

電子メール保存の未来

電子メールアーカイブズの技術的アプローチ
に関するタスクフォース報告書（仮訳）

2018年8月

助成機関：



COUNCIL ON LIBRARY AND INFORMATION RESOURCES

ISBN 978-1-932326-59-8
CLIR Publication No. 175
Published by:

Council on Library and Information Resources
1707 L Street NW, Suite 650
Washington, DC 20036
Website at <https://www.clir.org>

Print copies are available for \$20 each.
Orders may be placed through CLIR's website at <https://www.clir.org/pubs/reports/pub175/>



Copyright © 2018 by Council on Library and Information Resources. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Library of Congress Cataloging in Publication Data:

Names: Task Force on Technical Approaches for Email Archives, author. |
Andrew W. Mellon Foundation, sponsoring body. | Digital Preservation
Coalition, sponsoring body.
Title: The future of email archives: a report from the Task Force on
Technical Approaches for Email Archives, August 2018 / sponsored by The
Andrew W. Mellon Foundation and Digital Preservation Coalition.
Description: Washington, DC: Council on Library and Information Resources,
[2018] | Series: CLIR publication; no. 175 | "August 2018." | Includes
bibliographical references and index.
Identifiers: LCCN 2018027625 | ISBN 9781932326598 (pbk. : alk. paper)
Subjects: LCSH: Electronic mail messages. | Electronic records. | Digital preservation.
Classification: LCC CD974.4 .T37 2018 | DDC 004.692--dc23
LC record available at <https://lccn.loc.gov/2018027625>

(日本語版仮訳について)

本稿は、クリエイティブコモンズ 表示 - 非営利 - 継承 4.0 国際 (CC BY-NC-SA 4.0) ライセンスにもとづき、図書館情報資源財団 (Council on Library and Information Resources: CLIR) から刊行された上記報告書を翻訳したものです (表紙画像と図は翻訳対象外としました)。なお、日本語に翻訳し公開することについては、著者の Christopher Prom 氏らからも、電子メールにて快諾の連絡をいただきました (2019 年 4 月 17 日付)。心より感謝申し上げます。誤訳等ありましたら、ご指摘いただけると幸いです。

2019 年 4 月
聖学院大学基礎総合教育部准教授 塩崎亮

目次

謝辞	vi
電子メールアーカイブズの技術的アプローチに関するタスクフォースについて	vii
要約	1
1. 電子メールアーカイブズの潜在的な可能性.....	3
1.1 伝承者としての電子メール	4
1.2 電子メールアーカイブの特異性	6
1.2 技術の活用	7
1.4 変化への適応	8
2. 電子メールの管理ライフサイクル.....	10
2.1 組織の記録としての電子メール	10
2.2 私的記録・寄贈資料としての電子メール	11
2.3 電子メールのライフサイクル	12
2.3.1 作成・利用	14
2.3.2 評価・選別	14
2.3.3 受入	17
2.3.4 整理	17
2.3.5 保存	18
2.3.6 利用提供	19
3. 記録技術としての電子メール.....	21
3.1 電子メールの定義	21
3.2 システムアーキテクチャ	22
3.2.1 アーキテクチャの特性	23
3.2.2 利用機能	24
3.2.3 運用・管理機能	24
3.2.4 電子メールメッセージのデータモデル	26
3.2.5 メッセージの構成要素	27
3.3 アカウント	29
3.4 データ転送モデル	29
3.5 電子メールの脆弱性	31
3.6 ASCII メッセージ以外の付加要素	31
3.6.1 添付データ	31
3.6.2 リンク先・メッセージ外の資源	33
3.6.3 署名欄	34
4. 現行のサービスと動向.....	35
4.1 進化する電子メールの生態系	35
4.1.1 不正使用、不正防止、セキュリティ、到達可能性	35

4.1.2 マーケティングと電子商取引サービス	37
4.1.3. 一般向け電子メールサービス	38
4.1.4 企業向け電子メールサービスと運用	39
4.1.5 電子メールのストレージ、法令遵守、記録管理	39
4.1.6 法令遵守と法的ツール	41
4.2 保存機関の課題	43
4.2.1 電子メールの収集	43
4.2.2 真正性の保証	46
4.2.3 整理・保存の履歴管理	48
4.2.4 添付ファイルやリンク先コンテンツの保存	50
4.2.5 セキュリティ・プライバシーの保証	54
4.2.6 大容量・大規模コレクションの処理	55
5. ありうる解決策とワークフローのサンプル	60
5.1 保存戦略	60
5.1.1 ビット列の保存	60
5.1.2 マイグレーション	61
5.1.3 エミュレーション	62
5.2 柔軟なワークフロー設計を支える相互運用性	64
5.2.1 複数ツール間での処理	64
5.2.2 コミュニティで共有可能なデータモデルの策定	65
5.2.3 ファイル形式の要件定義	66
5.2.4 API と相互運用性	67
5.3 ワークフローと実装シナリオ	70
5.3.1 ビット列保存のワークフロー	70
5.3.2 マイグレーションのワークフロー	71
5.3.3 エミュレーションのワークフロー	77
6. 今後の方向性：提言と次なる段階	80
6.1 コミュニティ形成とアドボカシー	80
6.1.1 障壁の低い／短期的な課題	81
6.1.2 影響の大きい／長期的な課題	83
6.2 ツールのサポート・検証・開発	86
6.2.1 障壁の低い／短期的な課題	87
6.2.2 影響の大きい／長期的な課題	88
付録 A. システム処理の自動化	93
付録 B. 電子メールのツール：図書館、文書館、博物館向け	94
Archivematica	94
DArcMail (Digital Archive Mail System)	96
EAS (Electronic Archiving System)	97

ePADD (Email: Process, Appraise, Discover, Deliver).....	99
Preservica Standard Edition	101
TOMES Tool (Transforming Online Mail with Embedded Semantics).....	103
付録 C. 電子メール保存の調査プロジェクト	105
電子メール保存に関するシンポジウム	105
カーカネットプレスの電子メール保存プロジェクト	105
CERP (Collaborative Electronic Records Project)	106
DAVID (Digital Archiving in Flemish Institutions and Administrations)	106
Kaine Email Project@LVA.....	107
MeMail (Email Preservation at the University of Michigan)	107
PeDALS (Persistent Digital Archives and Library System)	108
TOMES (Transforming Online Mail with Embedded Semantics)	108
付録 D. 参照情報源	110

図一覧

第 1 図. 電子メールのライフサイクルの各段階	13
第 2 図. 電子メールメッセージのデータモデル	27
第 3 図. ビット列保存の基本ワークフロー	71
第 4 図. ハーバード大学図書館におけるマイグレーションのワークフロー	72
第 5 図. スタンフォード大学図書館におけるマイグレーションのワークフロー	74
第 6 図. スミソニアン協会における XML マイグレーションのワークフロー	76
第 7 図. エミュレーションのワークフロー例	77

謝辞

電子メール保存の技術的なアプローチに関するタスクフォースは、アンドリュー・メロン財団学術コミュニケーション局の支援を受けた。次の方々に感謝申し上げたい：メロン財団の Donald J. Waters、Patricia Hswe、Kristen C. Ratanatharathorn、Tasha Garcia、Molly McGrane-Cleary、Celia Bradley、英国デジタル保存連合（DPC）の William Kilbride、Sarah Middleton、Sharon McMeekin、そして参照情報源を管理してくれた、Courtney Cain（レイクフォレストカレッジ）、Shreya Udhani（イリノイ大学アーバナ・シャンペーン校）。

また、次の方々や機関にも感謝の意を記したい。Artefactual Systems 社には、Joel Simpson の参加をサポートいただいた。その他、ハーバード大学の Grainne Reilly、Skip Kendall、Keith Pendergrass、Preservica 社の Jon Tilbury、スタンフォード大学の Josh Schneider、Peter Chan、ウォータールー大学の Maura R. Grossman、Gordon V. Cormack、FWD:Everyone 社の Alex Krupp、全米歴史出版物記録委員会（National Historical Publications and Records Commission: NHPRC）の Nancy Melley、ノースカロライナ自然・文化資源局の Kelly Eubank、Jeremy Gibson、Sarah Koonts、米国・州文書館長評議会（Council of State Archivists）の Anne Ackerson である。

あわせて、デジタル保存連合が開いた 2 日間の会議に参加いただいた方々、ワークショップ「デジタル革命の後で」においてプレゼンテーションの機会をいただいた Lise Jaillant にも謝辞を述べたい。これら会議の参加者からは、有益なフィードバックをいただいた。本報告書にもその内容を反映させている。もちろん、誤りや不十分な箇所については、すべて我々の責任である。

電子メール保存の技術的なアプローチに関するタスクフォースについて

2016年11月、アンドリュー・メロン財団とデジタル保存連合（DPC）は、電子メール保存の技術的なアプローチに関するタスクフォースを立ち上げ、広報を行った¹。本タスクフォースの目的は、保存コミュニティの取り組むべき課題を提言することであり、本報告書はその成果物である。我々は、概念的・技術的な枠組みを提示した。電子メールを保存する現行の各取り組みは、互いに相反する個別の解決策としてではなく、必要に応じて柔軟に相互利用可能なツールとして捉えられる。また、コミュニティの発展に不可欠な要素・領域を本報告書内で記す。本タスクフォースの詳細については、ウェブサイトを参照されたい（<http://www.emailarchivestaskforce.org/>）。

委員会メンバー

Christopher Prom（共同委員長）イリノイ大学アーバナ・シャンペーン校
Kate Murray（共同委員長）米国議会図書館
Fran Baker マンチェスター大学
Matthew Connelly コロンビア大学
Wendy Gogel ハーバード大学

タスクフォースメンバー

Hillel Arnold ロックフェラー・アーカイブセンター
Courtney Cain レイクフォレストカレッジ
Euan Cochrane イェール大学図書館
Kevin De Vorsey アメリカ国立公文書記録管理局
Glynn Edwards スタンフォード大学図書館
Riccardo Ferrante スミソニアン協会アーカイブズ
William Kilbride デジタル保存連合（DPC）
Jessica Meyerson Educopia Institute [非営利組織]
Erin O'Meara アリゾナ大学図書館
Michael Shallcross ミシガン大学
Joel Simpson Artefactual Systems 社
Camille Tyndall Watson ノースカロライナ州立アーカイブズ
Richard Whitt グーグル
Julian Zbogor-Smith マイクロソフト

注意事項

本報告書の内容はタスクフォースメンバーが属する機関の見解を反映したものではない。

¹ 公式なプレスリリースについては次を参照のこと。 <https://mellon.org/resources/news/articles/mellon-foundation-and-digital-preservation-coalition-sponsor-formation-task-force-email-archives/>.

要約

電子メールは、1971年にレイ・トムリンソンが単純な文字列のメッセージ「QWERTYUIOPのようなもの」を自身に送信して以来、大きな発展を遂げてきた。1980年代後半から1990年代前半までには、仕事や日常生活上における紙ベースでのやり取りが、電子メールに取って代わり始めたといえる。私的なメッセージや考えを一瞬のうちに送信できる手段として、電子メールは単にデジタル上の人生を記録するだけのものではなく、人生そのものを記録してきた。

電子メールは、いわば伝承者のようであり、また語り部のようなものでもある。26億を超える人々が現在電子メールを利用しており、1日平均にすると2,150億もの送受信メッセージが流通しているともいう。日々のおしゃべりが交わされる裏で、電子メールという痕跡が蓄積されている。果たして、未来の歴史家は、電子メールアーカイブズを駆使することにより、明日には歴史となっている今日という日を紡ぎ合わせることができるようになっていくだろうか。

文書館や図書館が行動を起こすことにより、電子メールに刻まれた証跡は収集・保存され、利用に供されうる。しかし現状、体系的な取り組み例はわずかでしかない。問題の一端は、その複雑さである。電子メールは一つのシステムから成るのではなく、作成、送受信、閲覧、保管といったサブシステムが複雑に相互作用して成り立つ。電子メールを保存するには、多数のプロセスが絡む。アーキビストは、提供者との間で信頼関係を築き、コレクションを評価し、対象資料を複数箇所から収集し、データを処理し、プライバシーや法的課題に対応し、メッセージや添付ファイルを保存し、利用を促していかねばならない。

電子メールを保存する作業はまだ新たな試みだが、本報告書は、アーカイブズ機関がこのもっとも複雑な問題に対処し始めていることを物語ったものともいえる。公開は当面想定せず、単純に、データを投入・保管しておくアプローチを選択する機関もあろう。あるいは、エミュレーション技術を用い、元の環境で電子メールを研究者が操作できるように対応する機関もあろう。もっとも一般的なアプローチは、電子メールを標準的な形へ変換・正規化してしまうものである。各アプローチとも、それぞれメリットとデメリットをあわせもつ。相互に排他的でもなく、組み合わせて利用することも可能だ。共通しているのは、いずれも入り組んでいるという点である。電子メールの保存は、実行可能ではあろう。しかし、この電子的な通信記録を保存することが、アーカイブズコミュニティに共通の目標だとすれば、それを達成するには、実際に取り組んでいる機関の数が少なすぎる。紙資料についていえば、アーカイブズ機関が語られていない歴史を掘り起こす役目を果たしてきたが、デジタル資料についていえば、それと同じことができるまでには至っていない。

この状況を変えたいと望むのであれば、相互運用性が鍵となる。電子メールの世界を形作る各種プロトコルは、相当程度に標準化され、多様な電子メール環境間での相互運用性を実現してきた。それと同じように、電子メールを保存するツール類も、それぞれのライフサイクルとは関係なく、相互に動くようなものでなければならない。商用にせよオープンソースにせよ、それらツール類の主なものは、文化遺産コミュニティでも利用されている。だが往々にして、特に、より正確に直感的な検索ができるかという点と、また、データ量が膨大な

場合に削除・編集作業がやりやすいかという点、心許ない。追加的な投資がえられれば、APIの整備など、処理の自動化を進めることができる。それにより、ツール同士の連携が容易になり、よりシームレスなワークフローの実現につながろう。

本報告書で記されたワークフローの例は、多様な組織の考え方、方針を反映したものといえる。同時に、電子メールを評価、収集、処理、保存、提供し、コレクションを構築することは、困難を伴うが、実現可能であることも示している。保存機関の担当者は、様々なツールを組み合わせ、それらを組織固有のワークフローに落とし込んでいかねばならない。とはいえ、比較的資源が豊富にあり、技術力もある機関であれば現実味があるものの、大半のところでは、そううまくはいきそうもない。既存のツールで電子メールを保存できないからではない。端的に言えば、問題が複雑なためである。コミュニティもツールも生まれつつあるものの、まだ成熟しているとはいえない。多くの機関において、基本的な調査も未着手であるし、方針もまだ定まっていない。電子メールを保存し利用提供するための、相互運用可能なソフトウェアツールの開発、方針の策定が進みつつあるのが現状といえる。

本報告書の目的は、技術的なものと政策提言的なもの双方にまたぐ。何よりもまず、技術的な観点からいえば、本報告書では、a) 電子メールを保存する現行の取り組みについて検証・評価する。また、b) それらの取り組みを包含する概念的・技術的な枠組みについて述べたい。各取り組みは、互いに矛盾する個別の解決策としてではなく、必要に応じて柔軟に相互利用可能ないわばツールとして捉えられる。さらに本報告書では、c) この技術的枠組みを改善・洗練させ、既存のツールをこの枠組み内に位置付け、欠けた要素を埋めていくとともに、保存コミュニティが取り組むべき課題を明確にしていきたい。

補完的に、本報告書では、保存コミュニティがこの課題の解決に向けてより積極的に取り組んでいけるよう、政策的な提言を行う。電子メールを保存するということは、現代という歴史を記録し、調査研究を行うに当たり、必要不可欠かつ確かな投資といえる。大規模なデータ処理、センシティブな情報の自動抽出などの技術進展がなければ、また、電子メールの処理業務にそれら技術が組み込まれることがなければ、この大量の歴史記録は研究対象として扱えないまま取り残されていくことだろう。たとえ、歴史研究の関心対象になったとしても、諸権利が記録されたとしても、あるいは、説明責任が担保され、良質のサービスが提供されることになったとしても。

プレディクティブ コーディング（予測符号化技術）など、自動化技術によりコストを下げる事が可能となった一方で、制約要因は資金と時間に限らない。コミュニティ全体を通して、新たなやり方、アプローチを組み込む際の課題についても目を配る必要がある。新たなスキルと技術が求められ、資金調達やガバナンスにも変容が迫られる。端的に言えば、電子メールを保存する上で取り組むべき課題は、単純に技術・金銭的なものにとどまらない。

電子メールは、歴史的な記録として重要なものである。したがって、この記録を保存し、アクセスを保証することは、アーカイブズ機関やアーキビストの役割であり、その価値を決定づけるものともいえる。この課題に向き合わなければ、我々は、記録を保管する機関と位置づけられうるのか、責任を放棄していると見なされはしないか。何も手を打たなければ、この歴史記録が消失の危機にさらされることになってしまう。自ずと解消される課題ではない。取り組むべきは、「いま」である。

1. 電子メールアーカイブズの潜在的な可能性

年を経るにつれ、書類棚から大切なメモを、あるいは、靴箱から埋れていた手紙を発見するといった光景は徐々に見られなくなっている（訳注：米国などにおいては思い出の品を靴箱に入れる慣習があることから）。いまや、やり取りの記録は、ハードディスク、モバイル機器、そしてクラウドサービス上でひそかにファイリングされている。電子メールで交わされるメッセージの内容は、従来の紙ベースのコミュニケーションと著しく異なる。長い談話的なもの、理にかなった議論が展開されたもの、あるいは、私的な関係を記したものなどをも含む。だが電子メールは同時に、あらゆる表現手段を表すものともいえる。このメディアは、インフォーマルなやり取りを助長してきたといえる。仲間内でないと理解できない表現、不可解な脱文、遠回しな語り口など、多くのスレッドが会話に近しい。

後方には、未来の歴史家がそっと佇んでいる。歴史研究者らは、過去を描き出す際の一つの材料として、電子メールのアーカイブが利用できるようになる日を待つ。このような未来を現実のものとするため、今まさにアーキビストは、「送信」ボタンを押した後に残される、この電子的に価値あるものを収集、整理、保管していくことが求められている。

電子メールは、その終焉が何度となく唱えられてきたものの、実態は、広く普及し、日常生活上も重要なツールとなっている。我々の生活自体を記録したものであり、デジタル時代の初期世代で生じてきたことを反映するものでもある（Cerbain 2016）。ある研究者らの表現によれば、「生息域（habitat）」として電子メールは捉えられるという。

情報管理などの幅広いタスク、組織内の調整・共同作業に電子メールは用いられている。電子メールは、多数の作業が与えられ、まかせられる場所であり、また、オンライン出版物や情報サービスへアクセスする際のポータルの役割を担うようになっている。勤務日でなくとも、オフィス労働者が大半の時間を費やす場所となった（アプリケーションは常に動き続け、時に関心を集める）。（Ducheneau and Bellotti 2001, 30）

SMS（ショートメール）やインスタントメッセージのような他のコミュニケーション技術も利用され続けているが、もっとも広く受け入れられている形態は、現在でも電子メールといえる。2016 年時点で、世界中の電子メール利用者は約 26 億人であり、今後 5 年においてもその数は増え続けるであろうといわれていた（Radicati Group, Inc. 2016）。地球をも超えて利用できることは、その遍在性を物語ろう（Oberhaus 2016）。各宇宙機関は宇宙飛行士との連絡手段として電子メールを採用している。NASA は、工具や修復部品を現地で作成できるように、国際宇宙ステーションへ 3D の設計ファイルを送る際、電子メールを利用している（BSG Web Group 2017）。宇宙船を待つより、即座にデータを送付可能であることに利点があるのは明らかだ。とはいえ、この極度の遍在性、柔軟性、そしてスピードにより、ユーザがメッセージを作成・受信、または削除する際に実際起きている複雑さが見えにくくなってしまう。

電子メールのデータをアーキビストがどのように保存し、それらを利用に供するべきか、という課題は、研究者だけの関心事でない。家族史の調査を行う一般の人々、学校の課題に

取り組む学生、弁護士やジャーナリスト、政府の活動について知りたいと願う市民にとっても重要な問いである。しかし、我々アーキビストのコミュニティ、技術者、研究者らが、その保存と利用という課題に対する解決策を見出さない限り、さらに、その解決策を広く適用な形で普及させていかなければ、電子メールが有す豊富な情報は、日の目を見ないまま埋もれてしまう。

本報告書は2つの目的をもつ。

- 1) 新たな処理・保存・提供技術にもとづき、電子メールのコレクション構築をアーキビストが行いうる手法について吟味・推奨すること、さらに、対応できていない課題を特定し、追加の開発を提言すること。
- 2) 電子メールのコレクション構築を進める上で、技術的課題の解消に対して投資が必要なことを意識喚起していくこと。

想定読者には、保存コミュニティ、デジタル保存の専門家、技術者、ソフトウェア開発者、情報システム業者、歴史家、研究者、管理職者、資金提供機関、財団を含む。本報告書では、成功へ導くための具体的な道筋を描いている。とはいえ、ものごとを前に進めるには、ここで確認した課題や推奨事項について、幅広いコミュニティからの関与が欠かせない。

「手紙を保存してきた機関は、歴史研究にとって貴重な情報源となってきた。電子メールの利用が増すにつれ、この情報源の存在意義が問われることになる。電子メールを保存する標準的な手順を確立させるということは、現在の記録を将来に向けて残す上で欠かせない側面の一つだ。」Peter K. Bol ハーバード大学東アジア言語文化学科教授兼先端教育担当副学長（私信より）

1.1 伝承者としての電子メール

主たるコミュニケーション手段として、電子メールは、今日の個人と社会が織りなすストーリーを記録する。家族内の世間話や知人間のおしゃべりから、組織内の意思決定や政府の活動まで、世界中の電子メールアカウントに記録され続けている。New York Times の記事によれば、「避けがたいほど、不安定で、たまらなく便利だからこそ、電子メールは、気まづくなるくらいに私的な記録となりがちだ。我々がかき集めたかった話の詳細、つまりは時間、場所、気持ち、何を、いつ、どこで、誰がといった情報を内包している」(Manjoo 2017)。確かに、昨今のニュースで、個人・政治・ビジネス・学術的な話題が扱われる際、電子メールが情報源となっている例も多い。

ウィキリークス創設者のジュリアン・アサンジに関する電子メールの重要なメッセージを英国検察庁が 2014 年に削除してしまった事例は、国際的なニュースともなった (MacAskill and Bowcott 2017)。このデータ損失が法的にどのような影響をもたらすかまでは、よく分からない。英国検察庁側が、削除されたものが具体的に何か把握できていないためであり、「一度削除された電子メールアカウントの中身を知る術はもはやない」。

グプタリークスとして知られるデータセットは、南アフリカの富豪グプタ家の流出メール 10 万～20 万件から成るものだが、電子メールの価値を如実に示す例の一つといえる。グ

プタ家がどのように政治的影響力を駆使し、政府の方針決定に介入するとともに、当時のジェイコブ・ズマ政権に働きかけ、国営企業との有利な契約を結んでいたか、電子メールの内容から、その不正の実態が白日の下に晒された。「アフリカの内部告白者を保護するプラットフォーム (PPLAAF)」により、「取り調べ (起訴) の遂行を支援するため」いくつかのメッセージは2017年11月の10日間だけ公開された。電子メールが公開された元のウェブページはもはやないものの、ジャーナリストらに要望することでデータセットはいまでも入手できる (Organized Crime and Corruption Reporting Project 2017)。このデータセットからえられた情報は、南アフリカの政治・経済にとどまらず、国際政治・世界経済に多大な影響を及ぼした。ズマ氏は大統領を辞任せざるをえなくなり、関連していた大手PR会社ベル・ポッティンガー社は経営破綻に陥ることとなった (Alderman 2017; Onishi 2017)。

ビジネスの領域における例も挙げておく。フロリダのある法律会社では、受信者に知らせることなく、スパム扱いの電子メールを検出し、恒久的に削除するシステム設定にしていた。弁護士費用として最大100万ドルを支払う命令が下された内容を含む電子メールがスパムとして検出・削除されていたため、当該法律会社側は、期限までに異議申し立てする機会を失うはめになった (Weiss 2017)。

