ordner oder des gesamten Accounts (festgehalten in der Anlage zu Schenkungs- und Depositaverträgen des Vorlass- oder Nachlassgebers).
- Es sollten **Intervalle zur regelmäßigen Übernahme** der E-Mails ins Archiv mit dem Abgebenden vereinbart werden, um vorzeitige Löschungen aufgrund von Speicherplatzmangel o.Ä. zu vermeiden. Dies ist im Vergleich zur einmaligen Abgabe von papiernem Schriftgut eine Änderung im bisherigen Übernahmeablauf, der der Flüchtigkeit des Mediums geschuldet ist und einen Umgewöhnungsprozess nach sich zieht.
- Die **Kassation von privaten E-Mails** durch das Archiv ist nahezu unmöglich, da der Aufwand zum Filtern von kassablen und archivwürdigen E-Mails zu hoch ist. Mit Text Mining Werkzeugen könnten E-Mails mit bestimmten Stichworten herausgezogen werden. Jedoch müsste hierfür ein Thesaurus gepflegt werden, da man aufgrund der Heterogenität der Inhalte nicht auf Normdaten zurückgreifen kann. Außerdem könnten auch interessante dienstliche Inhalte in privaten E-Mails versteckt sein. Daher ist die Archivwürdigkeit von einzelnen E-Mails kaum zu überprüfen.
- Es ist nicht möglich, Wissenschaftlern die Nutzung des Accounts für private Zwecke zu verbieten oder eine Abgabepflicht festzulegen, da das geistige Eigentum und die Kreativität der Persönlichkeiten – aber auch die Haltung des Archivträgers – dem entgegenstehen.
- Die **Nutzung der archivierten E-Mails** im Archiv richtet sich wie bisher nach den archivrechtlichen Regelungen, d.h. Prüfen von schutzwürdigen Belangen Dritter (z.B. externe Korrespondenzpartner), Entgegenstehen vertraglicher Regelungen mit dem Vorlass- bzw. Nachlassgeber, Nutzung unter bestimmten Auflagen, etc. Die Bereitstellung der E-Mails erfolgt nach deren Erschließung im Archivinformationssystem im Virtuellen Lesesaal, beispielsweise über das Filtern nach bestimmten Stichworten oder Namen; diese Vorauswahl müsste aus datenschutzrechtlichen Gründen vonseiten des Archivs getroffen werden.

4. Vorstellung der bisherigen Verfahren zur E-Mail-Archivierung

Projekt Programmierung eines unterstützenden Tools zur E-Mail-Archivierung (Fr. Starkloff, MPG)

Gemeinsam mit einer Informatik-Professorin der Freien Universität Berlin hat das Archiv der Max-Planck-Gesellschaft ein Projekt initiiert, für das eine Gruppe Studierender gewonnen werden konnte. Im Rahmen eines Masterseminars soll ein Anwendungsprogramm oder zumindest die Grundlage entstehen, die den Übernahmeprozess von E-Mails in unterschiedliche digitale Langzeitarchive ermöglicht. Hierbei sind verschiedene Prozesse, wie das Filtern von vorab definierten Adressen oder Mailbox-Namen, die Erhebung von Metadaten sowie das Konvertieren in archivierungsfähige Formate einzubinden.⁴ Diese

⁴ Genaueres zur Anwendung kann dem Anhang zum Protokoll entnommen werden.

Prozesse müssen teilweise aus vertrauensbildenden Gründen automatisiert und ohne Sichtung der übergebenen Mails erfolgen können.

Das Masterprojekt wird in Form eines Scrum-Projekts durchgeführt. Ein Lastenheft, das die Ablaufprozesse, potentiell vorhandene Programme, die eingebaut werden könnten und Literaturhinweise listet, liegt vor. Innerhalb des regelmäßigen Austauschs mit den Archivaren soll eine gemeinsame Sprache entwickelt werden, um letztendlich das benötigte Resultat zu erzielen.

Seitens des Archivs der MPG wird ein stetiger fachlicher Austausch mit Informatikern und Archivaren, sowohl in der MPG als auch mit Archiven die über Erfahrungen im Bereich Langzeitarchivierung verfügen, gepflegt. Fachliche Unterstützung für das Projekt wurde von verschiedenen Stellen zugesagt. Das Endprodukt soll als Open-Source-Tool frei zur Verfügung gestellt werden. Einzige Bedingung ist, dass vorgenommene Erweiterungen wieder der Öffentlichkeit zur Verfügung gestellt werden müssen. Offene Fragen gibt es noch viele: Wann ist der beste Zeitpunkt, um an die Abgebenden heranzutreten und in welchen regelmäßigen Zeitintervallen? Führt man Gespräche zur Übernahme der E-Mails schon bei Einstellung der betreffenden Person?

Ergebnisse der anschließenden Diskussionen:

- Zukünftig muss der Archivar viel früher agieren. Die Vorfeldarbeit gewinnt wieder an Bedeutung, da klar sein muss in welcher Struktur die E-Mails vorliegen und um den Abgebenden von den Vorteilen zu überzeugen.
- Problematisch ist es, qualifiziertes Personal für solche Projekte zu gewinnen. Wünschenswert sind Personen mit einer Informatik- und Archivausbildung. Eventuell kann eigenes Personal geschult werden.
- Sperrfristen (auch die Archive betreffend) können über z. B. Rechenzentren oder gesicherte, vom Netz getrennte, Rechner garantiert werden.
- Anhang und E-Mail bleiben bei der Archivierung grundsätzlich zusammen, können aber für die Nutzung einzeln abgefragt werden. Es ist außerdem angedacht ein Validierungstool zur Überprüfung der Art des Anhangs auf Archivfähigkeit einzubinden.

Praxis der E-Mail-Archivierung des Archivs der IMU (Fr. Seeliger)

Die elektronische Archivierung der E-Mails wird von Seiten der IMU gefördert. Sachmittel werden zur Verfügung gestellt. Auch aufgrund geringer Personalkapazitäten wird ein hoher Automatisierungsgrad angestrebt, um so auch der Zeitnot der Abgebenden Rechnung zu tragen. Als Langzeitformat wird die PDF/A genutzt, um möglichst alle Daten langfristig zu sichern. IMU interne Guidelines werden erarbeitet (u. a. hinsichtlich der Sperrfristen). Bisher wurden 14 Accounts ins Archiv übernommen. Erste Dateirepräsentationen als Tar-file und PST-file werden auf dem Langzeitspeicher geführt, welche als erste Einträge mit weiteren Informationen (Provenienz, Daten etc) in der Archivsoftware verzeichnet werden.

Als Langzeitformat ist neben XML auch das PDF/A Format etabliert, das selbst durch Metadaten definiert ist und mit benutzerdefinierten Metadaten angereichert werden kann. Header, Body und Attachment werden ins Langzeitarchiv überführt, indem zur Konvertierung webPDF mit Open-Source-Software zu einem Programm verknüpft wurde. Dieses ermöglicht die originäre Struktur des abgegebenen Accounts in hierarchischen Ordnern, Unterordnern und den jeweils in ihnen enthaltenen E-Mails einschließlich ihrer Attachments abzubilden und in der gleichen Hierarchie zu speichern. Zusätzlich liegen neben den PDF/A- Dateien auch die Originale (der Attachments) vor. Nicht alle Formate sind langzeitarchivtauglich wie z. B. animierte Dateien und Fotos, aber die Priorität der E-Mail-Archivierung wurde von vornherein auf ihre Essenz gelegt, während die Repräsentation sekundär ist. Wichtig ist es, die Zusammenhänge zu erhalten.

Kern der Datenmigration sind neben Protokollen und Prüfsummen sogenannte Indexdateien (PDF/A-Format), die in jeden Ordner eingefügt werden. Sie stellen eine Art

Inhaltsverzeichnis dar, indem sie für jeden Ordner die Inhalte (weitere Unterordner sowie E-Mails mit ihren Attachments) in Kurzform aufzählen. Mittels der Archivsoftware FAUST kann über die Indexdateien eine Referenz zur ersten Objekt-Aufnahme (Provenienz) erstellt werden. Die Archivsoftware wurde so konfiguriert, dass sie alle Mailbox-Elemente (Originale und PDF/A) des gesamten Accounts automatisch als Stapel liest. Metadaten werden in die Felder der Software eingelesen, Text-Inhalte werden in einer Wortliste invertiert, Dateipfade werden automatisch aufgenommen und dadurch auf die entsprechenden Objekte verlinkt. Referenzfelder erlauben das Verfolgen des Kontextes einer Email oder eines Attachments. Der Nutzer kann also nach Objekten oder Namen bzw. allen Metadaten recherchieren und dabei verschiedene Strategien wie Index- oder Volltextrecherche verfolgen. Außerdem besteht die Möglichkeit, die archivierten Emails über Indices mit anderen Beständen zu verknüpfen. Es ist auch möglich, aus PDF/A XML-Files zu exportieren und damit (vorher und nachher) Schnittstellen zu schaffen.

Ergebnisse der anschließenden Diskussion:

- Verschiedene Formate (PDF/A oder XML) sind in Bezug auf den benötigten Speicherplatz unproblematisch.
- In der Archivsoftware der IMU wird jede archivierte E-Mail automatisiert als Datensatz angelegt. Es handelt sich hierbei um eine automatisierte Einzelblattverzeichnung. Jedes Archiv muss selbst über den Grad der Erschließungstiefe und der Automatisierung entscheiden. Die erfassten Metadaten wiederholen sich in jedem Objekt, sind aber nicht in jedem Objekt vollständig ausgefüllt. In der Archivsoftware werden nur ausgewählte Metadaten dargestellt.
- Die Archive müssen für ihre Einrichtung selbst definieren, welche Metadaten relevant sind und einen Metadatenkatalog erstellen.
- Grundsätzlich besteht die Möglichkeit einer Nachnutzung des Archivierungs-Verfahrens der IMU (bis auf webPDF ist alles javabasiert und damit Open Source).

5. Schluss des Workshops

Nach dem Dank an alle Teilnehmenden, die Organisatoren und insbesondere an Birgit Seeliger für die Bereitstellung der Räumlichkeiten, folgt ein Austausch über mögliche Themen für das nächste Treffen des Arbeitskreises (Bewertung, Personalakten, Erstellen von Findmitteln, Erschließungsrichtlinien, Erschließung von Fotos, Karten, Plänen, Bauakten, Flugschriften, Möglichkeiten einer Verbunderschließung). Zur Abstimmung wird es eine Doodle-Umfrage geben. Das nächste Treffen soll etwa im März 2017 stattfinden. Es besteht außerdem die Idee, einen Workshop zur Beratung bei Einführung von DMS mit Einladung eines Kollegen aus Regensburg zu veranstalten. Die Idee trifft auf breite Zustimmung. Das Landesarchiv Berlin kann möglicherweise Räumlichkeiten zur Verfügung stellen.