学術の世界においても事例がある。ニューメキシコ大学のスポーツ局長ポール・クレブスは、自身の電子メールから、寄付金の不正使用に関与していたことが明るみになった。2015年にスコットランドで行なわれたゴルフイベントでの不正使用についてであるが、これは州監査局が調査している案件の一つに過ぎない (Grammer 2017)。ニューメキシコ大学では、職員は州の行政法に準拠することを示している以外、記録保管に関する方針を特に定めていなかった。電子メールのメッセージは「一時的なもの」と見なされ、恒久的な記録管理の対象ではなかった (Dyer 2017b)。クレブスは自身に不利なメッセージを削除するよう職員に指示を出してさえいた。後に彼は、ニューメキシコ州が公的記録の保存と廃棄に関する方針を定めていることを知らなかったと述べている (Dyer 2017a)。

ブリティッシュコロンビア州のカナダ自由党事務局長兼選挙運動本部長だったローラ・ミラーは、オンタリオ州自由党が10億カナダドル以上のコストをかけて2つのガスパラントの建設中止を決定した件に関連して、2012年初頭に交わされた電子メールを削除したことにより、告訴されるに至った。公判においては、複数の行政高官により、通常の手順とは異なり、ミラーの電子メールアカウントが削除または消去されていないことが明らかとなった。ミラーとその同僚のアカウントが、ガスパラント2件の建設中止に関して、重要な記録となりえると考えられたため、当該アカウントは残されていたとのことである (Ferguson 2017)。

いわば「伝承者」、「ドキュメンタリー作家」、「仲裁者」として機能する電子メールの例は、政府内においてすら見いだせる。地域・国レベル双方の政治的次元を超えて事例は確認できよう。政府の活動・方針だけでなく、世論という法廷 (世論の下す判断) にも電子メールは影響を及ぼしてきた (Gearan and Rucker 2017; Cheney 2017; Rosica 2017)。有名な例として、30年以上も前になるが、イラン・コントラ事件を挙げておく²⁾。米国国家安全保障会議軍政担当に任命された米海兵隊中佐オリバー・ノースと米海軍中将ジョン・ポインデクスター

²⁾ イラン・コントラ事件の概略史は、参考文献 (Sabato 1998) を参照されたい。

は、ローカルストレージから、ニカラグアでの秘密工作が記録された 5,000 件あまりの電子メールメッセージを削除したが、そのコピーがバックアップのテープに残存していた (Johnston 1990)。このバックアップコピーは後に、米国国立公文書記録管理局 (NARA) へ移管された。米国政府の公的記録を保存する機関であり、ノースとポインデクスター両氏の行為を公のものとするため、証拠文書として利用に供されている。それには、意図的に世論操作し、法に反する政府活動を覆い隠すことを狙い、彼らが編集した誤った歴史年鑑を含む³。

1.2 電子メールアーカイブの特異性

電子メールのメッセージは、他の記録物と比較して、一過性の、はかないもののように見える。ここでいう他の記録物とは、かつて秘密にされていた日記、写真、私的な手紙などである。これらを読むのに技術的なスキルは (ほぼ) 必要とされない。電子メールは、一種のタイムカプセルともいえるが、長く残されるには、技術やシステム構成だけでなく、強い意志も欠かせない。明らかに保管すべき紙の記録は確かに存在する。長期保存するため、それらは図書館や文書館に移管されよう。しかし、電子メールやデジタル形式でのやり取りの記録が残されるためには、技術的な対応とともに、人の介入が欠かせない。他のデジタルデータと同じく、電子メールは大量に生み出される。だが、それらを長期的に保存することでえられるメリットは何か。これを明確に示すのは難しい。「情報時代」において、電子メール、ソーシャルメディア、あるいはクラウドベースのデータストレージといった形態で、デジタルのデータはますます生産されていく。アーキビストや歴史家は、我々の過去の大半が消失してきたことを十二分に知っている。記録を保管してきたものが誰もいないか、あるいは、誰もそうすることを望まなかったことによる帰結である。

電子メール保存の問題は、単にこの情報メディアが重要な情報を含むにも関わらず、あまりに一過性のものだからではない。むしろ、アーキビストが収集する前の段階と関係しよう。情報がどのように組織化されているか、あるいは、何も考えられていないか、という側面である。将来、研究者が電子メールのコレクションを利用できるようになるか、どのように利用できるようになるかは、最終的に、電子メールを送信・受信・管理する本人に依存してくる。

現状、アーカイブズ機関へ研究者が訪れると、アーキビストが資料探しを手伝ってくれよう。彼／彼女らは、当該コレクションについて深い知識をもつ。紙の資料やファイルを作成した本人が管理していたのと同じ秩序を維持しつつ、それらを保存しようとアーキビストは努めている。さらにいえば、各コレクション、そして下位の各フォルダの範囲と内容について記述した目録または検索手段 (finding aids) を作成してきた。こういった作業により、研究者は、もっとも適切と思われる記録史料を特定できる。それだけでなく、この知識の整理のされ方はまた、当該機関の「オフィシャルな考え方 (official mind)」を表すものともいえる。ボックス内の編成のされ方を含め、ファイル群自体から、研究者個人では思い付かなかったかもしれない関係性が見出され、しばしば予期せぬ発見へとつながっていく。

しかし、保存機関における電子的な記録、特に電子メールは、根本的に紙の世界と様相を

³ 歴史年鑑のコピーは、レーガン大統領図書館で確認することができる (Reagan Library 2017)。

異にする。旧来の紙でのやり取りの記録は、その内容と構造の点で統一的に扱いやすかった。他方、電子メールのコレクションには、フォーマルなものからインフォーマルなやり取りまで含まれ、中にはリストサードから送信された大量のメール、未承諾の広告もあり、量という点でもこれまでと違うことから、既存の記録管理のやり方では、不可能ではないにしても、対応が難しい。電子記録管理 (ERM) システムは、格納したデータを確実に保管し、高速な検索を実現できるよう設計されている。紙ベースのファイリングシステムや技術と同様、未来の研究者のために長期保存し、利用に供するという点は、EMR システム設計時には優先されないのが普通だろう。これらのシステムは、組織内の管理や監査において役立つようつくられている。おそらく遠い将来になるだろうけれども、組織の部外者である第三者が、ある記録の背景や意義を評価するのにサポートするためではない。

電子メールの検索に関する要件も、他の資料 (デジタル資料を含む) と異なる。既存のツールは、大量の電子メールから探し出す手段としては最適でないかもしれない。キーワード検索あるいは全文検索が効果的で主な検索手法となろうが、館内での利用に限定されることが多い。さらにいえば、その種の検索では、電子メールの特性を必ずしも活かしかれない。多数のユーザが時間をかけてやり取りした反応が記録された、スレッド単位での価値が失われてしまう。また、ある特定の単語に限らない曖昧な検索を行いたい場合にも、キーワード検索あるいは全文検索だけでは不十分である。電子メールのこの膨大な規模と量を前にして、アーキビストは、既存の検索手段、構造分析、編成記述では対応しきれない側面について、技術で補う必要性に迫られている。

1.3 技術の活用

電子メールが消失してしまうリスクが高いことをこれまで見てきたが、逆に明るい面も指摘できる。電子メールがボーンデジタルで、メタデータを伴って作成されるからだ。自然言語処理 (NLP) や機械学習など、コンピュータ科学の領域では、大量のテキスト集合を処理し、メタデータを抽出または生成し、各種コミュニケーションやソーシャルネットワーク上の何らかのパターンを明らかとするような、多数の新技术が作りだされてきた。データとしてのテキストを扱う技術により、新規かつ学際的な研究領域が生み出され、多くの社会学者、データサイエンティストの関心を引いている。

電子メールコレクションからの発見を改善するツールはいくつか見られるが、有力なものの一つとして、固有表現抽出 (named entity recognition) と呼ぶ処理技術がある。人、場所、組織などの固有表現を自動的に抽出し、機械処理しやすい形に解析するものだ。この技術により、もっとも頻繁に言及された地名または個人名などを研究者が特定できるようになる⁴。別の有力な技術として挙げられるのは、トピックモデルである。これは、同じメッセージ内で共起しやすい語の集合を特定することにより、あるコレクションのトピックを抽出する技術といえる。それにより、主題別にコレクションを編成することが可能となる。また、語の使用頻度を確認したり、同一メールにおいて異なる複数のトピックがどのように結び付

⁴ ePADD ツールでは、固有表現抽出の機能が組み込まれており、電子メールのメッセージ内に含まれた人名・組織名・地名でのブラウジングや可視化が可能となっている。固有名詞の外部データセットとして、Wikipedia、DBpedia、Freebase、Geonames、OCLC の FAST、米国議会図書館の件名標目表や名称典拠ファイルが活用されている (Stanford University 2018)。

けられているかを見たりすることもできるようになる (History Lab 2018)。

とはいえ、このようなアプローチに対する強い懸念もある。複数のアカウントへ自由にアクセスできてしまうことから生じる課題を払拭するのは難しい。プライバシーの観点から支持されにくいかもしれない。たとえデータの引き渡し前に作成者自身が機微情報をふり分けていたとしても。現状は、アクセス可能な範囲を制限する、あるいは、仲介的なアプローチとして、アーキビストが利用可能な資料を取り分けておくといった対応が取られることが多い。電子メールアーカイブズが成熟するにつれ、特に、人口知能や機械学習の技術が組み込まれていくにつれ、この状況は将来変わりうるかもしれない。だがその時まで、研究対象となる「データとしてのテキスト」は、アーキビストにより選択され、法的手続きに沿って公開された、あるいは、法の枠外で漏洩されたデータに限定される。

データサイエンティストは既に電子メールのデータセットを利用してきた。米エンロン社における電子メールのデータセットや IT 企業アボカドリサーチ社の電子メールコレクションなどが挙げられる。それらを用い、コミュニケーションネットワークの分析、社会階層の発見、著作者の同定、文書の自動要約といった実験が数千件にわたり実施されてきた (W. W. Cohen 2015; Oard et al. 2015)。電子メールのコレクションを保存しておけば、多数の様々な学問分野において、調査を実施できる機会が格段に増す。以前は不可能であった方法で、個人や組織に関する研究を行うことが可能となろう。アーキビストでは対処できない問題にデータサイエンティストを巻き込んでいくことも同様に重要だ。(法的または倫理的な観点から制限が必要な) センシティブなメッセージを特定すること、あるいは、利用しやすくするためにメッセージを自動分類することなどの課題が想定される。

1.4 変化への適応

電子メールは、既存の記録管理プロセスを揺るがす。電子メールは、多様な内容が扱われ、私的な要素が濃い。その量は膨大となり、内容は碎けたものだったり、その場限りの思いつきに過ぎないものだったりする。ある受信ボックスには、数千、数万、あるいは数十万件のメッセージが格納されうる。それらメッセージの内容は様々だろう。中には、センシティブな、秘密にしておきたい内容のものもありうる。あるいは、将来、法的に問題となりかねないものすらあろう。他のコミュニケーション手段以上に、電子メールを長期的に保存することは、恐怖心を引き起こしかねない。強い警戒感を示すのは、多くの人の自然な反応であろう。過度にメッセージを削除したり、公開猶予期間を長く設定したり、あるいは、端的に保存対象とされたくない嫌悪感を示したりする。保存コミュニティが負う責務は、資料の受入・評価選別・保存に関して、また、技術を活用し、機微情報の内容確認・編集・利用提供(後述)を行う上で、確固たる方針と行動を示すことにより、信頼関係を築いていくことにある。

既存の評価選別業務は、新たな処理形態に変容を見せ、普及しはじめている (Huth 2016)。データマイニングなどの新たな形態が出てきたことにより、未来の研究者はどのように電子メールコレクションを利用するか、これまで考えたことがないことをも考慮せねばならなくなっている。紙のコレクションと同じように、個々に取り上げれば、些末な記録に過ぎないようなものも、より広い文脈の中で捉えれば、重要で啓示的なものにもなりうる。機械

処理やデータマイニングは、出所の原則と結び付けられる際、多くの可能性をもたらす。例えば、それらの技術により、社会的な関係性やコミュニケーションパターンの全体像を、歴史を物語る上で欠かせない要素を描き出すことが容易になるかもしれない。

電子メールを長期保管し利用に供するために、電子メールのデータを取り込み、処理することは、いわばデータを飼いならすことともいえよう。具体的には、メッセージのヘッダ情報からメタデータを解析・抽出すること、添付ファイルのフォーマットを特定・分類すること、固有表現抽出・自然言語処理・トピックモデルの処理を施すことなどが挙げられる。これは、多段階のプロセス・判断を必然的に伴う、高いスキルを要す作業だ。各段階での判断に応じて、結果的に、データの保存のされ方、研究者への提示のされ方は変わってきかねない。そのため、一連の作業は、後からも十分に理解できるよう、記録されておかねばならない。

デジタル資料の受入が紙資料と同じように、あるいは、紙資料以上に行われるようになっていけば、ボーンデジタル情報のみ収集する機関が一般的な存在になる日も近い。実際、本報告書の執筆者の中には、メーリングリストなど、電子メールのみのコレクションを選別評価し保存できないか、依頼を受けたものもいる⁵。NARA の 2018-2022 年戦略計画では、2022 年以降、デジタル形式のみでの記録移管となることが明記されている (NARA 2018)。英国の Digital Lives プロジェクトでは、将来、日常生活のスナップショットを提供するため、パーソナルデジタルアーカイブのサンプルを収集機関は取得すべき、と提案されていた (John et al. 2010)。同じことが、電子メールアーカイブズによりはつきりと当てはまるのではないだろうか。スパムメールでさえ、社会史または文化史の研究者にとっては学術的価値をもつかもしれない (Brunton 2013)。フィンランドの Digitalia プロジェクトでは、パーソナルアーカイビングを促す試みが初めて試みられた。市民アーキビストに対し、低コストの使いやすいツールを提供し、自身らの私的な電子メールや他のデジタルコレクションの保存を促すプロジェクトである (Jääskeläinen, Kosonen, and Uosukainen 2017)。このようなアプローチにより、これまで記録史料として取り上げられることの少なかった多数の人々の電子メールが後世に残される確率を高められ、既存コレクションにおける社会階級の偏りを補いうるかもしれない。

しかし、保存対象が大衆にまで広げられたとしても、すべてのメールアカウントに等しくアクセスできるようになる訳ではない。他の記録史料と同様、その利用形態は、利用制限や公開猶予期間を定めた組織の方針に従うことになる。また、他のデジタルデータよりも、電子メールのコレクションは、技術的な制約がかかる。長年にわたり進化を遂げ、複数のシステムで稼働してきた電子メールの特性上、その保存と利用には課題を伴う。

⁵ 例えば、イリノイ大学により運営されているアメリカ図書館協会のアーカイブズにおいては、*Progressive Librarians Guild* のメーリングリストの記録が保存されている。
<https://archives.library.illinois.edu/alaarchon/index.php?p=digitallibrary/digitalcontent&id=361>.

2. 電子メールの管理ライフサイクル

本報告書の目的に照らし、電子メールのライフサイクルモデルは、組織の記録と私的な記録の2種類に分けたい。

2.1 組織の記録としての電子メール

組織の文脈において、電子メールは、事業を遂行するに当たりなされた議論・意思決定・各種活動について記録されたものといえる。したがって、電子メールの管理は、組織を取り巻く法規制により課せられた記録管理の範疇に入る。同時に、当該組織の管理・運営上、必要とされるものでもある（U.S. Securities and Exchange Commission 2002）。理屈の上では、電子メールは明確に定義された記録管理体系により統括されるべきであり、あらかじめ定められた記録の廃棄スケジュールに沿って評価選別はなされることになるだろう。さらに、正式な形で決定された記録管理計画では、保存目的で移管される対象が決められるだろう。また、長期管理するに値し、当該機関の記憶として、あるいは歴史的価値を有すものとして残すべき電子メールの範囲、逆に廃棄対象の定義がなされる⁶。だが、大規模かつ経営資源が豊富な組織でさえ、既存の記録管理手法を電子メールへ適用するには困難を伴う。ほとんどではないにしても多くの場合、ユーザの個々のアカウントが一般的な記録管理体系に位置付けられることはめったにない。この問題を解消するため、アメリカ国立公文書記録管理局では、Capstone と呼ぶアプローチを提唱した。より詳細については後述し、また、別の文脈で取り上げることになるが、州政府や大学の文書館でも採用されている考え方である。

電子メールのメッセージは、電子メール専用のアプリケーション以外において、具体的には、ある案件ファイル簿の一部として、収集・保持される場合もある。この場合、当該データは、印刷され、または PDF のようなフォーマットに変換された上で、写真や文書類など他の記録と関連付けられるかもしれない。エンベロープ（制御情報）やヘッダ情報、リンク先のコンテンツ、添付ファイルは、メッセージ本体とは切り離され、あるいは、完全に失われるかもしれない。この場合、使用されたアプリケーションや手順で、業務・法手続き上の要件を満たすに十分な情報をえられるか、考慮しておかねばならない。このニーズを満たす戦略がいくつか考えられうる。MBOX や EML などのネイティブ形式でデータを収集しておくこと、メタデータとしてのヘッダやエンベロープから情報を記録しておくこと、案件ファイル簿内に保管されたメッセージのコピーと一緒にリンク先コンテンツや添付ファイルを保持しておくこと、などが挙げられる。

紙の記録と同じく、法的責任が発生しうるものとして電子メールを認識している企業も多い。この場合、個人データや事業活動の記録を保存しておくということは、いずれ廃棄されることが仮定された上で、純粹にリスク管理の観点から理解されている。電子メールのメッセージが削除されるのは、単に安全を期して、当該情報が不利益をもたらしかねないと懸念してのことであろう。個人データ保護方針やその取り組みもまた、十分に理解されていな

⁶ 米国アーキビスト協会の用語集において、記録史料（archival records）という語は、電子的なものを含み、形式が何であれ、「それらが含む情報に永続的な価値があるため、あるいは、作成者の責任や役割を示す証拠として保存される」資料として定義されている（Pearce-Moses 2005）。

いことが多い。過度に慎重な態度が取られ、「自分の知らないことでは心は痛まない」姿勢が取られる。これは、保存に反した姿勢といえる。事業遂行にとって必要とないものとなれば、記録管理方針では通常、その記録は廃棄される。結果、歴史的な情報は失われてしまう。情報の損失により、負の影響が生じかねない。実際に記録管理を行う上では、一時的な価値しかもたない記録と、長期的な価値をもつ記録とを峻別せねばならない。それぞれ、保持と廃棄のスケジュールが定められることになるが、後者については、歴史的価値を考慮すべきことが推奨されている（NARA 1997）。

2.2 私的記録・寄贈資料としての電子メール

私的記録・寄贈資料としての電子メールの定義はやや曖昧である。一般的に私的記録としての電子メールは、自身のデジタルアーカイブの一部として、あるいは、家族のデジタルアーカイブの一部として形作られ、通常、組織における記録保管の枠組みに位置付けられることはない。もちろん、積極的にそれらデータを保存していない機関から寄贈された記録など、例外はあろう。とはいえ、そのような組織があったとしても、インフォーマルなボランティア活動として取り組まれてきただけで、事業活動の過程で、電子メールや他の電子的記録が半ば偶然に集積した結果に過ぎないかもしれない。このような組織の記録は私的記録ともなりえ、あるいは、寄贈資料として扱われることになるかもしれないが、いずれにせよ、あらかじめ定められた記録管理プロセス上で受け入れられる類のものではない。

私的な電子メールコレクションを集めようとする場合、アーキビストは、候補となる作家、科学者、政治家などを、そのキャリアの初期段階で見出し、長期にわたり彼／彼女らと仕事上の生産的な関係性を築こうとするのではないだろうか。資料が保存機関へ引き渡される前に、記録が適切に保管されるよう、アドバイスを与えながら。というのも、そういったコレクションでは、参照すべき記録管理計画が整えられていないためだ。このアプローチは、信頼関係を構築する上でも重要である。多くの人は、電子メールが自身のパーソナルアーカイブの重要な要素となりえることを自覚していない。逆に、自覚しているのであれば、プライバシーやセキュリティの問題を気にしているかもしれない。

初期の段階から作成者とアーキビストが共同して取り組めない場合でさえ、実際どのように仕事をしているか突っ込んだ話を聞き、貴重な背景的情報を集めようとするなどし、機会を伺うのがよい。そのような活動領域を、Digital Lives プロジェクトでは「拡張されたキュレーション（enhanced curation）」と表現していた。より具体的には、インタビュー内容を記録すること、作成者の仕事環境（PC 画面の場合もあれば、物理的な部屋の場合もあろう）の全景を撮影すること、個人の蔵書や持ち物を記録すること、ハードウェア、ソフトウェア、記録管理の方法に関したまばらな情報を収集すること、が含まれるかもしれない。このような作業を通じて、アーキビストと提供者は、選別・評価作業時のパートナーになりえる⁷。

収集機関が個人から電子メールのデータを寄贈、購入、貸与により受け入れる際には、法的拘束力のある契約が双方により結ばれる。この契約時、電子メールについて定めておくべき

⁷ Digital Lives プロジェクトでは、「アーカイブ志向の PIM（個人による情報管理）」研究の重要性が指摘されていた。情報ライフサイクル全体を捉え、個人のデジタルオブジェクトの真正性を保証すること、長期的な観点から作成者や所有者個人が利用・再利用できるようにすることを目指す研究が求められているという（John et al. 2010, x）。

事項については、本報告書の参考文献で詳細を記している（Task Force on Technical Approaches for Email Archives 2018c）。

2.3 電子メールのライフサイクル

長年にわたり、記録管理者やアーキビストは、業務を遂行していく上で、記録のライフサイクルという概念を用いてきた⁸。この概念では、記録のライフスパンの各段階が定義されている。作成から、廃棄あるいは長期保存、発見、アクセス、利用までの各段階である。ライフサイクルのモデルにはいくつか種類があるものの、大きくは変わらない。電子メールについていえば、次のようにまとめられるだろう。

- ・ 作成・利用：アカウント保有者
 - ・ アカウント保有者自身が作成すること（下書き・送受信を含む）、システムによりメッセージが自動生成されること。
 - ・ 日常業務または私的なコミュニケーションにおいて、電子メールやファイル添付を利用すること。
- ・ 選別・評価：記録管理者、アーキビスト、アカウント保有者
 - ・ 提供者と関係を築くこと。
 - ・ 電子メール廃棄の判断を行うこと（廃棄／削除、もしくはアーカイブズ機関への移管を含む）。
 - ・ 受入前の選別評価を行うこと。
- ・ 受入：アーキビスト
 - ・ 対象となる電子メールの受託責任や法的責任を負うこと。次の事項を含む。
 - ・ データの収集と移管について
 - ・ 現用資料から永年保存資料への移行について（政府資料や機関アーカイブズの場合）
- ・ 整理：アーキビスト
 - ・ 受入の作業を行うこと。
 - ・ 編成記述の作業を行うこと。
 - ・ 受入後の選別評価と内容確認の作業を行うこと。
 - ・ 制限事項の確認、公開猶予期間の設定、編集等を行うこと。
 - ・ 必要な保存レベルと利用形態を判断すること。
- ・ 保存：アーキビスト、技術専門家
 - ・ 電子メールのメッセージ、添付ファイル、メタデータを保存システムに投入すること。
 - ・ 保存活動を行うこと（マイグレーション、エミュレーションなど）。
- ・ 利用提供：アーキビスト、技術専門家、研究者
 - ・ 電子メールコレクションに関する情報（書誌データや利用条件などについて）を研

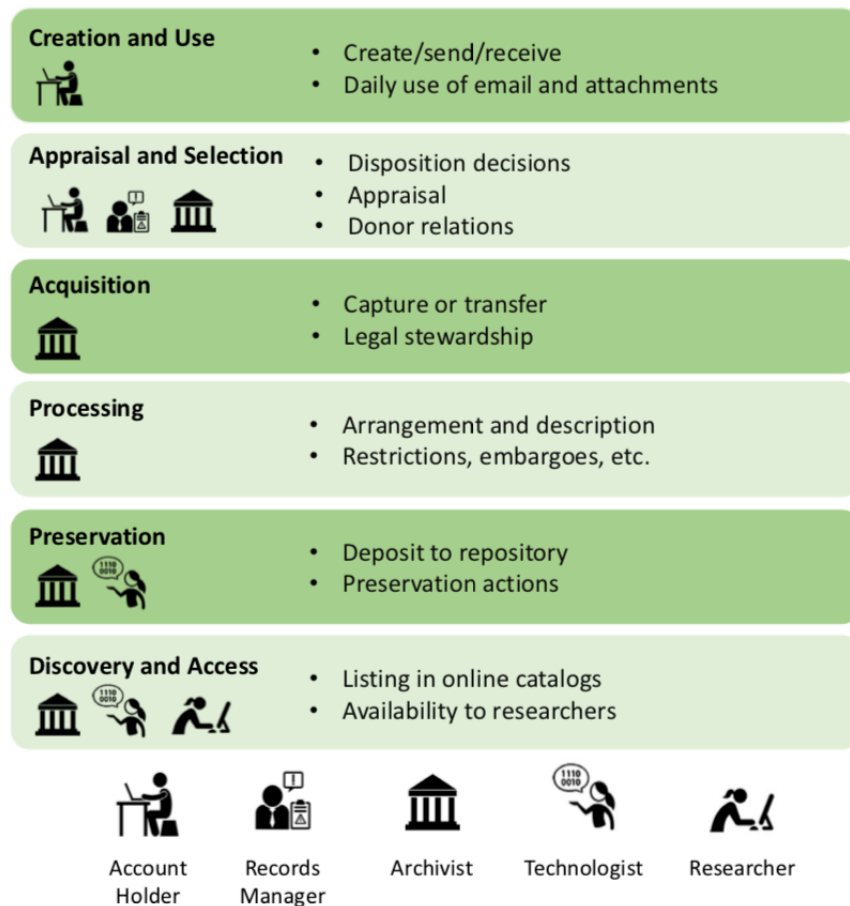
⁸ 英国のデジタルキュレーション・センターが策定したキュレーションのライフサイクルモデルは、あらゆるデジタル資料を管理・保存する上での高次モデルとして広く認められている。このモデルは、必要に応じて、ライフサイクルの各段階から検討を始めることが可能となっている（Digital Curation Centre 2018）。

究者に提供すること。

- ・ 電子メールのメッセージ、添付ファイル、メタデータを研究者に提供すること⁹。

電子メールのライフサイクルは、他のデジタルコンテンツのそれと異なる。初期の段階からアーキビストが関与する必要性が高いためだ。個人が晩年を迎えるまで、あるいは、亡くなられるまで、パーソナルアーカイブの収集を待つ必要があるかもしれない。理想的には、当該アカウントを作成者がまだ使用しているうちに、アーキビストやキュレーターが関与していくのが望ましい。このような初期介入が難しい場合は当然あろうが、作成者の私的なやり取りを恒久的に保存する上では重要な側面となろう。仮に、パスワードをアーキビストが入手できれば、直接または間接的に、そのアカウントへアクセス可能となる。それにより、事前に選別評価の作業ができ、保存計画も立てられるであろう。

第 1 図は、保存の価値があると判断された電子メールがどのようなライフサイクルの段階・プロセスを経ていくか、概観したものである。



第 1 図. 電子メールのライフサイクルの各段階¹⁰

⁹ 別のライフサイクルモデルとして、Paradigm プロジェクトにより策定されたものがある。パーソナルデジタルアーカイブに特化したモデルだ。デジタル情報のキュレーション活動が、アーカイブズ資料に関する既存のライフサイクル上にマッピングされている。使用されている語のいくつかは、OAIS 参照モデルから取られている (Paradigm Project 2008)。

¹⁰ 第 1 図で使用したアイコンは Noun Project (<https://thenounproject.com/>) のものだ。すべて、クリエイティブ・コモンズの CC BY 3.0 ライセンスである。ただし、研究者のアイコンはパブリックドメインである。各アイコンの帰属は次の通り：アカウント保有者 Gan Khoon Lay、技術専門家 Mazil from Noun Project の

2.3.1 作成・利用

作成・受信は、電子メールのライフサイクルの開始段階となる。モーリン・ペノックによれば、「キュレーションと保存は初期段階から開始される。したがって、電子メールのメッセージの作成者と受信者が、重要な最初のユーザとなる」(Pennock 2006, 15)。最終的に個々の電子メールが保存されるかどうかは、それぞれの送受信者が鍵を握っているが、その状況は様々であろう。企業等の組織で仕事上つくられた電子メールは、より厳しく管理されやすい一方、私的なものや、非営利組織やコミュニティ団体でつくられた電子メールの管理は、かなり緩いものとなるかもしれない。

この初期段階で重要なことは、どの電子メールクライアント（ソフトウェア）を対象とするか選択し、対象クライアントがメッセージをどこに保管しているか確認することだ。複数のプラットフォームを利用している場合、特に重要となる。また、電子メールクライアントによって、ファイル形式や添付ファイルの扱われ方も異なってくる。

核となる機能はコミュニケーション用途であるものの、電子メールは別の目的でも利用されてきた。個人による情報管理行動は多様であることを反映し、例えば、日々のタスク管理のため、会議予約などの日程調整管理のため、友人や同僚の連絡先を記録するため、などだ。

複数の調査で、バックアップや保存目的で電子メールを利用している人もいることが明らかとなっている。自分自身に対して、リマインド内容、ToDo リスト、重要なファイルをメール送信する場合などだ。あるファイルを失くしても困らないように、ファイルをメールに添付しておくことで安心感をえている人もいるかもしれない。この意味で電子メールは、事実上のパーソナルアーカイブとしても機能している。増え続ける大量のパスワードを電子メール内で保管している人さえいるという (John et al. 2010; Marshall 2008a, 2008b)。さらにいえば、電子メールの作成主体は人間に限らない。フォルダ監視やタスク管理ソフトでステータス変更が生じた際などに、自動生成される大量のメッセージを人々は受信している。こういったメッセージは、電子メールのアカウントと対象システムとが連動していることを示し、間接的にはあるが、アカウント保有者が活動していることを表す。

現用期間において、下書き、受信、送信フォルダでコンテンツをもっともいじるのは、作成者である。作成者は、このメッセージに何の情報が記されており、そのメッセージの文脈だとか、どのように整理されているか、もしくは整理されていないかを分かっている。受信ボックスがどのように整理されているかは、個々人によって大きく違う。大規模で複雑なフォルダ構成にしている人もいれば、検索や並び替えで用を済ます人もいるだろう。このような個人間での整理方法の違いは、使用しているシステムが影響している場合もある。あるいは、その人自身のパーソナリティを反映した結果の場合もあろう。いずれにせよ、アーキビストは、個人アーカイブの原秩序を尊重し、そこに価値を見出す。

2.3.2 評価・選別

相互排他的ではないものの、それぞれ異なる戦略を組み合わせることにより、評価・選別

は行われることが多い。米国政府機関において、整理対象となる電子メールを評価選別する初期のアプローチは、レコードスケジュールや他のガイドラインにもとづき、各メッセージの価値を検討するというものだった。政府機関の中には、保持する価値があると見なすメッセージを識別する（タグ付けする）よう、あるいは、適切なフォルダへメッセージを移すよう職員に指示していたところもあっただろう。機関によっては、保存用コピーとして電子メールのメッセージを印刷するように、あるいは、それらメッセージを別のフォルダへ移すように、オンラインシステム上へコピーを置くよう指示されている場合もあっただろう。

このアプローチは、組織における個人の役割を重視し、すべてのアカウントを評価選別の対象とするものであった。だが、個人に判断を求めるのは、時間を要し、徹底させるのには困難を伴うやり方だったといえる。2013年にNARAは、別のアプローチを取ることに変更した。これをCapstone（役職指定）アプローチと呼ぶ（NARA 2013a）。指定された記録作成者の電子メールアカウントが、長期的または恒久的な保持対象とされる。これにより、取り組むべき対象が絞られ、記録管理のガイドラインを遵守するのに必要な諸資源を削減することが可能となる。このCapstoneアプローチは、米国の政府機関がNARAへレコードを移管する文脈で提唱されたものではあるものの、州政府や大学のアーカイブズなど、他の機関においても導入されてきた（NARA 2013a）。ある特定個人（連邦政府機関の長官や直属の部下、大学の総長や学部長、あるいは多国籍企業のCEOなど）の電子メールが収集対象となり、保存されることとなる。この役職指定のアプローチは、次の3つの観点から擁護される。1) それら個人が作成・受信した電子メールの大半は公的な記録と見なせる、2) 電子メールの証拠価値はメッセージ内容によって異なる、3) 一定の役割を担う個人の電子メールをメッセージ内容にもとづき整理することは、実行不可能ではないにしても、その量の大きさからすると現実的でない、という3点である（U.S. Government Accountability Office 2008）。ある集団やプロジェクトに関するメッセージが、Capstoneアプローチの基準を満たさないものの、保存しておくのがよいと考えられる場合については、ある特定のフォルダにそれらメッセージを置くよう指示する組織もあるかもしれない。

受入前に、機微情報、私的な情報、制限されるべき情報が含まれていないか、ePADD、BitCurator、FTK（Forensic Toolkit）などのツールを用いて確認される場合もあろう。それにより、アーキビスト、あるいは資料の提供者にとっても、編集または公開猶予期間を設けるべき対象資料を事前に特定しておくことが可能となる。社会保障番号、電話番号、他の構造化データや名称を検索するのは容易といえるが、センシティブな話題、複数データを組み合わせられる事項など、曖昧な検索を要する場合は困難を伴う。より洗練されたアプローチが必要となり、時間を要す。対応期限に比較的余裕があり、機微情報を特定するツールが欠かせない場合、いったん保存機関へ移管し、後日、対応を検討する組織もあるかもしれない。

評価選別は、時間のかかる、労働集約的な作業である。そのため、コレクション単位でのみ評価選別を行い、より詳細な作業は受入後に実施とする場合も多い。大容量のストレージが低コストで用意できるのであれば、評価選別を必須としないと判断する組織があってもおかしくはない（The National Archives 2016）。電子メールの保管が法的に義務付けられていない場合であっても、歴史的または文化的な観点から、『すべてを保持しておく』ことでえ

られる便益が高まっている」(Pennock 2006, 18) とする見方は以前からあった。(作家などの) 個人アーカイブズの場合、評価選別作業は、いずれにせよ、ごくわずかなものにとどまるかもしれない。技術的なアプローチでできる範囲に限定されることもあろう。デジタルアーキビストやキュレーターは、対象となる個人のデスクトップ環境すべてを収集したいかもしれない。そこで、クライアント環境のデータを含め(とはいえ、ウェブベースのものは含まれないが)、対象コンピュータのディスクイメージを取得しようとする場合もあろう。この場合、評価選別の作業は、受入後に実施されざるをえない。

自身の電子メールを個人が管理するのと同じように、アーキビストが、対象者の電子メールアプリケーションを使用して、受入前に基本的な評価選別を実施できる場合もあろう。その他に、送信フォルダのみ保持しておくというアプローチもありうる。それにより、返信すべき、あるいは転送すべき重要なメッセージをすべて収集できるかもしれない。また、受信・送信双方の内容を一つのスレッドとして収集できる¹¹。同じような発想から、大規模な組織で見られる機能的なやり方を、個人や小規模組織でも適用できないだろうか。例えば、ある機能、プロジェクト、領域にもとづく記録を残すということが考えられる。

評価選別の一環として、記録の廃棄(保存機関へ移管するに当たり、それら記録を最終的に消去してしまうこと)に関しても多数のやり方が想定される。親機関の文書を受け入れる組織アーカイブズでは、移管または廃棄する前に、作成元ではどの記録を何年にわたり保持すべきか、記録の廃棄スケジュールが決められているところもあろう。電子メールのメッセージまたはアカウントが公的に位置付けられておらず、あるいは、歴史的な記録と見なされていない場合、それらはすぐに廃棄される対象となっているかもしれない。もしくは、職務遂行上、一時的に保持されているだけの位置付けかもしれない。組織における電子メールの管理規則が、このような記録管理の状況を生み出し、あるいは、影響を及ぼしてきたのかもしれない。なお、外部の資金提供者など、他の組織に対して記録管理の義務を負っている場合は、記録の廃棄スケジュールをあらかじめ定めておくのが望ましい。

理想的には、組織の構成員と IT 担当者として、記録管理スケジュールとその手引きが十分に検討された上で、電子メールの管理が実現されるのが望ましい。もちろん実際には、より柔軟な形で進められることが多い。例えば、IT 部門では、職員アカウントのうち、非常に古いメッセージを定期的に削除しているかもしれない。あるいは、組織内で定めた電子メールのデータ容量を超えてしまっている場合、不要メッセージをアカウント保有者が削除しない限り、それ以上のメッセージを受信しないよう、IT 部門は設定しているかもしれない。いずれのやり方とも、公的または歴史的に価値ある記録が失われてしまう危険性ははらむ。他方、廃棄スケジュールが整備されていない、あるいは、廃棄スケジュールで規定した廃棄を実行する手段が整備されていないために、退職した職員の分を含め、すべての電子メールを保持している組織もあろう。

パーソナルデジタルアーカイブが組織によって収集される場合、当該資料の廃棄(disposition)とは、初期調査や選別評価作業後に、受入対象の資料を移管する段階でなさ

¹¹ これは、ニューヨーク・フィルハーモニックのアーカイブズで取られた戦略である。とはいえ、ある美術館を対象とした調査では、「重要」と分類されたメッセージすべてを収集することは効果的でないと結論付けられた(Cocciolo 2016)。

れる作業を指す。通常、移管対象となる資料は、歴史的価値をもつと判断され、長期保存されるべきものとして扱われる。電子メールの場合、その対象範囲が問題となる。保存対象の電子メールは、個々のフォルダ、または全アカウント分のデータとなるかもしれない。この段階で、フォレンジックツールで特定された機微情報、あるいは長期保存される価値がない資料については、削除されることとなるかもしれない。

自身のデジタルアーカイブを管理してきた個人本人の側としては、この段階で、電子メールの一部または全部を残すことにするか、どのように処分するか決めることとなる。アーキビストの側としては、寄贈に関する覚書や他の法的文書にもとづき、電子メールの所有権をえなければならない。理想的には、提供者と保存機関との話し合いの上、利用制限または廃棄条件についても当該文書上で明記しておくのが望ましい。

2.3.3 受入

受入は、長期保存の対象となる電子メールが保存機関へ移される段階のことをいう。

電子記録管理システムを導入している組織であれば、当該システムを介し、対象となる電子メールを作成段階で収集できてしまう場合もある。そうでない場合、あらかじめ定めたレコードスケジュールに沿って、電子メールのデータは共有ネットワークドライブ上に出力され、保管されることになるかもしれない。保存機関への移行は、外付けハードドライブなどの物理媒体でなされることもあれば、FTP など電子的なやり取りでも行われるだろう。

記録管理システムが正式に導入されていない場合、あるいは方針が定められていない場合、職員の異動時または離職時にデータの移管が行われる場合も想定できる。この場合は、収集機関が個人の電子メールを収集するのと同じような過程を経る。具体的には、対象となるクライアントソフトから電子メールの該当データを出力するか、あるいは、PC 全体のディスクイメージを作成することになるだろう。

現用アカウントの公的な電子メールが組織アーカイブズへ移管される場合もあれば、第三者である個人のデータを収集機関が受け入れる場合もあるが、いずれの場合にせよ、確実に真正性を担保した上で移管を行うのが、主たる関心事となる。対象となる電子メールには、極めてセンシティブな私的情報を含んでいる場合があるため、移管作業は厳重になされなければならない。外部メディアは常に暗号化されるのが望ましい。保存機関で電子メールのデータを受領したら、まずはネットワークと接続していない検疫用のコンピュータに格納されることになるかもしれない。次に、機微情報が含まれていないかの確認やより詳細な評価選別作業が行われることになるだろう。受入後の評価選別作業が実施されてからも、切り離されたセキュアなネットワーク上で、データの処理や保存を行うことにしている組織もある。どのような移管手段をとるにせよ、移管後には、データの完全性など、問題がないか確認するのが望ましい。これには、電子メールのメッセージや添付のデータを対象とした、内容が変更されていないことを示す固定性情報に関するチェック、ウィルスチェックの実施を含む。

2.3.4 整理

電子メールが保存機関へ移管された後、整理作業が行われる。ある組織内の別の部門から

移管される場合もあれば、外部の第三者から寄贈または購入などで受け入れられる場合も想定できる。

対象となる電子メールの主な情報を集めておくため、この段階で、通常は受入記録が作成される。電子メールのみ移管されたのか、あるいは、保存対象資料（デジタル形式のみの場合もあれば、紙資料も含むハイブリッドなコレクションの場合もあろう）の一部に過ぎないのかも記される。この段階では同時に、受入前に実施された評価選別を見直し、より詳細な評価選別が行われることも多い。詳細な評価選別とは、機微情報の確認作業を含め、一部のサンプル調査にもとづき、資料作成者からの情報提供と、半自動化による確認作業とを組み合わせ実施される（例として次を参照：The National Archives 2016）。

詳細な確認作業が実施困難な場合、利用制限することを基本として、すべての電子メールアカウントのデータを保持しておく組織もあるかもしれない。いいかえれば、研究者から要望が出されるまで、それらデータは非公開としておき、いざ公開する段になってから、閲覧対象となるメッセージ内に機微情報が含まれていないか確認する、という対応である。あるいは、一定期間、（電子メールであろうが、デジタル形式でない資料であろうが）シリーズ単位で公開を猶予する方針を取る組織もあろう。形態はどうであれ、コンテンツを公開しないのであれば、内容の詳細を確認する作業は後回しにできる。例えば、スミソニアン協会のアーカイブズでは、会長、次長、部長などは重要な記録作成者として位置付けられ、それらの電子メールが収集されている。この電子メールの記録は、受入時点から 15 年後に公開される予定となっている。より長い公開猶予期間を設定している組織もあろう（Ferrante 2015）。ハーバード大学では、大学教員アーカイブズ [faculty archives]（以前は大学教員文書 [faculty papers] と呼んでいた）の利用は、個々の寄贈証書にもとづき管理されている。（アナログにせよデジタルにせよ）大学経営に関する記録は、その作成日から 50 年間は利用できない方針が定められているし、また、学生や職員など個人に関する記録は、少なくとも 80 年間は非公開扱いとなっている（Harvard University Archives 2018）。あるいは、プリンストン大学のアーカイブズでは、大学経営陣の電子メールを組織の記録として収集しており、40 年間は非公開の方針だという（Princeton University Department of Rare Books and Special Collections 2018）。

管理段階で行われるその他の作業として、ツールを用い、保存対象となる電子メールのメッセージや添付ファイルを分析することが挙げられる。ファイル形式やバージョン情報、作成・送信・修正日、作成者、ファイル容量などのメタデータが抽出される。また、電子メールアーカイブズの真正性や完全性を検証する作業も含まれる。

2.3.5 保存

ライフサイクルにおける保存とは、電子メールのメッセージ、添付ファイル、メタデータを保存用リポジトリに移す段階をいう。電子メールを格納し保存するにはいくつかの課題が避けられない。

- ・ 関連する電子メールのメッセージ群を単体の情報パッケージとして保存機関が扱ってもよいか。あるいは、各メッセージは別々の情報パッケージとして扱われるべきか。
- ・ 電子メールの添付ファイルは、受信時のとおり、メッセージ本文とあわせて管理される

のがよいか。あるいは、メッセージ本文との関係は別途記録した上で、個別に管理されるのがよいか。

- 電子メールのメッセージ間の関係性、スレッドという単位、フォルダなど他の分類情報を、保存機関側ではどのように扱うことになるか。
- 電子メールと分かるように、保存機関はどのような保存戦略を取るようになるか。

各メッセージを一つの提出用情報パッケージ（SIP）として扱うにせよ、あるいは、各アカウントを一つの提出用情報パッケージ（SIP）として扱うにせよ、いずれにしても、保管用情報パッケージ（AIP）には、元の形式の電子メールと、正規化した、または保存用の形式のものの双方とが含まれているのが望ましい。AIP を管理するため、件名、送信者のログ、添付ファイル、可能であれば、内容が変更されていないことを示す固定性に関するデータなどのメタデータがデータベース上に置かれ、理想的には、その AIP と関連ファイルとが一緒に格納されるのが望ましい。さらに、この AIP は、信頼に足るデジタルリポジトリへ移されることになる。

この段階で、持続的なアクセスを保証するため、保存機関により手が増えられる。出所履歴情報が記され、真正性を担保するために集められた（今後の保存活動を行う上でも必要となる）メタデータに、PREMIS で定められた要素、例えば保存活動、エージェント、その他付加データなどが追記されることになろう。その結果保存された電子メールは、元々受け入れていたもの（電子メールや添付ファイル）とは別に、新たな AIP として扱われることになる。信頼に足るデジタルリポジトリにおいて、元の AIP は、この新たな AIP に置き換えられることになろう。

2.3.6 利用提供

保存された電子メールは、未来の研究者、学生、家族、郷土資料家らにとって、貴重な情報源となる。多数の学問分野・領域において利用される可能性をもつ。それは、伝統的な研究領域にとどまらない。いままでにない、革新的な研究方法で扱われることになるかもしれない。

ごく基本的なことからいえば、個々のパーソナルアーカイブに興味をもつ人にとっては、手紙と同じように、電子メールのやり取りも確認したいものだろう。調査対象がデジタル情報のみから成るものにせよ、紙資料も含むハイブリッドなものにせよ、コレクション全体においてその電子メールがどのように位置付けられるかを見たい利用者もいるだろう。このような研究においては、現行のいわゆる目録（または検索手段）は変わらず有用なツールといえる。カーカネットプレスの電子メール保存プロジェクトでは、インタビューも実施されているが、その中で研究者らは、ハイブリッドなアーカイブにおいて、アナログとデジタルの要素を結びつけて考える有効な手段の一つとして、目録（または検索手段）が捉えられていたという（Baker 2015, 221）。もしくは、電子メールの本文とそのメタデータがあれば、目録の要素を自動生成する可能性が生まれよう。電子メールのヘッダからは、対象期間、送信者・受信者に関する情報をえられる。テキストマイニングや潜在意味解析の技術にもとづき、ある概念や名称を抽出することも可能だろう。逆にその結果を目録データに反映し、よ

り検索されやすくすることも考えられるかもしれない。結果として、目録データから、保存された電子メール自体を確認しようとする利用者も出てくるかもしれない。デジタルリポジトリに格納された電子メールは、別の検索インタフェースで提供されている場合もありえ、新たな検索や絞り込み検索などを行うこともできるかもしれない。

ある個人や組織の電子メールの内容に主たる関心をもつ研究者からすると、電子メールアーカイブの提供方法については特に強い意見はないかもしれない。とはいえ、少なくとも、クライアント PC 環境での閲覧、またはウェブブラウザでの閲覧に近い形で利用できることを期待しているのではないだろうか¹²。

あるいは、よる没入型の環境を望み、作成者本人のデスクトップ環境上で電子メールを見ているかのような体験をしたい研究者もいるかもしれない。エモリー大学ではサルマン・ラシュディのコレクションをそのような形で提供している¹³。対象となる電子メール環境が元々どのように動き、見えていたか、エミュレーション技術にもとづき再現できれば、この種の利用ニーズに応えられるだろう。また、今日の歴史研究者が当時の書簡類を解読するため古文書について学ぶのと同じように、未来の研究者は、旧式のハードウェアやソフトウェアの利用方法を習得する必要があるかもしれない。

電子メールのデータから主たるメタデータを抽出できるようになれば、それらメタデータが従来の「検索手段」に置き換わる可能性もあるし、あるいは、別の検索ポイントとして活用されうる。電子メールのヘッダ情報やメタデータからデータを可視化することも容易となろう。

著作権、個人データ保護、プライバシー、機微情報といった問題は、電子メールの利用・活用を制限する要因となる。保存機関や図書館で保管されることになった電子メールの場合、それらをオンラインで提供するとなると、事前に著作権者から許諾をもらい、かつ、機微情報が含まれていないか確認しておかねばならない。これは、特に規模が大きくなればなるほど困難な作業だ。保存対象のアカウントが1つだけだとしても、アカウント保有者以外に関係する権利者が数千人に及ぶ場合もある。ePADD のディスカバリモジュールなどのツールがあれば、編集加工した電子メールのデータをオンライン上で提供することもできる（送受信者名、地名、団体名などの「エンティティ」から検索できる）。だが現時点では、電子メールの本文が見られるのは施設内限定の機関が大半だ（Owens 2014）。

電子メールを収集保存する機関であれば、現在と未来の利用者がそれらデータをどのように利用したいと望んでいるか、あるいは望むであろうか、思い巡らせねばならない。さらに、未来の利用者が記録の真正性や完全性を確認したいと望む可能性もあることをも考慮すると、受入後に必要な処理を施したデータに加えて、受入時の元データをも保持しておくのがよいかも検討せねばならない。

¹² カーカネットプレスの電子メール保存プロジェクトでの調査結果によれば、送信先の指定フィールドである To や CC、日付、件名など、通常使用しているのと同じインタフェースを好む研究者が大半であったという。ただしこれは、何が何でも元の環境を再現して欲しいということではなく、差し障りのない程度に、という意味合いに過ぎない（Baker 2014, 28）。Preservica のようなリポジトリソフトウェアでは、電子メールをこのような形で表示できる。

¹³ エミュレーション環境に対する研究者側の見解については、参考文献（Rockmore 2014）に詳しい。

3. 記録技術としての電子メール

本報告書の主な目的は、電子メールの処理・保存・提供に関する手法を評価し、推奨案を提示するとともに、新たなツールやアプローチを導入することにより解決可能な課題を特定することである。このような目的から、電子メールとは何であり、どのように動くものなのか、明確に定義し、理解しておく必要がある。

3.1 電子メールの定義

電子メールとは、一つのモノを示すこともあれば、複数の意味合いを含むこともある。つまり、個々のメッセージを指す場合もあれば、メールボックス内にあるメッセージすべてを表す集合名詞ともなる。あるいは、動作動詞ともなる（例えば、「今日中にメールしておくよ」など）。こういった用法があるにせよ、電子メールは、あらかじめ定められた標準規則に従いメッセージを作成・送受信するシステムである。

電子メールは、構造化テキストであるメッセージを、ネットワークに接続されたコンピュータ間で交換するという比較的シンプルな道具として生まれた。だが今では、多様な機能をもつものとして進化してきた。多くの異なるネットワークやプロトコル間で電子メールは現状動いている。もちろん、文字で記されたメッセージと添付ファイルを単に送受信するためだけに電子メールを利用している人が大半であろう。とはいえ、個人による情報管理（personal information management: PIM）の文脈では、日程や連絡先の管理のために電子メールが活用されてきたともいえる。文字の代わりに音声のメッセージをやりとりできるボイスメールもある。タスク管理やプロジェクト管理のシステムにも、コメントやその反応、ToDo リストを送受信する仕組みとして、電子メールの機能は組み込まれてきた。編集用のシステムでも、コメントを送受信するために電子メールが利用されている（例えば本報告書も、Google Docs 上で共同編集された）。Twitter や Facebook などのソーシャルメディアにおいては、履歴を電子メールで残すことも可能だ。コンテンツ管理システムのなかには、データベースのバックアップをあるアカウントへ定期的に電子メールで送る設定が可能なものもある。このような電子メールの機能が組み込まれたシステムの場合、各機能の境界範囲は曖昧になるかもしれない。

しかし本来的に電子メールとは、送信者が受信者へ向けてメッセージを伝送する仕組みである。このやり取りが成り立つには、受信者側はメッセージの内容を理解できなければならないし、必要に応じて今度は逆に送信者側へ返信できなければならない。なお、実際のメッセージ本文をいう際に「メッセージ」という語を限定的に使用する場合もあるが、添付ファイル、外部コンテンツへのリンクや他の情報をもこの「メッセージ」という語に含める場合もある。

本タスクフォースのウェブサイトでは、電子メールの国際標準に関する情報をまとめている（Task Force on Technical Approaches for Email Archives 2018b）。だが、単に電子メールがどのように動いているかを確認したいだけであれば、それらの文書を読むのは苦痛だろう。使用される用語にも矛盾があることも多い。例えばだが、アドレスとメールボックスという用語はいずれも、メッセージが送られる宛先のことを指すとされている。RFC 5321 の 2.3.1

では、「電子メールのデータが格納されている場所（メールボックス）とその参照先（アドレス）とを分けて考える必要がない限り、通常それら2つの用語は同義に用いられる」との記述が見られる（Klen- sin 2008, 15）。

アーカイブズ領域の用語で、完全なレコードとは、必要な期間保存され、利用できることが担保されるとともに、証拠としての価値が維持されていることをも担保されるよう、十分な内容、文脈・構造情報を含むもの、と一般的には理解されている。ある活動を記録したものとして電子メールを保存することが目的なのであれば、保存の責務を負う担当者や組織は、対象となるメッセージやアカウントがどのような要素で構成されており、それら要素の関係はようになっており、システム上どのように格納されていて、メッセージの収集方法によって何が保持され、逆に保持されないことになるのか、ということ把握しておかねばならない。保存機関としては、完全なレコードが意味するものを理解した上で、電子メールの保存システムを開発し、研修の体制を整備し、方針を策定していく姿勢が求められている。それにもとづき、電子メールを保存し、利用に供する上で十分な情報を収集できるよう、方針と手順を実際に定めていくのが望ましい。

とはいえ、多くの要因から、この作業は困難なものとなろう。例えば、電子メールの標準規約では、メッセージを実際に伝送する際の制御データ（エンベロープ）は、メッセージ本体のメタデータに含まれたデータ（ヘッダ）と完全に同一でなければならない、とまで要求していない。標準では規定されていない、タグやニックネーム、別アドレスなどのメタデータを許容したアプリケーションも存在する。このような付加的なデータは、出力されないか、出力時に他との関係性が失われる。アカウント保有者の肩書きなどのデータは多くのメッセージから消されてしまう。また、記録管理や保存に必要な作業は、電子メールの見た目や挙動にも影響を及ぼす可能性がある。

3.2 システムアーキテクチャ

電子メールのコレクションを作成・編集、確認、整理、管理するためのアプリケーションは多数存在する。それらアプリケーションでは、それぞれ違うやり方でデータを蓄積しており、また、様々な機能が技術的な複雑さを生んでいる。したがって、そのアーキテクチャを理解しておくことにより、電子メールを保存する上での課題がより鮮明に見えてくると思われる。

既存の文献、特に技術的な解説がなされたものにおいては、関連するシステム全体がどのように動いているか説明されるとともに、電子メールのシステム概要について記されてきたといえる。例えば、InterPARES プロジェクトによる「電子メールの保管・保存」に関する報告書では、電子メールの機能が手際よく概観されていた（Pontevoipe and Salsa 2009）。クロッカーによる文献も同様に、電子メールのシステム概要を分かりやすく概括し、機能と設計の両面をまとめている（Crocker 2008）。本節では、より簡潔に電子メールの概要を見るところとどめたい。具体的には、電子メールのシステムアーキテクチャの特性に焦点を当て、このシステムがどのように維持管理され、受容されてきたかということについて述べる。次に、利用機能について記し、その後、アーキテクチャの目的やユーザのニーズを満たすために導入された運用・管理機能、各種標準について整理していく。

3.2.1 アーキテクチャの特性

技術的な観点からすると、電子メールの仕組みは、コンピュータのプログラミング言語と何ら異なることはない。ある入力と出力が繰り返され、一連の処理がなされる。想像しやすいイメージは、送信や転送のボタンをクリックすると、メッセージが気送管を通して指定した宛先に届く、というものだろうか。あるいは、郵便配達員にメッセージを手渡すと、配達員が郵便局まで持っていく、そこで仕分けされ（また他の郵便局へ回付され）、指定した宛先まで届けられる、というようなイメージが分かりやすいだろうか。RFC 5321 である Simple Mail Transfer Protocol (SMTP) では、すべてのコマンドを規定しており、具体的には、HELO（クライアントがサーバに身元を示す）、RCPTTO（受信者のアドレスを指定する）、DATA（メッセージと添付の伝送を開始する）などを含む（Klensin 2008, 15; Larramo 2018）。電子メールのクライアントソフトは、裏で仲介的な役割を演じてくれている。それにより、謎めいた名称のコマンド群をいちいち覚えなくても、私たちは電子メールの仕組みを利用できるようになっている。

電子メールは、異なるシステム間でメッセージがやり取りできるように設計されている。データ伝送が生じるまで、互いのことは知らない関係にある。そのアーキテクチャの特性として、次の点が挙げられる。

- ・ 世界規模でのアドレス指定：世界中のシステムにおいて唯一つとなるアドレスを用いなければならない。
- ・ 相互運用性：電子メールシステムの関係者間での事前調整は必要とされない。標準的なプロトコルにもとづき、送受信者双方ともメッセージの交換を安心して行うことができる。フォーマットも標準化されているため、メッセージが誤って解釈されることはない。
- ・ 非同期性：メッセージは中継して送られる。送信者と受信者がオンライン上で同時にやり取りする必要はない。
- ・ 冗長性：電子メールシステムのある要素に不具合が生じたとしても、システムとしては問題なく機能し続ける。ベストエフォートの原則にもとづき、メッセージは伝送されている。信頼度 100%のシステムを実現するには相当のコストと複雑な仕組みが必要となるものの、ここでは求められていない。
- ・ 分散型：電子メールシステムの管理と運用は、多数の異なるステークホルダーが分散して担っている。IETF（インターネット技術標準化委員会）や IANA（Internet Assigned Numbers Authority）などの管理団体は存在するものの、全体のシステムは集中的に制御されている訳ではなく、自己編成的な性質をもつ。例として、メッセージが伝送される経路は固定化されていない点が挙げられる。
- ・ 後方互換性：標準化されていることにより、システムの拡張と見直しを定期的に行うことができるが、旧標準との互換性を担保していかなければならない。後方互換性を担保することで、システムは徐々に変化させられ、また、別の機能同士で不具合が生じたり、予期せぬ挙動を引き起こしたりする可能性を避けられる。
- ・ 拡張性：通信プロトコルとデータ標準ともに、ある環境での機能拡張やカスタマイ

ズを許容している。それら拡張機能について、電子メールシステムの関係者全てが理解しておく必要はない（知らせる必要さえない）。この拡張性の性質により、世界規模での相互運用性を損ねることなく、自身らの要件を満たす変更を行うことができ、より革新的、実験的な試みも実現できよう。拡張機能は、完全に限定的なものとなるか、あるいは、IANA を介して登録することも可能である。

これらの核となる機能により、電子メールは広く普及したといえる。耐性があり、（インターネットの世界では）長く使用されてきた理由もそれら機能のためであろう。とはいえ、裏を返せば、トレードオフの関係といえるが、電子メールのシステムは、脆弱な側面をはらむ。具体的には、その複雑さ、なりすまし、スパム、フィッシングなどが挙げられる。

3.2.2 利用機能

電子メールのユーザは、メッセージを送信し、受信する人間、そして機械である。電子メールに対する関心の大半は、人間である利用者に向けられていると思われるが、電子メールのデータ内にはコンピュータが生成するメッセージも含まれ、電子メールのライフサイクルにおいては、それらもエージェント（動作主体）として捉えておくのがよいだろう。この場合のエージェント（動作主体）が人であれ機械であれ、電子メールには、一連の利用機能が含まれる。メッセージ本文を作成すること、添付ファイルを付けること、メッセージの宛先指定・送信・受信・閲覧・保管・削除・管理を行うこと、メッセージのメタデータを作成すること、自動処理を設定することなどが挙げられる。本タスクフォースのウェブサイトでは、これら機能の詳細情報を提供している（Task Force on Technical Approaches for Email Archives 2018d）。

3.2.3 運用・管理機能

電子メールの標準では、運用・管理機能についても規定されている。大半の電子メールサーバとクライアントに既にこれら機能は組み込まれているため、利用者とアーキビストの立場からすると、メッセージを長期保存する方法にも影響が生じうる。核となる機能には次を含む。

- ・ **メーリングリスト**：メーリングリスト運用の自動化は、電子メール標準で定められた主要機能である。利用者がメーリングリストのアドレスにメッセージを送ると、仲介機能によりさらにそのメッセージは伝送される。仲介機能においては、リストの各メンバー宛のメッセージを受信し、集約し、再送信が行われる（Crocker 2008）。ヘッダフィールド内のデータ（つまりメタデータ）も役立てられている。メーリングリストは、すべてのメッセージが保持され、メタデータにもとづき整理されるという意味において、保管機能をあわせもつことが多い。とはいえ、このメーリングリストの機能は、保存機関における長期保存の取り組みとは混同すべきではない。保存機関の主たる目的は電子メールを保存することであり、メーリングリストのような、後付けの副次効果ではない。
- ・ **配信機能**：メッセージを送信・受信・蓄積する電子メールのシステムでは、伝送状況を監視し、メッセージが確実に配信されたことを担保するため、標準的な機能がおおむね

組み込まれている。これらは、メッセージを「蓄積・転送」して伝送していく手法を支える機能である。ネットワークやインフラの一部に不具合が生じたとしても、この手法により、電子メールを確実に配信できる。DSN (delivery status notification) は、受信側のサーバが、送付されたメッセージの配信に遅延があった、失敗した、成功したといった状況を、送信者側へ通知する機能である。また、MDN (message disposition notification) は、当該メッセージが開封されたか、削除されたかを通知する機能である。なお、DSN も MDN も任意の機能だが、広く実装されてきた。

- ・ 独自管理・ポリシー強制：電子メールのプロバイダは、独自またはカスタム機能を提供していることがあり、独自のポリシーや規則に沿って実装されている。メッセージは、そこで定められた形式に変換されているかもしれない。ウィルスやスパムメールは隔離または削除されているかもしれない。メッセージは監視対象となっており、不適切な、あるいは、いかがわしい内容の場合はマスキング (redact) されるかもしれない。また、メールボックスの容量には上限が設けられているかもしれない。これは、上限を超えている対象者に、メッセージが送信できないことを通知するなど、配信機能と組み合わせで用いられるかもしれない。

運用・管理機能により、メッセージの内容やメタデータ（ヘッダフィールド）は書き換えられる、そして実際に修正されることも多い。結果として、まったく新しいメッセージが機械的に生成されている場合もあるかもしれない。とはいえ、見た目には、人間が作業した結果として理解されるだろう。この点には多大な利点があるものの（大規模なシステムになるほど後方互換性の原則が守られるため、この機能を取り除くのには限界があろう）、明らかに欠点も多々ある。真正性を担保すること、出所来歴を追えるようにすること、文脈情報を正確に収集しておくこと、などの課題をより複雑化させてしまう点が分かりやすい例となろう。

実際には、目に見える範囲でのヘッダ内のデータ、あるいは各メッセージの本文データは、システムやバージョンが異なっても、まったく同じではないにせよ、非常に似通っている。利用者側からすると、ヘッダに修正が加えられておらず、参照先の署名システムもまだ機能しているのであれば、メッセージの真正性を検証する手段は多数考えられる。だが、電子メールアーカイブズの場合、将来的には、そのような技術に頼ることはできないだろう。あくまで電子署名の仕組みは、対象メッセージが確実に配信されたことを担保することが主目的であり、その記録を永続的に残すためのものではないためだ。真正性の担保は、電子署名システムや証明書の有無に依存している。

2016 年 10 月、話題にはなっていないものの、電子メールの認証に関するニュースが取り上げられた。ウィキリークスにより公開されたキャッシュデータ内に存在する送信メールについて、送信者とされたものの中に、自分は送信していないと主張するものもいた、というニュースである。技術関連のブロガーであるロバート・グラハムは、真偽が問われたメッセージの真正性を検証するため、ヘッダ内にエンコードされた電子署名を確認したという (Graham 2016)。だが、グラハムが述べているように、証明書が元のサーバに残っていなければ、電子署名は照合されない。証明書が長期にわたり元のサーバに残っ

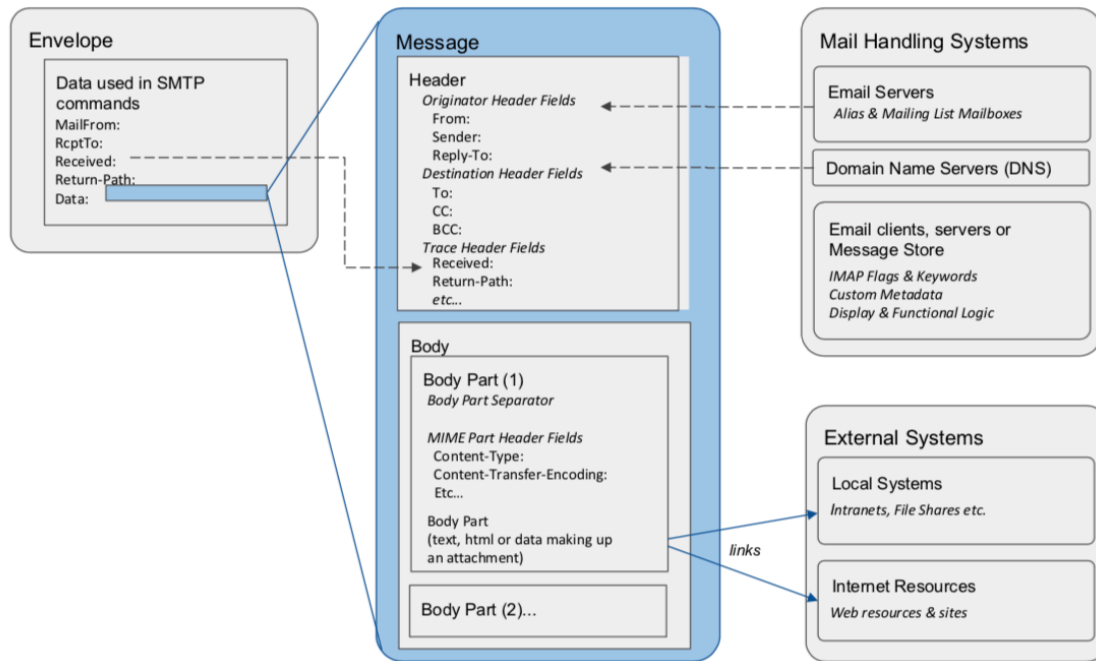
ていることは少ないと思われる。

3.2.4 電子メールメッセージのデータモデル

電子メールのシステムやインフラの標準的なアーキテクチャは、メッセージ、アドレス、処理プロセスにおけるデータモデルが標準化されていることで成り立っている。IETF は、技術仕様などの多数の文書を発行しているが、その中には、電子メールのメッセージはどのように構造化されるべきか、それらメッセージはどのように伝送されるべきか、開発者はそのプラットフォームをどのように拡張できるか、といったことが定義されたものも含まれる。なお、成熟度合いに応じて、それら文書群の位置付けは異なっている (Bradner 1996)。長年、草案の位置付けのままであるものも多い。とはいえ、総体として見れば、事実上の標準として機能しているといえる。

メッセージ構造を規定する標準は、現在のところ、RFC 5322 である (Resnick 2008)。当該標準では、『『電子メール』のメッセージを定めた枠組みにおいて、コンピュータのユーザ間で送られるテキスト型のメッセージ構文』を規定している。電子メールのデータを保管する際に用いられる形式については、当該標準では扱われておらず、使用されたクライアントまたはサーバ側のアプリケーションに依存する。利用者側が選択できる設定といえ、例えば、サーバ上にメッセージを残すか、利用者のデバイス上にコピーを残すか、もしくは双方に残すかなどの選択がありうる。電子メールは通常、プロプライエタリな形式で保管され、バックアップやデータ出力時に MBOX、PST、EML などのフォーマットに再構築されている (hMailServer 2018; Microsoft 2005; Novell Documentation 2018; Vogel and Cazabon n.d.; Zdziarski 2008)。

第2図は、電子メールメッセージのデータモデルであり、個々のメッセージやメッセージ文字列の構成要素を描こうとしたものだ。電子メールのメッセージは、紙媒体の等価物の構造と似た形で作られている。紙の手紙のように、メッセージの中身と添付ファイルとを含むエンベロープ、正しい宛先へ配信されることを担保するためのアドレス情報から構成される。宛先と戻り先アドレスはエンベロープのアドレス情報に含まれるため、ヘッダ内のフィールドは、参照目的のみに使用される。いわゆるレターヘッド (便箋に記載する宛名) に相当する。ヘッダ内のメタデータフィールドは、エンベロープにあるものとおおむね合致するが、(郵便配達員が手紙を配達する際、実際に見るのは封筒に記載された住所であるのと同様に) 電子メールシステムがメッセージを配信する際、実際に用いられるのはエンベロープのアドレスである。エンベロープにある情報の大半は、利用者から見えない。当該メッセージが確実に配信されるために、どのくらいの量のエンベロープ情報が必要となるかは、メッセージの作成、送受信、保管を行うシステム側で決められている。RFC では、エンベロープのフィールドすべてを維持することは求められていない。だが、真正性の確度をある程度担保するためには、ヘッダの分析が有効となる。



第2図. 電子メールメッセージのデータモデル

3.2.5 メッセージの構成要素

送受信される各メッセージには、必ず含まれるべき要素と任意の要素とがある。RFC 5321 では、次の点について要件を定めている。

- **エンベロープ**：RFC 5321 のセクション 2.3.1 で定められている (Klensin 2008)。エンベロープは、発信者アドレス、1 つ以上の受信者アドレス、任意のプロトコル拡張の要素から構成される。電子メールのシステムでは、当該メッセージ自体に含まれたメタデータとして (次のヘッダについての説明を見よ)、エンベロープで用いられた情報を通常記録する。しかしこのことは、必ずしも標準で規定されていない。
- **ヘッダ**：RFC 5321 で定められている。すべての電子メールには、ヘッダフィールドが含まれる。ヘッダフィールドは、メッセージのフォーマット仕様 RFC 5322 の通りに構造化されたヘッダ名、コロン、データから構成される (Resnick 2008)。InterPARES 3 の報告書では、ID、配信、スレッド、MIME (Multipurpose Internet Mail Extensions : 多目的インターネットメール拡張) について記した箇所において、このヘッダについて詳しく解説されている。RFC 2392 で定められている通り、各メッセージには、アプリケーション側で生成される ID が付与されるのが望ましい (Levinson 1998)。メッセージ ID は、理論的にはユニークなもので、ヘッダ内の Reply-To や References フィールドで参照されている一連のメッセージをスレッドに関連づけるために用いられる。このメッセージ ID の生成法は、明確に定義されておらず、アプリケーションごとに異なるアプローチが採られている。それらメッセージ ID により、メッセージとそれらを生成したアプリケーションとを結びつけることが可能となる。仮に、対象メッセージを生成する際に用いられたアプリケーションに紐づくメッセージ ID を偽造者が挿入し損ねたとしたら、そのメッセージ ID にもとづき、偽造された電子メールを見分けることもできる。加え

て、最終的な宛先へ送信されるまで、メッセージはあるサーバから別のサーバへ次々と伝送されていくが、その過程で発生する経路情報・配信情報もまた、システムによって保存される場合もあれば、されない場合もある。郵便物の消印やパスポートの印にいくらか似て、この情報は、保存されたメッセージのうち、ヘッダの **Received** フィールドに記録されることが多い。

- メッセージ本文：RFC 2045 で定義されている (Freed and Borenstein 1996)。メッセージ本文には、メッセージの文字列、HTML、添付ファイル、埋め込みコンテンツ (画像など) が含まれる。
- 添付ファイル (MIME)：電子メールの添付ファイルは、MIME について定めた RFC 2045、RFC 2046、RFC 2049 で規定されている。電子メールに含まれる画像、音声、他の構造化データを伝送する仕組みが定められている。
- データモデル拡張：エンベロープ、ヘッダ、ヘッダフィールド、本文、本文の要素といった核となる概念は詳しく定義されているが、RFC 5321 は、拡張モデルについても明示的に説明しているため、電子メールシステムの開発者にとって、拡張に対応していないシステムに影響を与えることなく、メタデータや機能を追加する際の枠組みとなる (Klensin 2008)。したがって、ある電子メールシステムでは標準にもとづく拡張 (例えば、電子署名に対応したヘッダフィールド) に対応している場合もあれば、システムによっては、ボイスメールに対応したメタデータなど、プロプライエタリな拡張機能を含む場合もある。

一つの記録として電子メールを保存したいと望むアーキビストの立場からすると、電子メールのデータモデルがもつこの分散的な性質と拡張機能は、さらなる課題を生む。例えば、カスタマイズされたメタデータは、メッセージ自体にではなく、ローカルのユーザエージェントかメッセージストアに保管でき、また、えてしてそうなっていることが多い。対象となる電子メールをそのシステム外で収集しようとする場合、このメタデータは保存されているかもしれないし、保存されていないかもしれない。EML や MBOX などの形式により、ひとかたまりのメッセージ群として電子メールを保管するシステムはあるものの、大半のシステムでは、各商用データベースそれぞれのテーブル内にメッセージの各要素 (例：送信者、受信者、日付、件名) を保管しておき、必要が生じた場合にのみ、メッセージ群を一つのフォーマットにまとめあげる仕組みとなっている。データモデルが存在するということはまた、メッセージ内にはメタデータが含まれうるということを意味している。メタデータには、広く使用されている (必然的に、広く知られ、文書化もなされてきた) ものもあれば、めったに使用されない (文書化もなされていない可能性が高い) ものものもある。同様に、外部システムとヘッダフィールド間の関係はゆるやかなものに過ぎない。ヘッダフィールドに含まれる電子メールアドレスは、実際にはエイリアス、またはメーリングリストを指している場合もありえ、その場合、宛先となるメールボックスはないこととなる。また、表示に関するプロパティや挙動ロジックは、ローカル側で規定されることが多い。ローカルのプロパティは、オープン標準 (例えば、電子メールフィルタリング記述言語の SIEVE) に準拠している場合もあれば、カスタマイズされたり、プロプライエタリな仕様であったりする

(おそらく、その場合が多い)。とはいえ、いずれにせよ、各メッセージの表示のされ方は、ローカルのプロパティ内容に左右される。具体的にいえば、Gmail では「広告」電子メールを別タブへ機械的に振り分けており、また、アップルの電子メールアプリケーションでは、メッセージ内の情報にもとづき、「連絡先」や「予定」を識別している。さらに、本文には外部にあるデータ（例：ウェブページや画像）へのリンクが含まれうる。それらは同一画面に埋め込まれているように見えるが、実際にはただ参照して表示させているに過ぎない。

まとめると、メッセージモデルの分散的で柔軟な性質により、独立した単体の記録として切り分け、それらを収集・保存することが難しくなる。さらにいえば、込み入ったアカウント体系と伝送システムから成る環境で、メッセージが自由に流通・再生成されていくことを考え出すと、問題の複雑さはさらに増す。

3.3 アカウント

利用者が送信ボタンを押した後、別のシステム群が一連の複雑なやり取りを開始し、最終宛先または複数の宛先にメッセージを伝送していく。もっとも基本的なレベルでは、アカウントやアドレスを頼りにしてこれらやり取りは行われる。電子メールのアカウントを作成するには、プロバイダに利用者を登録し、電子メールのアドレスに紐づくメールボックスを整備せねばならない。アドレスは、標準的な構文に従い、[利用者名]@[ドメイン名]という形となる。電子メールのアカウントは、個人を表す場合もあれば、個々人が属す集団を表す場合もある。また、個人は複数の異なるアドレスをもつ場合もあれば、一つのメールボックスに紐づくエイリアスを複数持つ場合もありうる。例えば、`help@samplecompany.com` というアドレスをつくることにより、各個人のアドレスを共有しなくとも、適切なスタッフ宛にメッセージを展開することが可能となる。同様に、メーリングリストや配信用リストを用意することにより、単一のアドレスに送信したメッセージを複数のメールボックスが受信できるようにする。

電子メールアカウントの利用はプロバイダ側で管理されるが、RFC 3501 で定められている通り、ユーザ名（通常は電子メールアドレス）とパスワードを必要とする（Crispin 2003）。利用者が使うアプリケーション側でそれらパスワードなどのデータを保持する設定にしておけば、ブラウザやスマートフォンなどで逐一入力せずとも電子メールを利用することができるようになる。

3.4 データ伝送モデル

プロムが述べるように、電子メールの核は、メッセージ伝送エージェント（MTAs）とユーザエージェント（UAs）間でなされる転送・保管の技術といえる。

メッセージが配信されるには、一つ以上の MTAs と一つ以上の UAs 間でやり取りがされねばならない。電子メールのサーバ（例：Microsoft Exchange、Postfix、Sendmail、qmail、Lotus Domino）は、一つの MTA として機能する。最終的な宛先に届くまで、電子メールのメッセージをあるコンピュータから別のコンピュータへと複数の MTAs が動かしていく。指定されたアカウントで対象メッセージが受信されれば、利用者は UA を用いて

当該メッセージへアクセスする。UA には、Microsoft Outlook などのクライアントアプリケーション、ウェブベースの電子メールアプリケーション、あるいは、モバイル機器のソフトウェアを含む。UAs では、メッセージを表示・作成・管理するとともに、一つ以上のMTAs にメッセージを送ることができる。實際上、UAs は、利用者が直接制御するクライアント側のアプリケーションであり、MTAs は、UA により間接的に制御されるサーバ側のアプリケーションといえる。(2011, 9)

この転送・保管技術という特性は、アーキビストが電子メールを保存する際に影響を及ぼす。電子メールは、ほぼ絶えず遷移中の状態にあるものと捉えるのが望ましい。エンコード・パッケージ化され、デコードされ、開封され、分解され、保管され、再構成され、出力されていく。しかし、保存対象となる記録をどのように考えればよいのか。作成され、送信された電子メールが対象となるのか。送信者のアプリケーション内にある送信フォルダに保管されたものが対象となるのか。受信者側の電子メールサーバに残されたものか。ダウンロードされたものか。PST 形式でローカルに保管されたものとなるか。あるいは、iPhone 上に残されたものか。

アーキビストができるのは、ある時点で存在していたメッセージを、また、ある目的で処理されたメッセージを、収集・保管することである。どのような種類のメッセージを保存するか意思決定せねばならないが、それは重要な意味合いをもつ。

今のところ、アーキビストの立場からすると、現用期間終了後にサーバや機器に残された電子メールのメッセージを収集することが多かった¹⁴。一方、コンプライアンス対応用の商用電子メールツールでは、送信時または受信時のメッセージとエンベロープのすべてを収集している。双方のアプローチとも、強みと弱みを有す。「終末期」収集により、対象となる機器やシステムに残されたメッセージについて取られた行為の記録を保存することが可能となる（例えば、メッセージ開封）。ジャーナリング機能では、電子メールシステムとは別の場所にデータの書き込みを行う際、電子メールデータを複製し、削除や修正がなされないようにするが、これにより、利用者が削除したメッセージを含め、作業履歴の完全な記録を残すことができ、その履歴を後から辿ることも可能となる。

BITS セキュリティ ワーキンググループは 2013 年に次の通り報告している。「フィッシング攻撃が広まったことにより、消費者の電子メールに対する信頼性が低下した。米国の非営利団体である ID 窃盗資源センター (Identity Theft Resource Center) の 2010 年時調査結果によれば、消費者の 81% が、オンライン上でのやり取りの際、個人的・金銭的な情報のセキュリティに関した懸念事項として、電子メールのフィッシング詐欺を挙げていたという。マルウェアの多くは、電子メールを介して蔓延したと見なされている。2011 年の調査によると、PC がマルウェアに感染したと回答した数は、米国の成人のうち 5,400 万人にのぼるという。ガートナー社は、米国の消費者のうち 40% 以上が、継続的に脅威さらされてきたことにより、電子メールやオンラインショッピングに対する信頼性に疑問を抱いている、と推定している。」(BITS Security Program 2013, 8)

¹⁴ このアプローチが意味するところを分析した興味深い例として、参考文献 (Bearman 2017) を参照されたい。

3.5 電子メールの脆弱性

電子メールが設計された初期時点において、セキュリティは重要な関心事ではなかった。のちに、この開放的で柔軟なアーキテクチャに起因して、様々なセキュリティ問題が生じていき、多くのアプローチが提唱されてきた。だが、あらゆる問題に適用可能な解決策はなく、逆にいえば、対応策は進化し続けてきたといえる。

電子メールは、人間によるものにせよ機械によるものにせよ、実質的に個々のメッセージのあらゆる側面（本文、ヘッダフィールド）が修正可能な設計となっている。結果として、乱用、誤用、あるいはエラーへの耐性が低い。次に例を挙げておこう。

- ・ 未承諾メール、またはスパム
- ・ 悪意ある内容のもの：機械を標的としたもの（マルウェア、ウィルス）、人を標的としたもの（フィッシング、利用者の認証データを盗もうとする行為）
- ・ 偽造：電子メール（本文、ヘッダフィールド）の真正性は多くの手段で危険にさらされる。実際の送信者からでなく、なりますました他者から電子メールが送られたように見せかける行為、いわゆるスプーフィングは、その最たる例である。
- ・ 窃盗：電子メールを用いた場合、公開されたインターネット上でテキストのデータが送られるため、そのメッセージやメタデータは、アクセス権限のない部外者に見られる可能性がある。何らかの対策が取られていない限り、通信経路の過程で、個人的または私的な情報は容易に盗まれてしまう。
- ・ 完全性：正当な目的からも電子メールは改変されうるため、（意図的にせよ、意図的でないにせよ）電子メールが改変される、または破損してしまう可能性は多々想定できる。
- ・ 否認（または拒絶）：前述した脆弱性を前提として、メッセージが改ざんされたと、あるいは、メッセージを送信していないと主張するものがいても、ありうる話として捉えられかねない。

これらの基本的な脆弱性が複数突かれる場合もありうる。例えばフィッシングは、悪意のあるものが信頼されている主体になりすまし、機微情報を奪おうとする試みである。これは明らかに真正性の侵害であるが、手口が洗練されればされるほど、同時に他のいくつかの側面を侵害する行為ともなりうる。

3.6 ASCII メッセージ以外の付加要素

ここまで述べてきたメッセージ構造は、電子メールの全体システムの核となる部分となっているが、個々のメッセージ、スレッド、アカウントに補足的な要素が加えられ、複雑さは増していく。いいかえれば、デジタルオブジェクトの集合体にそれら付加要素が足されていくことで、新たな保存問題が生じる。二次的な資料、例えば添付データや外部資源のリンクは、メッセージ自体と同じ程度に価値をもつかもしいない。とはいえ、それらは失われやすく、あるいは、元の文脈と切り離されてしまいやすい。

3.6.1 添付データ

元々、1982年に定められた RFC 822 において、電子メールのメッセージは、ASCII コー

ドのテキストでエンコードされなければならない、各メッセージの長さで行数には制限が設けられていた。だがすぐにユーザらは、音声ファイル、画像、文書を含めたがり、やがて電子メールは、そのようなバイナリのオブジェクトを包含する進化を遂げていった。伝送・配信手段を変更することなく、この目的を達成するため、仕様上、バイナリファイルはメッセージ本体内に MIME の形式で含められなければならないことが定められた¹⁵。ASCII テキストに限定されていた制約を克服するため、MIME 規格において、バイナリファイルは Base64 でエンコードされ、伝送時に非文字データは ASCII テキストへ変換され、MIME タイプが付与され、それにより、受信時、適切にデコードされ、利用できるようになることが規定されている。Base64 という名称から分かる通り、このエンコーディング方式では、ASCII のうち、64 の文字記号が使用される。元のバイナリデータを 6 ビットずつに分解してエンコードし、受信時に、デコードされたものが利用者へ表示されることになる。EML など、保管用のフォーマットでは、Base64 でエンコードされた形で添付ファイルを格納しているものもある。

その後、デジタル形式のボイスメールやファクスなども MIME で扱うことができるように、RFC は拡張・修正されていった。

- ・ ボイスメールの通信手段としての電子メール：RFC 3801 (G. M. Vaudreuil and Parsons 2004)。次も参照：RFC 3773、RFC 4239、RFC 6381、RFC 4393。
- ・ その他：
 - ・ RFC 4142：インターネットを用いたファクスメッセージ伝送
 - ・ RFC 1767：電子データ交換 (EDI) 用オブジェクトのカプセル化

電子メールの添付ファイルは、保存に関する問題を多数生じさせる。他のバイナリデータを保存することと同じ課題があるということを述べるつもりはない。添付されるファイル形式が無数にありうるというだけでなく、そのデータを保存し、電子メールのメッセージとの関係性などをどのように維持していけばよいのかといった課題もあろう。さらに、囲み記事にもある通り、有害なコンテンツが添付内に隠され、電子メールを介して広くばら撒かれる可能性も指摘できる。

システム上、投入・出力できるという点から、添付データも収集したいとアーキビストは望むであろう。だが、問題は複雑である。なぜなら、電子メールは異なるクライアント、異なるサーバで生成・処理されているためだ。よりいえば、RFC では、内部処理の仕組み、保管用フォーマット、エクスポート／インポート機能について規定していないためである。各システムでは、添付ファイルをそれぞれのやり方で処理しており、それがプロプライエタリな場合もある。クライアントアプリケーションの中には、メッセージ自体に MIME 形式で添付ファイルを埋め込んでいるものもあれば、ネイティブ形式のバイナリフォーマットとして別に保管しているアプリケーションもあろう。後者の場合、メッセージ内にポインタが埋め込まれているものもあれば、プロプライエタリなデータベース内にポインタが置かれ

¹⁵ 複雑な標準である MIME は複数の RFC で規定されており、それぞれ数回更新されてきた：
<https://tools.ietf.org/html/rfc2045>、[rfc2046](https://tools.ietf.org/html/rfc2046)、[rfc2047](https://tools.ietf.org/html/rfc2047)、[rfc2048](https://tools.ietf.org/html/rfc2048) (訳注：現在は、[rfc4288](https://tools.ietf.org/html/rfc4288)、[rfc4289](https://tools.ietf.org/html/rfc4289) と思われる)、[rfc2049](https://tools.ietf.org/html/rfc2049)。

ているものもあろう。

American Lands Alliance (訳注: 森林政策のロビー団体) のコレクションには電子メールが含まれており、米国議会図書館のアーキビストは、その整理を行った際、*Secret.zip* といった名称の添付ファイルを多数発見したという。初期の選別評価時、それらメッセージと添付ファイルは脇に置かれ、後日処理することとされた。それら圧縮ファイルの中身をプレビューできなかったためである。後日、それら添付ファイルには、ウィルスが含まれていることが確認された。この体験談は、Zip ファイルや圧縮ファイルには、未知の内容から成る隠し機能が含まれている可能性もあり、注意して開くのが望ましいということを示すものである。

3.6.2 リンク先・メッセージ外の資源

前述した添付ファイルが課題であるなら、リンク先資源もまた問題だ。ここでのリンク先資源とは、ハイパーリンク先にあるコンテンツ、電子メールのサーバやクライアントの環境外で保管されているコンテンツのことをいう。実際のところ、メッセージ以外の領域を扱うとなると、これはまた別の調査が必要となってしまう。その調査を今後誰が行うにせよ、本項での整理が議論の出発点として活用されることを願う。

脚注のような役割を担うリンクもある。その場合、所在（典型的には URL）が提示され、指し示す資源のタイトル名が示されることもある。この情報をもとに、ハイパーリンクをクリックして、参照された資源を読み手は見つけることができるかもしれない。だが、電子メールのクライアントアプリケーションまたはサーバは、リンク先のコンテンツを再作成する訳ではなく、あるいは、それらにアクセスできることを保証してもいない。この脚注という例えは不十分であろう。ネットワーク上にある資源は、メッセージを受信後すぐに示された資源へアクセスできる場合は例外であるが、旧来の脚注で参照された資源よりも、極めて不安定なものである。参照先が物理的な本であれば、この世に一つしかないものでもなく、（正しい版が入手できるとして）内容も変わっていないであろう。参照する必要がある、通常はその複製を見つけられるだろう。

他の外部コンテンツは、脚注というよりも、手紙に入れられた写真に近い。いいかえれば、添付書類のようなものであるが、大きな違いが一つある。そのコンテンツは、電子メールシステムを介して伝送されてくるのではなく、あくまで参照先として位置付けられるものに過ぎない。写真ギャラリーへ、あるいは Google ドキュメントへリンクが貼られている場合、様々な理由から、それら資源へたどり着けない場合もあろう。リンクが壊れているかもしれないし（外部資源への確かなポインタが存在しなくなる状態）、リンク先の資料をアーキビスト側では利用できない可能性もある（セキュアな環境、または遠隔地に保管されているなど）。リンク先資源が見つけれられたとしても、囲み記事で記されているように、それは一時的なものでしかない場合もある。文書や画像などのオンライン上のコンテンツは同じ性質をもつ。そのため、保存目的で電子メールを収集する場合、リンク先資源の真正性や証拠的価値についてどのように考えるべきか、課題が生じるといえよう。

2017 年 4 月、(訳注: ニューヨークの動物園にいる) エイプリルというキリンが出産するまでのライブ

動画配信が注目を集め、1400 万人に視聴されたという (Spangler 2017)。この麒麟の様子を配信したページへのリンクは、誕生の瞬間を共有すべく、電子メール、Facebook、チャット、インスタントメッセージなど、インターネット上で広く拡散された (Animal Adventure Park 2018)。現在、当該リンク先では、広く視聴された誕生時の様子でなく、現在の様子がライブ動画として配信されている。

3.6.3 署名欄

名目上、メッセージ本文の一要素と位置付けられているが、電子メールの署名欄には、送信者に関する基本的な、だが構造化されていないデータを含んでいることが多い。送信者の所属機関、連絡先、ソーシャルメディアや他のネットワーク資源へのリンクが具体的には挙げられ、あるいは、ロゴや画像が含まれる場合もある。こういった理由から、署名欄は、名称や関連性を探る上で宝の山となりえ、文脈情報を提供するものともなりうる。仮に、構造化された形で署名欄のデータを集めることができれば、典拠レコードの URI と紐づけていくことも可能となるかもしれない。裏返すと、署名欄の繰り返しデータは、自動分類の品質に影響を及ぼしかねず、今後の課題となりうる。

電子メールは、比較的単純な構造化テキストから複雑な形式のものに発展を遂げ、それが新たな保存問題を生んできた。電子メールを保存し、利用に供するという目的をもつ組織の場合、そのための方針やシステムを形作っていく上で、メッセージの技術的な性質を理解すること（具体的には次のような事項を含む。それらがどのようにフォーマットされ、アカウント間で伝送されているか。電子メールを規定する多数の標準、電子メールのアプリケーションがメッセージを管理するやり方など）は重要な作業となる。

4. 現行のサービスと動向

文化遺産コミュニティが電子メールを収集・整理・保存し、利用に供する上で、電子メールを取り巻く人と機械の関係性は、そのあり方に影響を及ぼす。電子メール保存の主な課題については、これまでも文化遺産コミュニティにおいて部分的には取り組まれてきた。だが、まだ整理されていない課題は多々あり、アーキビストやデジタル保存専門家は、それらすべてに対応していく必要がある。まずは現状把握を始める段階にあるが、電子メールサービスの動向を確認・評価するとともに、広く IT 業界、そして社会の流れについても概観していきたい。

4.1 進化する電子メールの生態系

電子メールは広く普及し、重要なものと位置付けられるようになった。IT 企業やサービス事業者からすると、電子メールを生成・管理・保存・処理するための市場が形成されてきたといえる。電子メールの生成・流通・消費、他のコミュニケーションサービスとの統合、法令遵守対応など、多様な話題があるにせよ、いまや数十億ドル規模の IT 業界が提供する電子メールシステムを多くの組織や個人が利用している。IT 企業が構築したプラットフォームは、電子メールの開放的で柔軟なアーキテクチャを活用したものと見なせるが、一方で、人類の過去の記録として電子メールを保存する責務を負う文化遺産コミュニティからすると、これは好ましい方向であると同時に、問題を複雑化させる遠因とも捉えられる。

4.1.1 不正使用、不正防止、セキュリティ、到達可能性

電子メールは、その初期段階から、開放的で分散的なアーキテクチャであったが、それゆえ、ヘッダの改ざん、スパム、スプーフイング、フィッシングなど、不正使用の脅威にさらされてきた。やがて電子メールが普及するにつれ、悪意ある利用を阻止するセキュリティ技術や標準が新たにつくられてきた。だがこれはまた別の問題を生み出し、正当な電子メールまでもが配信されない場合が生じてしまった。電子メールの不正使用や犯罪は、それを阻止するセキュリティ手段を飛び越して発展してきている。新たなセキュリティ手段が導入されるたびに、電子メールの到達可能性を向上させるための取り組みが必要となり、その技術が開発される状況となっている。

最大限の成果を生み出すため、電子メールのセキュリティ対策は広く全体で受け入れられねばならない¹⁶。ステークホルダーと産業界は、共通の目標を達成するため、利害を調整または最小化し、共同して標準の策定に取り組んでいる。そのような活動を行なっている団体は多数挙げられる（例：Messaging, Malware and Mobile Anti-Abuse Working Group (M3AAWG)、Email Sender and Provider Coalition、Online Trust Alliance、CAUCE、その他のアドボカシー団体）¹⁷。重要な標準のいくつかは、専門団体により支えられており、例えば、Domain Message Authentication Reporting & Conformance (DMARC) が挙げられ、これは、偽

¹⁶ 送信ドメイン認証技術など、多数の取り組みがなされてきたが、広く普及せず失敗に終わっている (Leiba 2013)。

¹⁷ <https://www.m3aawg.org/>; <http://www.esppcoalition.org/>; <https://otalliance.org/>; <http://www.cauce.org/about.html>.

装ドメインでのなりすましメールを防ぐことを目的としたものだ (dmarc.org 2018)。多数の組織が様々な情報を公開している。それらは、電子メールのセキュリティ技術や運用方法、標準の使用や最良事例、採用率や導入評価について概観する上で役立つ¹⁸。

電子メールの不正利用を防止するには、少なくとも負の副次効果を許容せざるをえない。それは、正当な電子メールの到達可能性を低下させてしまうという側面だ。ある調査報告によると、スパム扱いされるなどした結果、(企業体から消費者へ向けられた) 商用電子メールのうち、宛先で受信されたものは 79%に過ぎなかったという (Return Path 2015)。電子メールのサービス提供事業者からすれば、そのサービスを利用しているものが正当な電子メールを送信していることが保証されねばならない。そのため、多くのサービス提供事業者は、不正防止を推進する団体へ積極的に関わっている。

電子メールのサービスを提供するコミュニティにおいては、セキュリティ問題に対応するため、次の手段をこれまで採用してきた。

- ・ 暗号化: 暗号技術により、情報を、暗号鍵でないと解読できないコードに変換すること。暗号化では、悪意ある第三者がデータを収集する行為までは防げないが、必要な鍵がなければ、データを解読または利用することはできない。
- ・ 電子署名: 暗号化と似た別の暗号技術であり、公開鍵が用いられる。署名は、署名者だけが使用できる秘密鍵とともに生成される。当該署名は、公開鍵を用いて復号化できる。電子署名により、署名者の認証を確実に行うことができる。
- ・ 通知機能: 電子メールシステムの各アクター間で何らかの通知がなされることであり、多くのセキュリティ機能で採用されている。例えば、利用者側があるメッセージをスパムと判定した通知を送信できる機能が挙げられる。この場合、送信元 (または当該ドメインすべて) はブラックリストに載せられ、以降のメッセージはブロックされる。
- ・ 内容分析とフィルタリング: 電子メールの内容を分析する技術は多数あり、そのトピックや真正性を判定可能とする。ウィルススキャン、(例えば「ぼろ儲け」などの) キーワードでのフィルタリング、機械学習の適用なども含まれる。

産業界から生まれたセキュリティ機能のなかには、文化遺産コミュニティにとって価値あるものもある。

- ・ エンドポイントでの保護: サイバーセキュリティ市場は巨大であるが、セキュアな電子メールゲートウェイやエンドポイントでの保護プラットフォームもその中に含まれ、スパムや詐欺メールの削除など、電子メールに関連した機能も挙げられる (Firstbrook and Wynne 2015; Morgan 2015)。このようなサービスの大半は、リアルタイムで検知・保護するものだが、機能の中には、電子メールのアーキビストにとって有用なものもある。例えば、ブラックリスト (電子メールサーバで遮断するドメイン) とホワイトリスト (認定/認可されたドメイン) が挙げられる。また、これらサービスにより、電子メールメッセージの位置付けを立証する、あるいは反証する上での有用な証拠がえられる。例えば、ブラックリストに含まれるドメインから発信されたものは、信頼性が低

¹⁸ 例として次を参照。 <https://www.m3aawg.org/supporting-documents>; <https://www.m3aawg.org/for-the-industry/published-comments>.

いと判断できるだろう。とはいえ、このブラックリストまたはホワイトリストを保存する先行事例、ましてや、この継続的に更新されていく修正履歴を残すような先行事例はおそらくないと思われる（本タスクフォースのメンバー内で見聞きしたものはいない）。

- アドレス検査：電子メールのセキュリティを扱う企業では、対象顧客の評判に悪影響がでないよう、電子メールのアドレスリストを検証・整理するサービスも提供されている（Informatica 2018; Never Bounce 2018）。これらのサービスでは、いかがわしいアドレス、例えばスパムメールや詐欺メールを以前に送信したと見られるアカウントが特定され、それにより、悪意ある第三者と不要に接触する（つまりは攻撃に晒される）危険性を回避できる。そのようなアカウントのリストを保存しておけば、文化遺産コミュニティにとって役立つかもしれない。大規模な電子メールアーカイブズのうち、いかがわしい送信者を特定するため、ブラックリストと同様、それらアカウントのリストが事後に活用される場合もありうる。繰り返すが、こういったリストを保存する取り組みは現在のところないと思われる。まして、そのようなリストが利用者に提供されたり、システムの一機能として組み込まれたりしている例もない。
- 電子署名：認証技術は、迷惑メールの動きを一步先んじようと開発されてきた。別の報告書（Task Force on Technical Approaches for Email Archives 2018b）では、重要なセキュリティプロトコルのいくつかについて解説している。具体的には、送信ドメイン認証技術である SPF（Sender Policy Framework）、DKIM（Domain Keys Identified Mail）、DMARC（Domain-based Message Authentication, Reporting, and Conformance）であり、いずれも RFC となっている。SPF と DKIM が広く普及し、DMARC も実装されれば、アーキビストや利用者からすると、対象メッセージの真正性に関する信頼性が高まりうる。あるいは、見方を変え、基礎・応用研究が進めば、そこから知見がえられる可能性もあるだろう。例えば、エンベロープの情報、特に電子署名を各サーバが保持しているか、どのように保持しているか、将来的にそれら署名を検証する上で十分な情報が残されているか、といったことを理解していくに当たり、エンベロープに関する調査が役立つこととなろう。ヘッダフィールドに含まれる履歴を残しておくことが十分にできるようになれば、おそらく、アーキビストや利用者は、真正性を検証する上で、DMARC やエンベロープ記録の保持機能を活用することができるようになるだろう。

4.1.2 マーケティングと電子商取引サービス

電子メールの多くは、領収書の送付や注文の更新など、マーケティングや商取引目的で作成されている。マーケティングや小売り目的で使われる技術やツールは、モバイル機器の普及、消費者行動の変容、新製品の登場などを要因として、進化し続けており、結果、大量の電子メールが不運な利用者の受信箱に送信される事態となってきた。そのような記録に関心のない文化遺産機関が大半であろうが、逆に、それらに証拠的価値を見出す機関もあるかもしれない。とはいえ、これらの基盤を成す技術自体は、ニュースレターや他の情報を配信する際にも用いられており、この意味ではアーカイブズ関係者にとっても興味深い側面といえよう¹⁹。マーケティング、顧客サポートシステム、商取引向けアプリケーションで生

¹⁹ 電子メールによるマーケティング技術や関連サービスを提供する企業について分析した報告書も興味深

成された電子メールを保存する上で考慮すべき点について見ておこう。

- ・ 利用者の電子メール環境に適したコンテンツを生成するレスポンシブデザイン技術にもとづいて電子メールがつくられている。この場合、画面の小さいスマートフォン上での表示のされ方と、デスクトップ画面上での表示のされ方とは異なる（Email Design Reference 2018）。
- ・ 所在場所、時間、ソーシャルメディア上でのやり取りなどをリアルタイムに反映した動的なコンテンツを提供している。例えば、電子メールに含まれたニュース、天気、ソーシャルメディアのフィードデータが、確認する度に更新されて表示されるなど²⁰。
- ・ 動画メールは、また新たな技術的課題をもたらす。大半のアプリケーション、クライアント環境では動画コンテンツを処理でき、また、動画をメールに埋め込み可能なマーケティング向けシステムも増えている²¹。

この種のコンテンツを保存する上での課題は、動的コンテンツを保存する場合と似ている。このようなシステムがどのようなものかを理解する上で、アーキビストにも役立つ情報が企業から提供されてもいる。例えば、電子メールによるマーケティングの手引きを公開している企業もあれば²²、あるいは、顧客シェア（利用者の機器または電子メールアプリケーションの種類別）などの指標を測る手法、到達可能性（対象者が確かに受信したかどうか）、クリック率（あるウェブサイトへのリンクをクリックしたかどうか）に関する情報を出している企業もある²³。

4.1.3 一般向け電子メールサービス

Gmail、Outlook、Yahoo!メールなどは、個人向け、または小企業向けの電子メールサービスである。オンライン上での消費活動が増すにつれ、これらのサービスも利用が増加している。依然として、電子メールが使用される機会も多いためだ（Radicati Group, Inc. 2016）。核となる機能や標準はさほど変わっていないものの、電子メールのサービスは進化し続けている。

- ・ ストレージ：電子メール使用領域の容量上限は上がり続け、多くのアカウントが実体なきアーカイブズになりつつあるが、保存という点では疑わしい。2004年にGmailが開始された時、容量上限は1GBだった。これは、Yahoo!メールの無料版の250倍、マイクロソフトのホットメールサービスの500倍の規模であった（Pogue 2004）。本報告書執筆時点で、Gmailの無料版で提供される容量上限は15GB、米Yahoo!メールの上限は1TBにのぼる²⁴。

い。ガートナー社からは、有償だが、電子メールマーケティング市場に関する年次報告書が出されている。大手企業による報告書は有料のことが多い（J. Cohen 2015; Hopkins and Sarner 2015; SparkPost 2015）。

²⁰ <https://www.marketingcloud.com/blog/real-time-email-examples/>; <http://www.realtime.email/rte-resource/exploring-the-benefits-of-realtime-email-white-paper/>.

²¹ <https://www.campaignmonitor.com/resources/guides/video-in-email/>.

²² <https://www.campaignmonitor.com/dev-resources/>; <http://templates.mailchimp.com/>; [https://litmus.com/resources for examples](https://litmus.com/resources-for-examples).

²³ <https://mailchimp.com/resources/research/email-marketing-benchmarks/>

²⁴ 参考文献（Wikipedia 2017b）と次を参照されたい。<https://overview.mail.yahoo.com/>.

- ・ 整理と検索：2010年、Gmailでは、フォルダ内の電子メールを整理する従来の手法に代えて、ラベル機能を導入した。受信箱内の電子メールにラベルを付与すれば、検索性能を大幅に向上させられる（Rodden and Leggett 2010）。類似の機能は、自動分類、電子メールのソート機能などを含め、他のサービスでも提供されていることが多い。なお、従来のフォルダ機能はまだ残っている（Casserly 2017）。

このような機能向上は、個人による電子メールの管理方法、アーキビストによる処理方法にも影響を及ぼす。電子メールのコレクション規模は増加し続けており、このような大規模コレクション内を目視で辿って探すには、大半の表示形式が準拠する標準的なメタデータでは不十分である。ナビゲーションを可能とするメタデータは、新たな発見の手立てとなろう。裏返せば、研究開発が必要な領域といえる。

4.1.4 企業向け電子メールサービスと運用

企業向けの電子メール製品・サービスとは、所属社員・職員に電子メール機能を提供したい組織用のものである。オンプレミスで構築される場合もあれば、クラウド上で実現される場合もある。一般向け電子メールサービスと同じ傾向が見られるものの、ベンダーは、リテンション管理、法令遵守対応、e-ディスカバリ（電子情報開示）機能を組み入れていくことで、自身らの企業向け製品・サービスを拡大させ、市場の需要に応えようとしてきた。電子メールのアーカイブサービスと位置付けられることも多い。ガートナー社の推計によれば、2016年時点において、法令遵守対応から電子メールを保存している組織のうち10%が企業向けの電子メール製品・サービスを使用しており、この数値は2021年までに35%に上がるであろうという（Dayley et al. 2016）²⁵。電子メールの保存に長期的な価値を見出す組織にとっては、関連する機能が主な電子メールサービスで既に提供され、または組み込まれていれば、追加のコストをかけることなく、ただその機能を流用することができよう。とはいえ、そのようなサービスにおいては、長期的なデータ保持ではなく、短期的な法令遵守対応が主たる狙いであるため、別途研究開発すべき課題があろう。

長期的に電子メールを保存する取り組みの一環として、こういった企業向けサービスを採用する場合、組織内のIT部門やサービス提供者の支援・協力が欠かせない。だがまずは、何を確認すべきか知ることから始めるべきだろう。政府のアーキビスト、記録管理者、IT専門家にとっては、既存のジャーナリングソフトウェアを評価し、長期的な保存に必要な追加機能の要件を開発業社と固めていくことが求められているかもしれない。

4.1.5 電子メールのストレージ、法令遵守、記録管理

電子メールのシステム外で電子メールのデータを管理・保管することに対し、市場の需要は無視できないほどにある。電子メールの保管に絞った営利サービスがある一方、多くの記録管理システムでは、他の電子記録とあわせて電子メールを扱う。いずれにせよ、短期的ではなく、長期的に電子メールを保存することに重きを置く組織にとって、外部のストレージシ

²⁵ ガートナー社では、マイクロソフトの電子メールソフトウェアのアーカイブ機能や電子情報開示機能を評価した詳細な報告書も出している（Landers, Harris, and Zhang 2017）。

システムを使うことには、利点もあれば欠点もある

ビジネス上の需要の高まりのほか、規制遵守に取り組むつつ、技術進化に継続的な対応が求められているけれども、この入り組んだ状況が曖昧な用語を生んできたともいえる。画像情報マネジメント協会 (AIIM) は、電子記録管理 (ERM)、ドキュメント管理 (DM)、エンタープライズ コンテンツ管理 (ECM)、あるいはその他の関連用語をそれぞれ切り分けて理解する上で有用な用語集を作成している (AIIM 2018)。こういった概念は、金融などの規制された業界においては、「法令遵守」の名のもとで議論されることが多い (Lin 2016)。ベンダー側も、「記録管理」や「コンテンツ管理」といった機能を説明する際、「法令遵守」の語を使う場合が多々見られる (Microsoft 2011; Mimecast 2018)。意味的な違いは置いておくとして、保存コミュニティにとって特に関連深いのは次の側面だ。

ストレージ管理は、初期のアーカイビングシステムにとって第一の課題であった。(他の種類のデータについてはいうまでもなく) 電子メールの受信箱が肥大化するにつれ、電子メールの一部を予備の (より安価な) システムへ移行させることで、ストレージのコストを下げようと組織は試みてきた。だが逆に、検索速度は落ちてしまう。ガートナー社の「企業向け情報アーカイビング」に関する報告書によれば、ストレージ管理機能は、バックアップやリカバリ機能とあわせて、大半のベンダーが提供し始めたという (Dayley et al. 2016)。

リテンション管理とは、ある記録を確実に保管しておく最短期間を設定することである。これは通常、記録の分類作業に基づきなされ、あらかじめ定められた期間またはスケジュールに沿って記録を扱うことが合意される。そして、削除されるまでの一定期間、当該記録を保管しておくことが保証されることになる。いいかえると、記録管理システムでは、記録が保管される最短期間を指して、この「リテンション」という語が用いられている。

残念なことに、電子メールのシステムでは、まったく逆の意味合いで、この「リテンション」という語が定義されているように思える。つまり、システムから自動的に削除されるまでの、記録が保管される最長期間のことを指すものとして、当該用語は用いられてきた。例えば、マイクロソフトの Office 365 (他のアプリケーションに加えて、電子メールも含まれる) では、電子メールの「リテンション」に対応している。だがこの意味するところは、システムが自動的に電子メールを削除する、あるいは、アーカイブへ電子メールを移動させるまでの、電子メールが保管される最長期間のことを指している。この最長期間が終わるまで、利用者は手動で電子メールを削除することができない。利用者用画面で単に見えなくなるだけで、データは残されたままとなる (Palarchio 2015)。Outlook 365 上で、電子メールを残すのは最短期間とし、かつ削除されないようにするためには、「ホールド」機能を使わなければならない。

ホールド機能 (インプレース保持や訴訟ホールドともいう) とは、保管すべき記録や電子メールを管理するためのものである (Microsoft Exchange Online 2017)。ある特定期間後にデータを削除する設定が可能だが、実際には設定されないことが多く、電子メールのメッセージは、訴訟が終了してから (あるいは、訴訟の恐れが解消されてから) 削除される。

ジャーナリングは、組織であらかじめ定められたルールに従い、受信または送信メールを複製し、それを保持する機能のことである。通常、保持された複製は、電子メールシステムの外で保管され、利用者が削除したり改変したりすることはできない (Microsoft Exchange

Online 2016)。別の形態として、エンベロープ ジャーナリングと呼ぶ機能もあり、これは、メッセージ自体に加えて、そのエンベロープを保存するものだ。これには、Bcc（ブラインド カーボンコピー）の宛先や配信リストなどの情報も含まれる。

監査ログまたはアクティビティログは、各メールボックスへの不信なアクセス、システム管理者（あるいは全利用者）の作業履歴を追跡するものだ。アーカイビングのソフトウェアベンダーの中には、データが改変されないようにするため、このアプローチを採用しているところもある。監査ログは、データのセキュリティを担保するための手段の一つと見なされることも多い。

権利管理（デジタルライツ管理とも呼ばれる）の機能は、誰が機微データへアクセス・転送・印刷・コピーできるかを、電子メールソフトウェアか第三者のプラットフォーム上で管理するためのものである（Kekre 2015; Microsoft Developer Network 2011）。

アーキビストにとって、これら機能は有用なものとなろう。NARA は、電子メールの管理手引きにおいて、ジャーナリングやクローリングについて解説している（NARA 2013b）。ジャーナリング機能により、全メッセージや予定を送受信時に収集できる。クローリング機能は、システム管理者が事前に定めた記録通知ルールを適用し、対象データを定期的に収集するものである。クロールされたデータはアカウント保有者でも見られるが、ジャーナリング機能で集められたデータは見られない。フィッツジェラルトは、デジタル形式の記録を特定・抽出し、デジタルリポジトリへ投入することを目的とし、アーカイビングソフトウェアがどのように利用できるかを論じている（Fitzgerald 2013）。なお、電子メールの出所来歴を記録しておく上で、監査ログは極めて有効な手段となりえようが、先行事例までは確認できていない。こういったツールが保存業務にどのように役立てられるかは、さらに調査研究を進めねばならない。

また、上述した機能には課題もいくつか含まれる。外部システムに保管された電子メールは、最適なフォーマットでは保管されていないかもしれない、あるいは、メタデータも十分整備されていないかもしれない。メッセージを確認するには、使用説明書を読まねばいけない、あるいは、アクセス権を設定せねばならないかもしれない。デジタルライツ管理システムでは、ある種類のコンテンツ（電子メールの添付ファイルなど）に制限がかけられるかもしれないが、当該システムでないと設定変更できず、アーカイブズ職員では扱えないという状況もありえよう。結局現状の権利管理システムにおいては、電子メールは歴史的記録の一部として保存されるというより、リスク管理の一対象物（削除されうる対象）として捉えられている。

4.1.6 法令遵守と法的ツール

電子メールは、法的手続き上、重要な記録となり、司法の場で証拠として扱われうる。加えて、電子メールを含む電子的に保管された情報については、その法的有効性に関する判例もいくつか確認できる（Pratt, n.d.; Wikipedia 2016）。この 10 年で、e-ディスカバリ（電子情報開示）関連のサービスや技術の市場は拡大を見せており、その規模は年間 18 億米ドルにのぼるとの推計もある（Zhang, Logan, and Landers 2014）。e-ディスカバリシステムでは、電子メールの長期保存に関連した技術的課題も扱われ、具体的には、大規模メッセージストア

の管理、法廷に提出すべき電子メールの特定、機微情報の保護、証拠的価値のある電子メールの保存が挙げられる。

多くの電子情報開示システムで、(証拠としてのソーシャルメディア メッセージなどを含め) 送受信時の電子メール収集を担保するため、ジャーナリング技術が実装されている。こういったソフトウェアにより、ファイルやメタデータの完全性が保証され、管理過程が記録されていく。e-ディスカバリシステムは、文化遺産の保存目的というより、法的義務への一対応手段として広く普及している。

電子メールの証拠的価値は、法務上、重要なものとなっているため、ジャーナリグ機能で集められたデータ、現用の電子メールサーバ、クライアント PC、ハードドライブ上のデータを含め、集積された電子メールをデータマイニングするツールも出回っている。PinPoint Lab 社の Harvester のようなツールを用いると、多様な電子メールサーバに接続し、ハードドライブ上のデータをスキャンし、元ファイルのメタデータを変更することなくデータを複製するとともに、ハッシュ値でデータの完全性を検証できるという (Pinpoint Labs 2018)。また、コレクション構築を支援し、データ管理の作業過程を自動記録してくれるともいう。

e-ディスカバリシステムには、文章内のある情報を人間が見つけ出せるよう、テキストマイニングや解析技術が組み込まれている。このようなアプリケーションは、機械と人間がフィードバックし合うことから、機械支援型レビュー (technology-assisted review)、あるいは縮めて TAR と呼ばれている²⁶。多くの企業でテキストマイニング技術が利用されているけれども、法曹・法務業界が求める品質は高い。裁判で正当性を主張せねばならないためだ。TAR の目的は、過去の事象に関し、証拠を挙げて主張の正しさを論じたい検察官や弁護士に役立つ記録を探し出すことであろう (Attfield and Chapin 2018)。

別の手法は目視で文書を確認することだが、デジタル文書の量が膨れ上がった今、あまりに負荷が高く、時間も足りない。したがって、法曹・法務業界としては、技術の効果的な活用を模索しているといえ、実際、学習・検証用に、電子メールの大量データセットを提供し始めている (The Coalition of Technology Resources for Lawyers 2016; Duke Law Center for Judicial Studies 2018)。e-ディスカバリソフトウェアを先導する企業の中には、電子メールのコーパスから機微データや個人を特定可能な情報を識別・削除するため、前述したような解析技術を適用した事例を公開したところさえある (Nuix 2018)。またガートナー社では、e-ディスカバリや TAR 技術の市場・ベンダー動向をまとめた年次報告書を出している (Zhang, Logan, and Landers 2014)。

e-ディスカバリ技術に今のところ確たるものはなく、研究・議論の蓄積が求められている (Grossman and Cormack 2014)。とはいえ、これまでの調査によれば、機微情報、機密情報、部外秘情報を目視で確認する行為は、間違いを起しやすいたことが示されてきた。TAR や自動分類による作業結果の方がよかったとする調査結果も出されている (Cormack and Grossman 2017; Grossman and Cormack 2011)。機械学習アルゴリズムがもたらす成果はまだ明確でないかもしれない。このことからしても、機微情報や個人情報特定の上で、保存

²⁶ キーワード検索、フィルタリング、サンプリングの手法を用いた、自動分類やプレディクティブ コーディング (予測符号化技術) は、e-ディスカバリの半自動化を支える技術と位置付けられる (Exterro 2018)。

コミュニティが求める基準をこれら e-ディスカバリ技術が将来満たしうる、とは断言できない。その逆のこと、長期保存に必要な基準をこれら技術では満たしえない、ともまた言い切れない。大半の文化遺産機関にとって、e-ディスカバリ向けのツールは費用面から手が届きにくいものの、近年、イリノイ大学では、州政府・大学アーカイブズの文脈でこれら技術が有用なものとなるか評価を行うプロジェクトが進められている (Illinois State Archives, and Records and Information Management Services, University of Illinois at Urbana-Champaign 2017)。

4.2 保存機関の課題

民間企業で蓄積された成果とは別に、アーカイブズ機関や図書館では、電子メールの保存に関して独自の取り組みを進めてきた。現状の選択肢の範囲内で、アーキビストらは諸問題に向き合い、判断を迫られているといえよう。おおまかに、次の問いに分けられる。

1. 我々が収集すべき電子メールは何か。どのように収集すればよいか。
2. 収集した電子メールをどのように管理すれば、その証拠的価値を高めうるか。
3. 担当者は、どうやって、電子メールを受け入れ、選別評価を行い、処理し、保存していくことになるか。
4. 添付ファイルやリンク先コンテンツを処理するには、どのような技術を用いればよいか。
5. セキュリティやプライバシーの問題を回避するにはどのようにすればよいか。
6. 大規模または大量のコレクションを処理する際に生じる課題はどのようなものか。

これらの問いに対する知見がこれまで蓄積されてきたけれども、以下の記述はあくまで暫定的なものに過ぎず、今後もさらなる取り組みが求められ、そのような活動が、コミュニティ全体を支えていくことになる。

4.2.1 電子メールの収集

保存対象となる電子メールが生成された元のシステムは多様なため、保存機関における最初の作業、つまり、後続作業に適した形で電子メールのメッセージ類を収集する作業は、複雑なものとならざるをえない。そのため、各システム固有の形式の電子メールを保存に適した標準形式で収集する手段は、現状、文化遺産コミュニティにおいて確立されていない。

組織においては、リスク回避の傾向がまずもって最大の障壁となろう。世評（レピュテーション）に敏感な組織においては、電子メールを保存することから生じる不利益が、潜在的な便益を上回ると見なしかねない。この場合、担当アーキビストは守りの姿勢となろう。例えば、会社の顧問弁護士は、廃棄が認められているのだから、電子メールを削除してしまえば法廷に提出しなくてよくなる、と主張するかもしれない。

とはいえ、技術的な制約が事態を深刻化させてしまい、文化的な意味合いから電子メールを保存することを、より困難なものとしてしまう可能性もある。ほぼ文書化されていないアプリケーションを使用せざるをえない場合もありえ、これは大きな制約となろう。一般に利用されている電子メールは多数のプラットフォームで提供されており、対象となるものがアーキビストにとっては未知のものの場合もありうる。一口に電子メールのアプリケー

ションといっても、サーバ、デスクトップ、ウェブ、モバイル向けそれぞれで、出力方式は異なり、形式やメタデータも違う可能性がある。利用可能な形式の範囲内で、電子メールのメッセージは、採用されたツールで扱える形に変換される必要があろう。このような変換作業は可逆と限らず、メタデータの欠損がありうる。良心的な寄贈者やIT部門が支援してくれる場合もあろうが、それらの知識が偏っていることもありえ、話し合いの場が設けられたとしても、生産的なものになるとは限らない。

別の課題として、添付ファイルや電子メールのスレッドの収集も挙げられる。単に添付ファイルを収集するだけでは事足りない。本文と別添付ファイルの関係性をも維持管理せねばならない。だが、添付ファイルが保管され、各メッセージと関連づけられる手法は多数存在するため、この作業もややこしいものになってしまう。添付ファイルが関連付けられていないことに、収集後しばらくしても気付かない場合もありえる。あるいは、元のやり取りの文脈を保存するためには、返信メッセージを束ねたスレッドが維持されるのが望ましい。

さらに、人的・時間的な制約が二つの課題を生む。第一に、対象となる電子メールアカウントへアクセスするには制約を伴うため、保存機関の担当者には即時の判断が求められる。結果として、価値ある記録とともに、明らかに収集範囲外の資料までもが混在してしまうことがしばしば生じる。この問題を回避する一つの策として、寄贈者側に選別評価してもらうことが考えられるものの、それによりまた別の問題が生じうる。理論上、入念に選別評価する寄贈者もいるであろうが、最後まで行う時間、気力、あるいは知識に欠けるものの方が多い。寄贈者主体の選別評価を進めると、対象は、全メールボックス、ごく一部のみ、選別されたものになるか、またはまったく寄贈されないという結果となりかねない。様々な電子メールシステムで適用可能な、受信箱の整理をしなくても選別評価を実施しやすくするツールなど、仮にあったとしても、期待できない。アーキビストと寄贈者とが協力して作業するのがよりよいやり方であろう。

時折、予期しない形で、ハードドライブやリムーバブルメディアなどの物理媒体上に、電子メールのデータが発見される場合がある。以前にバックアップを取った結果として残されたファイルの場合もあれば、システムに詳しいもの以外にとっては見えないライブラリフォルダ内にそのようなファイルが格納されている場合もありうる。こういった類のデータは、滞貨資料の受入処理がなされるまで、あるいは、利用に供されるまで、気付かれないことが常であろう。この場合、収集時に選別評価を行うことはできず、寄贈者と話し合うこともできない。

このような問題に対応するため、電子メールコレクションの収集手引きを作成してきた機関もある。保存機関の多くは、手近にあるツールをやりくりし、その場しのぎで電子メールを収集してきた。同じような取り組みが各組織それぞれ個別的になされたところで、業界全体の底上げにはつながりにくい。その作業の困難さが際立ってしまえば、電子メールを収集する優先順位も低まりかねない。このような問題が解決されないことには、状況はさらに悪循環に陥り、課題の解消には長い時間がかかりかねない。

現状採用されている収集シナリオを見ると、おおむね4つにまとめられる。「まとめられる」傾向にあるという意味では、明るい兆しともいえる。いずれのシナリオも、前述した課題からはまぬがれえないが、共通の課題も浮かび上がってくる。

- ・ システムからの出力：可能であれば、電子メールを収集する際、システムから直接データを出力してもらうのが一番よい。とはいえ、IT に詳しいものの助けが必要となることが常であろう。この傾向は、個人からの寄贈時というより、組織で電子メールを記録として残す際に顕著だ。ツールによっては、IMAP (Internet Message Access Protocol) 接続に対応している場合もありうる（例：ePADD、Preservica）。その場合、アーキビストは IT 担当者と協力し、採用しているツールが出力形式に対応していること、出力対象のメタデータ内に必要な項目が含まれていること、を確認しておくことになる。また、出力条件をアーキビストが指定し、IT 担当者がそれをシステム上、設定できるのであれば、収集前により詳細な選別評価が実施できることになる。
- ・ ウェブサービス上での出力：アーキビストまたは寄贈者が、ウェブベースのアプリケーション上で電子メールを出力し、保存機関の環境へそのデータを送る場合もある。この手法は、システムからの出力と比べて、メタデータ欠損や変換エラーなどの問題が生じやすい。添付ファイルやスレッドも、元の状態のままでは収集できていない場合がありうる。しかし、収集前に詳細な選別評価を行いうるという点で、この方法はアーキビストにとって有用といえる。
- ・ クライアント側での出力：マイクロソフトの Outlook では PST 形式で、アップルのメールプログラムでは MBOX 形式でデータを出力できる。Gmail でもデータのダウンロード機能が用意されている。とはいえ、出力されたデータは別の形式へ変換が必要となるかもしれない。現状、データ欠損が生じない出力、添付ファイルの確実な出力を担保してくれるようなツールは存在しない。
- ・ ディスクイメージの作成：電子メールを含む記録媒体へアクセス可能な場合、その媒体からディスクイメージを作成し、電子メールのデータを取得することも選択肢となる。ただしこの場合、アーキビストが関与できる範囲は限られており、出力形式、メタデータ、添付ファイル、スレッドについてはコントロールできない。生成されたデータは、内部の、おそらくプロプライエタリなデータストアではうまく動くであろう。収集前に選別評価することはできず、寄贈者とやり取りできることも限定的となる。

これら 4 つのアプローチにはそれぞれ特色があるものの、出力形式は、MBOX、EML、PST など、ある程度決まっている。固有のメタデータが出力されるツールもあれば、そうではないツールもある。プロムが報告しているように、移行ツールの多くは、ヘッダや本文のプロパティ（プロジェクト InSpect [Investigating the Significant Properties of Electronic Content Over Time] で定義されたものを想定している）をうまく保存できることが多い（Prom 2011, 10; Grace, Knight, and Montague 2009）。しかし、ある特定ツールの動作検証をした例はほとんどなく、ないしは、コミュニティ内で重要なプロパティ項目を定めるのがよいと前述の報告書では推奨されているが、それに従った取り組みもほとんど見られない。一つ単純な例を挙げると、あるツールを利用しようか検討している場合、そのツールが、既読メッセージであるか否かといった情報を含んでいるか、また、(RFC5322 で定義されている) フラグやキーワードなどのメタデータを保存しているかを知ることができるとありがたい。あるいは、この、拡張的な機能をプロパティで扱う方向性もありうる。例えば、RFC5451 では認証結果

をヘッダ `authentication-results` として付与することを定めているが、現状、広く使用されており、あるメッセージの真正性を証明する上で重要な技術と位置付けられている (Kuchera 2009)。よりいえば、ローカルに定められた、またはカスタマイズされたヘッダフィールドが活用される場合もあるかもしれない。仮に、音声メールが使用されている機関（特に政府機関など）を例として取り上げてみよう (Vaudreuil and Parsons 2004)。そこで導入されているシステムでは、メッセージを残した人の電話番号を記録する上で、プロプライエタリなフィールドを採用していることが多い。

また、上に挙げた 4 アプローチのいずれも、電子メールが使用されなくなった時点で取られる措置であり、蓄積されたコピーから電子メールが収集できると仮定している。このことは、次の二つの可能性をはらむ。

第一に、残される電子メールの記録が減ってしまいかねない。対象アカウントが使用されなくなるのを待っているだけでは、作成者により削除されるか、規則上定められたリテンションルールに従い削除されるか、あるいは、システムのデータ容量制限上削除されることになり、わずかなメッセージしか残されない結果となるかもしれない。非現用の電子メールが収集される場合、データ損失の可能性はさらに高まる。システム環境変数が削除されてしまっている場合もあるだろうし、電子署名も有効でなくなっているかもしれない。定期収集（定期的に対象電子メールを収集）すれば、他の資料とあわせて、より集めやすくなるかもしれないが、重複をどうするか、スレッド関係をどう保持するかなど、技術的な課題が新たに出てくる可能性もある。ジャーナリングソフトウェアを用いれば、すべての伝送データをリアルタイムに収集できる。定期的に受け入れること（例えば、現用アカウントから PST ファイルを年次で収集）にしても、対象アカウントがその期間中になくなってしまうリスクもあるだろう。そうなると、差分処理や重複排除などの厄介な作業が生じるし、あるいは、定期的な収集実行時に不正処理が生じ、データの欠損が起きてしまう場合もありうる。このような理由から、データ移管の方法や頻度が、長期保存の目標に影響を及ぼし、出所来歴の記録作業を複雑なものとしてしまう。

第二に、使用されなくなった時点で電子メールを収集するとなると、それによる生じるトレードオフは深刻なものとなり、また、民間企業では標準的な、伝送時に電子メールを収集するというアプローチの利点を活かさない、ということが指摘できる。米国州文書館長会議 (Council of State Archivists) / 米国歴史出版記録委員会 (全米歴史出版記録委員会) による電子メールアーカイブに関するシンポジウムで議論になったが、政府情報を扱うコミュニティにとって、ジャーナリングシステムの基本要件や改善点を探るプロジェクトは非常に有用なものとなるだろう。具体的に、ある対象データを特定したり、リアルタイムにデータ収集したり、保存価値をもつと思われる、またはもちそうなメッセージの長期的管理をしたり、といった機能改善につながっていくことだろう。こういったプロジェクトは、州の最高情報責任者 (CIO) と電子メールサービスを提供する主要企業の代表者（例えば、グーグルやマイクロソフト）とで共同で行われることにより、さらに有用なものとなる。

4.2.2 真正性の保証

電子メールのようなデジタルオブジェクトは、それがあべき形で提示されるのであれ

ば、本物であるということができよう。真正性、あるいは、ローゼンバーグによる CLIR 報告書中の表現にならえば、デジタル環境における真正性とは、「完全性、網羅性、正確性、妥当性、元データと比した場合の忠実性、有意性、本来の目的からした適切性を含むものである」(Rothenberg 2000, 52)。

電子メールが偽造、修正、削除、脱文脈化されやすいのは、その標準が開放性とセキュリティとのバランスに重きを置いているとはいえ、驚くべきことではない。法曹・法務業界に目を向けると、電子メールの真正性に関する議論は、1900 年代にまで遡る (Bearman 2017)。判例 *Armstrong v. Executive Office of the President* においては、一義的複製は電子的複製の方であり、電子メールを印刷した複製は、構造・文脈情報、ヘッダ・リンク・タイムスタンプなどのメタデータを欠いているため、便宜的複製の扱いとなり不完全なものとみなされる、との見解が示された。

電子メールコレクションの場合、真正性は、事前に定められた手順に沿って確認されることになると思われる。例えば、業務上の電子メールが、組織内での利用が認められた電子メールのアプリケーションを用いて、通常業務の一環として作成・管理されているのであれば、また、公的なアドレスやメッセージの主要素 (例：ヘッダ情報、添付、署名) がその電子メール内に含まれていれば、さらに、記録のリテンションスケジュールに従い管理され、許可なく修正できないシステム上で保存されているのであれば、当該メッセージを真正なものとするのは妥当なことと思われる。とはいえ、実世界はそう単純でない。犯罪科学の観点から真正性について判定を下すには、多大な労力を必要とする (Banday 2011)。あるいは、電子メールのシステムに関する記録やサーバログは、受信時にメッセージを認証するために用いられる証明書と同じく、真正性の検証時に貴重なデータとなりえよう。電子文書記録管理システム (EDRMS) の環境では、真正性を担保するための業務フローや基盤を提供してくれるが、少なくともその目的から、こういった記録管理システムを使用している保存機関はほとんどない。

つまるところ、真正なものかどうかは、証拠の評価を下す人間、価値判断を行う人間によって判定される。プロジェクト InterPARES では、人々が真正性に関する判断を行う際の一義的な尺度として、電子メールの文脈情報や機能特性はもとより、既存のアーカイブズ機関や図書館で保存された資料であれば真正なものであるに違いないと認識される傾向にある、と結論づけられた (Bunn et al. 2015)。

出所来歴や真正性について考える際、ウェブアーカイブも興味深い事例となる。いくつかの研究では、アーカイブされたウェブ情報について、出所来歴に関する確たるメタデータ (例：収集理由、方法、時期) がないため、学術研究目的で利用する場合、その品質に問題が残ることが示されてきた (Ankerson 2012; Ben-David and Huurdeman 2014; Maemura, Becker, and Milligan 2016)。このような知見は、よりよい保存管理が必要な理論的根拠となりえよう。調査結果が示すように、電子メールやウェブアーカイブなど、ネットワーク上で形成された大規模なデータの集積体を調査対象とする研究者らは、どのように文脈情報を提示できるか、調査結果の妥当性について他の研究者が検証可能な環境をどのように提示できるか、といったことに悩まされている。研究者らは、分析処理手法について記録しておく役割を明らかに担っているが、保存機関もまた、保存・検索システムの要件や設定について考慮

し、出所来歴のメタデータを提供できるよう取り組むことが期待されている。

4.2.3 整理・保存の履歴管理

保存機関の担当者が、メッセージの真正性に影響を及ぼしかねない外部要因を統制するのは難しい場合も多いだろう。一方、保存機関の信頼性を高める取り組みは意識的に行う。例えば次に挙げる取り組みを行っておけば、コレクションの出所来歴が記録され、一連の記録管理工程がしっかりとなされ、作業履歴を後から追うこともできるということが、第三者にも伝わりやすいと考えられる。

- ・ 所有権の移転、または資料管理主体の変更記録を登録しておくこと。
- ・ 文脈情報の保持を担保すること。例えば、ある対象アカウントのフォルダ構造、電子メール間の関係性、添付ファイルとの関係性、他に移管されたデジタル資料と電子メールとの関係性、対象資料に関する補足的なメタデータなどの属性を保存機関側は保存しておくのが望ましい。
- ・ 対象資料について取られたすべての取り組み、その取り組みに責任をもつ人またはシステムに関する監査証跡を保管しておくこと。
- ・ あるストレージ所在場所または媒体から、別の所在場所または媒体へ、複製または移行された場合に、対象資料の固定性について確認が行われること。
- ・ 保存機関でなされた取り組みを保存メタデータの一部として記録しておくこと。なお、保存メタデータは、対象電子メールとともに管理される。

このリストは、出所原則が電子メールにも適用可能なこと、それは他の形式のデジタル資料についても当てはまることを示唆している。とはいえ、大半のアーキビストからすると、専用の本格的な統合ツールは手が届かない代物だろう。そういった自動システムが利用できない機関では、囲み記事「アーキビストによる作業履歴に関するメタデータモデル」で示したような情報を手動で記録した上で、その記録をコレクションの管理データベース上で電子的に保管していくこととなろう。

このような取り組みを進める観点から、機械処理の結果を記録可能な保存システムが導入されている大規模組織もあるかもしれない。理想的には、こういった記録がデジタル来歴の一部となれば、長期保存の点から、デジタルオブジェクトの履歴管理をシステム管理者はしやすくなるだろう。また、真正性についての判断もしやすくなるだろう²⁷。付録 A には、ハーバード大学で実装された、メタデータの自動収集事例をまとめた。

できれば、デジタル保存システムからえられる記録がコレクション管理システムと統合できること（あるいは、その逆）が望ましい。そのような統合機能により、当該資料に関する作業記録を一つのまとまったデータセットとして生成・編集することが可能となろう。もし統合システムが開発されるのであれば、対象デジタルオブジェクトとその関連作業を結びつける ID が付与されることも重要となろう。

²⁷ PREMIS とは、保存メタデータの標準で、コミュニティ内で広く受け入れられている。PREMIS データ辞書では、デジタルリポジトリの保存機能に関する語彙を定義しており、データモデルでは各実体の定義がなされている。PREMIS の活用に関わる文献として、次を挙げておく（Caplan 2009）。

このようなツールがまだ普及していないという事実は、いいかえれば、電子メールコレクションの保存作業に関する情報を収集する機能を組み込むため、コレクション管理ソフトウェアが改修されていく可能性を示唆している。選別、評価、受入、編成、記述、そして保存といった業務フロー全体を通して、関連データは収集されるのが望ましい。システムによっては、あるイベントや作業記録を、単一メッセージ、複数メッセージ、あるいはアカウント全体と結びつけるためのモジュール機能が含まれているものもある。

保存機関でデジタル保存システムが利用されている場合、当該システムで収集されたメタデータには、ファイル形式の識別、マイグレーション、データ取り込みなどの保存関連作業が記録されていることが多い。これらは、バックエンドで動いているシステム上の記録が機械的に集められ、PREMIS や他の形式で記録されたものである。しかし、保存システムの履歴機能では集められていない類のメタデータ項目が、利用者にとっては重要な情報となる場合もあろう。例えば、評価選別結果や記述作業について記した記録、プライバシーに関する確認結果は、コレクションの真正性や文脈を利用者が読み解く上で有用な情報源となる。コレクション管理システムとデジタル保存システムとの連携ができていない保存機関（あるいは、デジタル保存ソフトウェアを保有していないアーキビスト）からすると、このような記録は、作業ノートとして残しているかもしれない。これは、ライトらが今までにないアプローチとして提案した、目録データの中にコメントアウトを注釈や奥付などの形で記しておくやり方と類似している (Light and Hyry 2002)。もう一つの囲み記事は、そのようなデータを人間が読める形態でどのように記せばよいか示した例である。

アーキビストによる作業履歴に関するメタデータモデル

電子メール作成の文脈

- ・ 電子メール作成に用いられたシステム (例: ウェブベースの Gmail やマイクロソフトの Outlook アプリケーション)
- ・ 用いられたプラットフォーム (コンピュータ環境)
- ・ 使用種別 (業務用、私用、家族と共用、組織内で共用、その他用)
- ・ アカウント使用者 (例: 個人または団体)

記録管理の文脈

- ・ アカウントの詳細
- ・ アカウント名/ユーザ名、書類上の公式名称
- ・ アカウント使用期間
- ・ 整理状況の体系 (例: フォルダ構造の有無、どのようにフォルダが用いられているか、ゴミ箱フォルダは下書きまたは確認用に用いられていたか)

保存やキューレーションの文脈 (この種の作業の多くは PREMIS の events として記録)

- ・ 対象資料を受領した担当者
- ・ 対象資料収集時の受入プロセスの詳細

- ・ 対象資料の検査、選別評価、棚卸、検証、記述時に用いられたツール
- ・ 対象資料に適用された長期的な保存方針（いつ、誰が、どのような目的で）

個人文書または外部組織から提供された記録の場合、必要なメタデータ項目が追加される場合もある。例は次のとおり。

- ・ 選別基準
- ・ 提供者が用いていた電子メールシステム
- ・ アーキビストによる当該記録の収集方法

2017 年 10 月 9 日、デルタ大学で実務を担当するアーキビスト、ジェーン・ドーは、ジョン・ドーのコレクションについて、寄贈契約の範囲外であった資料を特定するため、当該コレクションに含まれた当人の電子メールを確認するに当たり、BitCurator を使用していた。寄贈者から提示された検索語にもとづき、Bitcurator 上で対象アカウントのデータを検索し、寄贈の対象外となるべきメッセージ 35 件を特定した。ジェーンは、対象アカウントから削除対象を抽出し、MBOX ファイルを再作成した。ジェーンは、JHOVE と呼ぶファイル形式の識別ツールを再度動かし、新しい MBOX が適切に生成されたか確認を行った。ファイル形式は問題ないと判定されたので、今度は MBOX ファイルを仮領域に移した。10 月下旬に、別のアーキビストが追加の作業を開始予定である。このような作業はすべて、デルタ大学のコレクション管理システム上で、ジェーンが作業ノートとして手動で記録していた。

4.2.4 添付ファイルやリンク先コンテンツの保存

電子メールのコレクションを処理する上で、添付ファイルはアーキビストにとって悩ましい存在である。初期の作業は割合単純なもので、添付ファイルが元のメッセージと結びついているか、問題なく表示できるか確認することだ。厄介なのは、URL を介してメッセージと結びついているコンテンツである。URL は、明示的に記されている場合もあれば、あるいは、実体は本文内でなく別サーバに存在するが、メッセージ内に埋め込まれている場合もある。

添付ファイルやリンク先コンテンツについて、アーキビストは次の課題に取り組まねばならない。

変換時におけるデータ損失・破損の可能性

電子メールのワークフローにおいて、変換ツールがうまく添付ファイルを処理できず、データの欠損や破損が起きてしまう場合もありうる。電子メールを収集し、後続の処理を行う際、プロプライエタリな形式から、EML や MBOX などの標準形式へ変換が必要な場合も多い。同じように、対象となるクライアントまたはサーバのアプリケーションによっては、添付ファイルが出力できないものもありうる。この問題は、品質管理機能が含まれたアプリケーションを用い、変換結果を監視することにより、人力で解決することも可能だろう。とはいえ、そのようなやり方では多くの時間と資源を費やす必要がある。半自動化された検証方法でないと、一組織内で継続もできなければ、複数機関で実施されることもないだろう。

処理上の問題

電子メールコレクションを扱うアーキビストやキュレータは、複数のリスクが存在することを認識しておくのが望ましい。次の側面に留意することにより、保存対象ファイルの真正性が担保され、未来の研究者に役立てられることだろう。

- **削除の可能性：**

添付ファイルのみが保存対象となる場合、電子メール全体を誤って削除してしまうことも容易に生じうる。添付データがメッセージ内に埋め込まれている場合、当該データはメッセージ自体とは別に保存されるものであるため、抽出される必要がある。このことは、メッセージ自体の履歴に加えて、添付ファイルの処理履歴も記録しておくのが望ましいということを示している。もっとも重要なことは、添付を削除したり、メッセージ本体と添付の関係を切り離したりしたことをシステム側で記録しておくべき点である。メッセージの返信または転送時に添付の削除が自動的になされる場合が挙げられる。また、コレクションやスレッドの重複削除を行う際、アーキビストは注意が必要である。

- **ファイル形式の問題：**

添付は、様々な形式の場合がありうる。識別・評価にはそれぞれ異なるツールが必要となるかもしれない。とはいえ、技術的・構造的な観点から添付の確認を行うには、Quick View Plus や（Quick View Plus が内包されている）FTK などの商用ソフトウェアで対応できる。

- **ファイル容量：**

動画など、添付ファイルの中には非常に大きいものがありえ、その場合、ストレージに負荷がかかる。また、ファイル伝送時にはネットワーク負荷がかかる。電子メール送受信時に大規模ファイルを外部に置くシステムも存在することに、アーキビストは留意しておくべきだろう。例えば Gmail の設定では、Google ドライブへ大規模オブジェクトを自動的に配置し、本文にそれへのリンクを生成するといったことが可能である。アップルの Mail Drop 機能でも同様のことができる（Gmail Help 2018; Apple 2018）。いずれのシステムでも、データの保持期間が定められている。サービス向上を目的として新機能が絶え間なく追加されていくため、保存という観点からするとリスクが持続的に発生しているといえる。

- **ウィルスと悪意のあるコンテンツ：**

添付には、ウィルスや悪意のあるコンテンツが含まれている可能性もある。それらは検出されねばならない。また、対応手順の詳細が定められていなければならない。

2016 年に実施された NARA の調査プロジェクトにおいて、主要なウィルス検出ソフトウェアでは、PST ファイル内のウィルスを検出できなかったことが確認されている。実際、どのツールでも、この課題をうまくこなすことができなかったことを NARA は確認している。NARA では、PST から個々の EML ヘッダファイル形式を変換するのがよいと判断され、EML ファイルと添付データを対象としてウィルスチェックが行われている。

保存機関／保存の問題

ある電子メールアカウントまたはメッセージ群が処理対象となったら、ストレージと長期保存の問題に保存機関は直面せざるをえない。次の2点に集約される。

- ・ **添付のストレージと処理：**

添付が MIME 形式でメッセージ内に埋め込まれている場合、対象メールと当該データが関連づけられていることを保証するのは比較的たやすい。とはいえ、添付データは、メッセージ内に埋め込まれた MIME コンテンツとしてではなく、ネイティブのバイナリ形式で監視・格納される方がたやすい。ファイルが問題なく格納されていることを保証するため、元のメッセージ内にポインタが置かれているのが望ましい。多くのクライアントアプリケーションでは、ほぼ目につかない形で、これを自動的に処理する。デジタルリポジトリの文脈では、コンテンツを長期間保存しなければならないが、メッセージと添付を問題なく関連付けておくには、未解決の問題が多数ある。しかし何にせよ、添付とメッセージとの関係を長期間維持できる機能を保存システムに組み込むのが望ましいことは確実だ。極めて単純に考えれば、メッセージと添付との関係を、スプレッドシートか METS ファイルに記録しておくことも可能だろう。より複雑には、対象ファイルの新たな配置場所が参照できるよう、メッセージ本体を修正するといった解決策が考えられる。いずれの解決策であろうと、ストレージ上の配置場所は、処理過程やメンテナンス作業時に変更される場合もあるため、計画は慎重に練っておく必要がある。

- ・ **マイグレーション：**

メッセージ本体は、基本的に ASCII や UNICODE のテキストデータであるが、それと比較すると、添付データはまた別のファイル形式であることが多いため、保存するに当たっては、新たな課題が生じる。電子メールのファイル形式に関する保存方針（例えば、メッセージ自体を保存する際のファイル形式を決める際の方針）では、添付データの保存問題にはうまく対応できないだろう。受領後、添付データは保存時のファイル形式へマイグレーション（変換）されることになっているかもしれない。だが、これは新たな問いをもたらす。元の添付データは保持しないのか。対象ファイルはメッセージとどのように関連づけられるのか。ファイル形式は長期にわたり監視対象となるか。もし監視対象となるのであれば、次期のマイグレーションはどのように行われることになるか。

リンク先コンテンツ：

メッセージや添付に含まれたコンテンツが対象となる場合、リンク先コンテンツの保存もまた課題となる。多くの職場において、ファイルは共有管理システム（SharePoint や BOX など）上に置かれ、それへのリンク情報が電子メールでやり取りされている。この場合、電子メールは、データ間を結び付ける役割を担っており、データ自体を含んではいない。基本的には、容量の節約にもなり、同一ファイルに対して多数の関係者が作業する上で、データコリジョンを回避できるという経営判断といえるが、保存の問題とも深く関係する。利点と

して、リンク先が共有ファイルサービス上の認証付きフォルダであれば、流出リスクを下げられる。仮に対象メッセージを不注意に転送してしまったり、あるいは、対象アカウントが盗まれたりしても、当該ファイルへは指定された受信者しかアクセスできない。だが、セキュリティの問題はもろ刃の剣である。作成者側が対象フォルダへのアクセスを認めていなければ、秘匿しておく必要性がもはやなくなったとしても、アーキビストや利用者はそのフォルダに含まれる必要なファイルへアクセスできない。このような理由から、外部システムへの依存を未然に防ぎ、対象ファイルとは別の複製データを収集できるよう、提供者側とアーキビストは共同して作業する場合も考えられる。

リンク先コンテンツの保存の仕方が、対象組織のコレクション構築方針や諸資源に応じ、より柔軟な場合もあるだろう。リンク先コンテンツはメッセージと別に保管されているため、保存しない判断をするアーキビストもいるだろう。アナログの世界でもこのような戦略は当然見られた。調査を進める利用者の中には、あるフォルダや書簡内で言及された他の文献や出版物をも追いかけようとするものもいるかもしれない、とまでは考えるだろうけれども、実際のところ、そういった文献までアーキビストが探し出しておくことは期待できないかもしれない。とはいえ、コレクション分析の結果にもとづき、あるいは、他の要因から、そういった文献を保存しておくことが重要だと考えるアーキビストも中にはいるかもしれない。

電子メールが、Sharepoint、Slack、Box、Dropbox、Google Drive などのファイル・メッセージ管理システムなどと連動していくと、リンク先コンテンツの問題は、より複雑なものとなる。リンク先コンテンツは、メッセージに不可欠なもの、少なくともメッセージより重要なものとなる場合もありうる。オフライン環境で写真や文書ファイルへアクセスしようとした際、配置場所が特定できない場合は、リンク先コンテンツへ単純にアクセスできず、メッセージ内でうまく処理されない。ネットワーク上にあるコンテンツの場合、問題は悪化してしまう。置き場所は変わりうるし、コンテンツも変わりうる。あるいは、対象コンテンツは単純に保存されていないかもしれない。参照先のネットワーク情報源は、内部ネットワークとの間にあるファイアウォールでアクセスできないものかもしれないし、認証が必要なネットワーク内に置かれているものかもしれない。リンクの中には、短縮 URL サービスや bitly のようなリンク先管理プラットフォームが用いられているものもありうる。それらサービスでは、実際の所在場所と関係がない URL を生成できるため、丹念に調べたとしても、生成された URL から元の URL を導出するのは不可能に近い。生成されたリンクが永続的に維持されることを謳ったサービス（例：bitly）はあるものの、それらサービス自体が存続する保証はない。

電子メールの保存アプリケーションには、現在のところ、メッセージ内の URL で示された情報資源をハーベストする機能は組み込まれていない。まして、埋め込まれた画像、文書、他のコンテンツを保存する機能もない。このような問題を最小化するには、既存サービスの知見を取り入れ、対象となる情報資源が実際に置かれている場所を記録できるよう、そのリンクを辿っていく機能をアプリケーション側にもたせることだろう。そのために電子メールの保存アプリケーションは、主要サービスの進歩に遅れないようついていかねばならないし、それにより、元の所在情報のみ提供できるようになろう。もう一つ別の、より包括的

な解決法として、電子メール保存アプリケーションと、ウェブアーカイブ上で、あるいは（添付に近い形の）リンク先情報資源として、対象資源の検索を行えるようなソフトウェアとを連携させることが考えられる。こういった連携機能は、実用的な観点からすると、変更が生じる前にリンク先コンテンツを収集できるよう、ジャーナリング機能のように、タイミングよく実行されねばならない。主たる問題は、電子メールを保存する上で用いられるツールが、ウェブのリンク情報を抽出すべきか、リンク先のウェブページをクロールするか、保存したウェブページやウェブサイトに DOI や他の永続的識別子を付与するか、保存された電子メール内のリンクを新たに生成された識別子に書き換えるか、といったことである。これは、電子メール保存における文脈や規模の問題が絡むと、さらに厄介となる。

電子メールのコレクションにリンク先コンテンツも含めることになると、それらも保存されるということに当然なり、そうすると、ストレージの費用はあがるし、複雑さも増す。他方、ウェブ上の情報資源のすべてを把握するのは不可能なため、ウェブアーカイビングの範囲もまた問題となる。繰り返しになるが、リンク先の情報資源は、移動または削除されるし、内部ネットワークに置かれる場合もありうる。結局、そういった状況に起因する課題について、すべて解決してくれる方策を描くのは困難を伴う。内部ネットワークにデータが置かれている場合、ハーベスタが収集可能なように認証設定しておいてもらう必要がある。

まとめると、電子メールの添付やリンク先コンテンツの収集は難しい作業となる。組織や開発者にとっては解決すべき問題の一つといえよう。前述した諸課題は逆に、取りうる手段が多数あることを示唆してくれている。

4.2.5 セキュリティ・プライバシーの保証

セキュリティとプライバシーの問題は、対象アカウントが実際に使用されている場合に重要な関心事となる。保存機関へ対象電子メールが引き渡された後であっても、この問題が消えるわけではない。逆に、電子メールのコレクションはセキュアな環境にあり、私的または機微情報が漏れることはないことを示すためには、様々な手段を講じなければならない。

私的・機微・制限・機密情報：

一時的または永続的にアクセス制限がかけられた情報を含む電子メールコレクションの場合も、悩ましい問題が発生してしまう。政府などの組織においては、制限・秘匿・機密情報扱いのメッセージが存在しうる。公式にそう位置付けられていなくとも、扱いに慎重を期すべきメッセージ、学生の記録、健康情報や個人の記録など、法や規則上制限された情報を含むメッセージもあろう。とはいえ、制限が必要と認定されたメッセージは、あったとしても少数に過ぎないことが多い。制限が必要ない他の大量メッセージにまぎれ、特定するのは難しく、分離までしておくとなるとなおさらだ。

個人的な側面、仕事上の側面双方が入り混じっているため、電子メールには、私的な情報が含まれていることが多い。家族の氏名、電話番号、他の連絡先が記載されていることも多い。健康や体調などの機微なことがらについても記録されているかもしれない。ある対象者のみしか知るべきでない内容を含むメッセージもあろう。このような問題は、職場のアカウントだけに限ることではなく、保存機関へ移管された個人コレクションについても当ては

まる。第三者が絡むと（例：機微情報を含むメッセージで言及された第三者が、その電子メールが保存機関へ移管されたことを把握していない状況）、問題はさらに複雑さを増す。

私的・機微・制限情報に絡む諸問題は、受入・保存・利用に至る電子メールの管理全般に影響を及ぼす。対象資料は、著作権法、その他法規類、所有者との法的合意または契約に沿って扱われることになる（Whitt 2017）。大半の電子メールには、私的な内容、雑用の記録、仕事上のやり取りが混在しているとすれば、対象データを収集・整理する上で、アーキビストはセキュアな手段を講じなければならない。セキュリティ対応が求められるということは、アーキビストが電子メールを収集、選別評価、整理、提供、保存する上でのワークフロー、方針、ツールに当然影響を及ぼす。私的なデータをどのように収集・保存することになるか、そしてその規模がどのくらいになるかは、提供者または電子メールのアカウント保有者との交渉次第であろうし、また、記録管理の指針、組織の方針、国および地方レベルの法制度、管理・提供システムとの関係にも依存する。本タスクフォースでは、電子メールのセキュリティ標準について詳細な分析を実施した。参考情報源を確認してほしい（Task Force on Technical Approaches for Email Archives project website 2018b）。

暗号化：

セキュリティに関する別の問題として、メッセージ、フォルダ、アカウントレベルでの暗号化機能についても考慮せねばならない。プライバシーの問題が重要視される業界においては、暗号強度の高いファイル形式（例：IBM 社のグループウェア Lotus Notes では、「電子メールのメッセージは、送信前、受信後または保存前に暗号化が可能である」）が広く採用されている（Sustainability of Digital Formats 2015）。Outlook の PST のような形式でも暗号化には対応しているが、各システムで様々である（Sustainability of Digital Formats 2013a, 2013b）。基本的に暗号化されたファイルは、暗号鍵が提供されていない限り保存しにくい。寄贈時に暗号鍵も保存機関へ提示され、暗号化していない形式で対象資料を保存しない限り、この問題は解決されないだろう。

4.2.6 大容量・大規模コレクションの処理

電子メールアーカイブの規模は、学術・政府機関の間でも様々である。小規模な組織では、一件のアカウントですら、その対応に苦慮しているかもしれない。学術研究の領域で扱われる電子メールアーカイブの平均的な規模は約 4 万メッセージと想定されるが、大きなものになると、100 万メッセージ程度になると思われる。連邦政府機関になると、100 万メッセージでも小さい方になるだろう。

電子メールコレクションは規模が大きくなる傾向にあり、ほとんどではないにしても、その多くに対して、機微情報の有無をある程度まで確認せねばならない。これを目視で行うのは現実的でない。米国連邦最高裁判所判事エレナ・ケイガンの電子メール、7 万 5 千ページ分を確認するのに必要な推定時間は 6 千時間を要す、といわれていた（NARA 2010）。バージニア州知事事務局の電子メール、総量 130 万件のうち 15 万件のメッセージを確認する作業についても、同様の試算が示された（Library of Virginia 2016）。

NARA は、この規模の問題に対処する一つの手段として Capstone アプローチを考案した

(NARA 2015)。連邦政府の各機関では、それぞれ異なる形で受容されている。一方では、重要な部署や個人を特定する際に、他方では、すべての送受信メッセージを保存する判断を行う際にこのアプローチが活用されている (Plante 2015)。あるいは保存機関の中でも、あるプロジェクトや展示に関する電子メールを、永続的に残すものとしてタグ付けしたり、それらを別フォルダとして管理したりするための手引きが作成されたところもある (Rockefeller Archive Center 2006)。

しかし、この対象としたアカウント内のメッセージすべてを収集するというアプローチにおいても、そのメッセージ、参照された実体、含まれる内容は、膨大な規模となる。スタンフォード大学で集められた電子メールアーカイブズの中には、70 万件のメッセージを含むものもあった。ただし、そのうち 38 万 5 千件は未読であり、重要と見なされたものは 7 万 8 千件に過ぎず、20 万 3 千件は送信フォルダにあり、残りは作成者によって個別のフォルダで整理されていたものだという。同様に、マンチェスター大学によるカーカネットプレスの電子メール保存プロジェクトでは、17 万件のメッセージ (16GB) が処理されている (Baker 2014)。各アカウントには、数万の固有表現が含まれる。比較的小規模な電子メールアーカイブですら、ここではスタンフォード大学によるロバート・クリーリーの電子メールアーカイブを例にすると、5 万件のメッセージのうち、自然言語処理の結果、2 万 4 千件の宛先、5 万 7 千件の人名、4 万 2 千件の組織名、2 万 2 千件の地名が抽出されたという。また、多数の形式で、1 万件以上の添付ファイルが含まれていた。電子メールアーカイブズの複雑さ、添付ファイルの形式の多様さ、宛先や固有表現の数の多さに、思わず目がくらんでしまう。

デジタル保存の作業規模：

膨大なコレクションであることに起因する問題は、デジタル保存のライフサイクル全般に関係してくる。選別評価を例に挙げよう。単純に、永続的な価値があるメッセージを判定する上で、大規模になるほど時間がかかってしまう。カーカネットプレスの電子メールが処理された際、スパムやジャンクメールの扱いが重要な問題になったと担当者は記しており、具体的には、正しい送信先から受信していた数百件のメッセージが、件名からスパム扱いされてしまっていたという (Baker 2014)。

電子メール、あるいは、他の種類のデジタル形式のコレクションでも、重複はかなりの程度避けられない。送信者と受信者間で処理がなされるという性質上、少なくともメッセージは 2 件作成される。それに加えて、リストサブやメーリングリストから送信されたメッセージ、第三者が送信したメッセージの複製を受信したもの、あるいは、「全員に返信」されたメッセージを受信したものなどもある。加えて、件名や内容が変わっていても、返信時に前の本文を消さずに残したままにしておく人も多い。このため、一つのアカウントにおいても、複数のアカウントにおいても、同じメッセージが多数重複して存在してしまう可能性がある。

業務用の電子メールアカウントについて、定められた記録管理の手引きに従い、重複メッセージまたは保存対象でないメッセージは、各人で削除されていくのが理想的ではあろう。だが、そういった手引きは無視されるのが常で、多数の重複メッセージが、選別評価、そし

て整理業務に流れ込んでいくのが現実かもしれない。評価選別作業を円滑に行うには、重複メッセージの特定に役立つツール類が欠かせない。内容、作成者、言語の点から似ているメッセージというよりは、実際に同一のメッセージを特定してくれるツール類である。Treesize Proのような重複データ除去ツールの多くは、ファイル管理に焦点が当てられており、電子メールコレクションに特化したものではない。

電子メールコレクションの規模が膨大となる側面は、編成記述作業においても課題を生む。アーカイブズ領域で培われてきた理論と実践に従えば、電子メールのコレクションは、アイテムレベルでなく、集合体として記述するのが望ましいといえる。このアプローチは、アーカイブズのコミュニティで広まった「より多くの成果を、より少ない作業で (More Product, Less Process : MPLP)」という考え方に沿うものだ。とはいえ、電子メールの整理作業ではソフトウェアというツールを活用することができ、それにより、コレクションの内容をアーキビストが確認しやすくなる可能性もあろう。

ジョージア工科大学の情報技術通信研究所におけるプロジェクトでは、様々な資料を対象とし、UNIX コマンドを用いてファイル形式の特定が可能か検証された。ただし、信頼性・正確性ともに結果は芳しくなかったという²⁸。プロジェクトでは、電子メールなど、デジタル資料に含まれた内容を機械的に特定する手法の開発も行われている。もしこれが可能となれば、記述作業の自動化が促進され、情報公開請求への対応など、アクセス機会の増大につながられるだろう。

同時に、大規模な紙のコレクションでは実現が困難であった、作業の自動化や記述の仕方の面などで、電子メールやデジタル資料特有の性質を活用する方向性もあろう。例えば、ePADD では、メッセージの重複除去機能だけでなく、送受信先の名寄せも行えるため、エイリアスや個々の宛先を集約すること（編集も可）もできる。電子メール内で言及された固有表現（人名、団体名、地名）の抽出も可能である。TOMES (Transforming Online Mail with Embedded Semantics) プロジェクトでは、政府機関の電子メールを対象とし、同じような機能の開発が行われている (NHPRC 2018)。固有表現抽出技術の使い道は様々だ。出現頻度のデータにもとづき、対象コレクションに含まれるもっとも重要なテーマを見極めることも可能となろう。さらに、そのキーワードを対象コレクションのアクセスポイントとして付与することも考えられる。典拠データと固有表現の比較を行い、国や国際レベルの典拠レコードへリンクを貼ることも可能であろう。

コレクションの提供、レファレンスサービスという面からすると、対象となる電子メールは人間が読める形でないといけない。未来の研究者は、個々の電子メールを見たいと思うであろう。とはいえ、スニペットを表示させる形から、(ePADD で具体化されたように) コーパスレベルで提供する形態に変わりつつある。全文が表示され、添付も確認でき、ヘッダ情報も利用できるといった世界だ。研究者の中には、データの一形態として電子メールを扱う

²⁸ ジョージア工科大学の報告書では次のようにまとめられている。「ファイル形式の情報を管理し、file コマンドで参照されるマジックファイルを生成するためのデータベースシステムについて述べたい。ファイルのシグネチャを含むファイル形式に関するメタデータは、(マジックナンバー [ファイル識別子] を含む) マジックファイル内でなく、データベース内で容易に変更可能としている。これは、UNIX の file コマンド、マジックファイルの柔軟性が向上したことを意味する。GUI でも file コマンドを使用できるようにしている。これまで 800 種類以上のファイル形式を対象とし、それらのシグネチャを検証した。また、file コマンドとシグネチャのファイルの信頼性に関する評価を実施した。」(Underwood et al. 2009)。

ものもいる。例として、History Lab とバージニア州知事の電子メールを対象とした研究を挙げておく²⁹。

大規模になるほど多くの資源が必要：

他の形態のデジタル資料と比較して、電子メールはその規模が大きいため、それらを保存するにはより多くの資源を必要とする。大量のデータを扱うには、ストレージ、マシンパワー、ネットワーク帯域といった面で、相当のコストがかかる。大規模コレクションをシステムへ投入するには、相応のコンピュータ資源、ネットワーク環境、そして忍耐を要する。Google のバックアップツール Takeout を使用した本タスクフォースメンバーの一人は、自身の電子メールデータのダウンロード URL をえるのに 2 日かかったという。

インフラに係る費用と並んで、労働コストも無視できない。提供者、アーキビスト、そして研究者に至るまで、データを処理するには時間とスキルが欠かせない (Rouse 2018)。ランド研究所による e-ディスカバリに関する報告書によれば、「自動コード化技術 (predictive coding) により、法律家が (目視で) 文書を確認する時間の約 86% を節約できた可能性がある」という (Pace and Zakaras 2012, 68)。サンプル文書 47,650 件を対象とした調査で、目視による確認では 686 時間かかっていたが、e-ディスカバリ関連技術を適用したところ、98 時間にまで短縮できたと報告されている。しかし、短縮できるとはいえ、いまだ莫大な時間を人がかけねばならないことに変わりはない。このことは、技術面だけでなく、それを利用する担当者にも同程度の投資が欠かせないことを示唆していると思われる。

イリノイ大学の研究者らは、NARA の Capstone アプローチで収集された電子メールを整理する上で、e-ディスカバリの自動化を進める目的から、自動コード化技術 (「電子情報開示時に文書を確認する作業の一部を自動化するため、キーワード検索、フィルタリング、サンプリングを行うこと」) の検証を行なっている (Illinois State Archives and Records and Information Management Services, University of Illinois at Urbana-Champaign 2017)。e-ディスカバリ (電子情報開示) は、「民事・刑事訴訟上、証拠として利用される目的から、電子データを探し、特定し、確保し、検索する手続き」であるが、これは複雑で厄介な作業となり、膨大な量のデータを処理せねばならない (Rouse 2018)。ランド研究所の報告書によると、電子情報開示の対応コストは、1 ギガ当たり、約 22,480 ドルにのぼるという (Pace and Zakaras 2012, 28)。大規模データの処理費用は高額となるが、今後、自動コード化のような技術の適用により、コストを低減させていくことは可能だろう (Pace and Zakaras 2012)。また、技術開発が進み、整理業務にそれら技術が組み込まれていかない限り、電子メールという歴史的記録の大半は、いつまでも利用できないままとなってしまうかねない。歴史研究者からの支援があろうが、権利関係の記録がしっかり管理されていようが、説明責任の圧力があろうが、サービス向上が図られようが、そもそも技術的に対応できない可能性もあるということだ。

自動コード化などの技術で作業コストを低下させられるとしても、金銭・時間的な側面以外のことについても考慮せねばならない。保存機関のコミュニティでは、未知の技術やアプローチの導入を広く検討していかなければならない。新しいスキル・技術が必要とされ、既存業

²⁹ 次を参照。 <http://www.history-lab.org/>; <http://www.virginiamemory.com/collections/kaine/>

務、財源、ガバナンスの見直しも求められよう。端的に言えば、電子メールを保存するには、単にソフトウェアやその購入費用があれば事足りるわけではない。

5. ありうる解決策とワークフローの例

解決すべき課題は残っているものの、電子メールを収集・保存し、利用に供する取り組みは進んでいるといえ、関連ツールが開発されるとともに、コミュニティも醸成されてきた。取りうるアプローチやツールにはいくつかの選択肢があるため、個々の状況に応じたワークフローを組み立てていく上では、そこから何を選ぶべきかが問題となる。本章では、保存アプローチとワークフローについて、いくつかの提言を行う。付録 A と付録 B では、アーキビストにとって有用と思われるツールやサービスについて、詳細情報をまとめた。より網羅的なリストについては、本タスクフォースのウェブサイトを参照されたい（Task Force on Technical Approaches for Email Archives project website 2018e）。

5.1 保存戦略

保存機関が取りうる保存の対策には選択肢がある。ビット列の保管、マイグレーション、エミュレーションを挙げておこう。保存計画では全体のライフサイクルを考慮せねばならないが、実際には保存機関の主導で進められることが多いだろう。その際には、計画を策定するだけでなく、対象ファイルの持続性や提供可能性に影響を及ぼしかねない変化を監視することも含まれる。

どのようなアプローチが採用されるにせよ、おそらくネイティブの API が既存ツールの拡張機能を利用することになるだろうが、対象電子メールと適切な保存作業に関するデータを共有できるようなツールが望ましい。当該組織内の電子メールや添付ファイルを扱う機関においては、組織で定めた方針に従わせることにより、作成・保存される形式を比較的管理しやすいといえる。他方、外部の団体や個人から電子メールを収集する機関においては、所蔵対象となるものはより多様なものとなるだろうし、そのため、保存計画を実行するにはより多くの資源が必要となるだろう。

5.1.1 ビット列の保存

ビット列の保存とは、ビット腐敗（bit-rot）や意図せぬ削除・変更のような脅威から、コンテンツを保護する手法やサービスのことをいう。分かりやすいものとして、全米デジタル管理連盟（National Digital Stewardship Alliance : NDSA）が定めたデジタル保存の推奨レベルというレベル 1 に対応する活動が挙げられる。具体的には、チェックサムの生成と固定性のチェック、保存用バックアップの作成、ファイル形式の解析などだ（National Digital Stewardship Alliance 2013）。ビット列保存では、デジタルオブジェクトの表示、文脈、解釈といった側面は考慮しない。むしろ、研究者や他の利用者に提供する必要が生じるまで、管理対象のシステムまたはデジタルリポジトリにデジタルオブジェクトをひとまず投入しておく形をとる。

ビット列保存は、電子メールを保存する準備がまだ整備されていない機関にとって、一つの選択肢となることは疑いえない。実際に整理され、利用できるようになるまでの期間、電子メールのメッセージや添付ファイルが保護されることになる³⁰。いずれ、法制度や提供者

³⁰ プロプライエタリなシステムの電子メールの場合、特有の技術に依存しているため、保存機関側のシス

との契約条件に反しない形でデータが提供できるようになれば、電子メールのメッセージすべて、あるいは一部を手作業により利用者へ開示することになろう。このアプローチは単純で分かりやすいが、リスクも高い。もっとも根本的な形でデータが保存されているとはいえず、扱いが容易になるとはいえないからである。

5.1.2 マイグレーション

マイグレーション戦略には2つの選択肢がある。そのいずれかのみ行う場合もあれば、双方を組み合わせて行う場合もある。

- ・ 不安定な形式から、より安定した形式へデータを移行または変換すること。
- ・ デジタル資料を作成するのに用いられた様々なファイル形式から、より少ない管理可能な規模の、かつ、オリジナルのデータがもつ複雑な構造や形態をエンコードできる標準的なファイル形式へ、デジタルオブジェクトを変換すること。

電子メールの場合、後者の戦略を取ることが多い。なぜなら、既存ツールやワークフローが、ある特定のファイル形式を前提としていることが多々あるためだ。ローゼンバークはこのアプローチを推奨している。

新しい媒体上にデジタル情報を複製し（新たなファイル形式へ変換することもありえ、「マイグレーション」とも呼ばれる）、再生させる必要があることは、科学や企業といった領域だけでなく、図書館学やアーカイブズの領域においても認識されてきた。これには継続的な取り組みが欠かせない。記録媒体は、物理的に読めなくなったり、旧式化したりしてしまう。移し替えた先の媒体がそうなる前に、データを複製し続けていかねば、いずれ利用できなくなるだろう（Rothenberg 1999, 11）。

ファイル形式の変換とは、あるファイル形式を別の形式へと変換する作業をいう。基本的に、ファイル形式の旧式化リスクを減らすために行われるが、ファイル管理上、ソフトウェアとの互換性担保やファイル形式の正規化のためにも実施される。おそらく、現在、もっとも採用されている保存戦略だろう。電子メールコレクションのマイグレーションについて補記しておくべきこととして、メッセージのファイル形式のみが対象となり、添付のファイル形式は考慮対象外とされることが多いことに触れておく。

電子メールについて、マイグレーションの戦略が取られることが多いのはなぜだろうか。一番大きな理由としては、よく利用されているツールで扱われるファイル形式が限られているためだろう。デファクトの保存ファイル形式として、MBOX（アカウント単位）とEML（メッセージ単位）が挙げられるが、これらは文書化が十分になされており、プロプライエタリなものではなく、オープンなファイル形式である。多様な電子メールで読み書き可能であり、ePADDやハーバード大学の電子アーカイビングシステム(EAS)でも扱える(Prom 2011)。ePADDでは、MBOX または IMAP のアカウントデータのみ投入でき、それ以外の場合はMBOXへ変換することが推奨されている。EASの場合は、多数のファイル形式で投入可能

テムへ投入する以前の問題として、そもそもデータ抽出がうまくいっていない可能性もある。

とされているが、投入後の処理時に EML へ正規化されている³¹。

別のファイル形式として XML が挙げられ、採用している機関もいくつかある。XML は、ソフトウェアやハードウェアに依存しないマークアップ言語であるため、保存という意味では理想的なものだ。ただ、利用の観点からアカウント単位で XML 形式の電子メールを処理可能なツールはほとんどない。電子メールの XML スキーマで国際的に普及しているものはないが、スミソニアン協会とノースカロライナ州立文書館はアカウント単位の電子メール用 XML スキーマ (Email Account XML Schema : EAXS) を策定している。米国の文書館で採用しているところもあるという。名称が示唆する通り、EAXS は、アカウント単位で単一の XML 文書を作成する。とはいえ、色々と設定することが可能であり、メッセージと添付を切り離したり、添付を外部のストレージに保存したりすることもできる。

マイグレーションの本質は変化であり、変化には常にある種のリスクが伴う。様々な理由から、移行時にデータの損失が生じる可能性は避けられない。例えば、互換性の問題がある (変換後のファイル形式では正確に実行できない場合もあれば、移行前のデータを完全に識別できていない場合もある)。ビット列が失われること、ファイルが壊れることなどもありうる。組織は、他のデジタル保存戦略と同様に、テスト結果、利用可能なツール類、知識、スキル、経験にもとづき、メリットとデメリットを比較せねばならない (Waters and Garrett 1996)。

もしマイグレーションを実施することに決めたのであれば、元のファイル形式でも電子メールを保持しておくのが望ましい。これが元データとなる。将来の技術進化でより効果的な保存手段が取りうることになれば、元のビット列に立ち返り、直接それをいじることも可能となろう。

5.1.3 エミュレーション

状況によっては、電子メールを保存する上で、エミュレーションが適切かつ有効なアプローチと考えられる場合もある。特に、著名な個人の電子メールを元の文脈内で体験したい、当該個人の作業環境に没入したい場合にはエミュレーション技術が重要となってくる (Loftus 2010)。

エミュレータとは、あるハードウェア環境を別のハードウェア環境上で再現するソフトウェアアプリケーションをいう (Daintith and Wright 2008; Wikipedia 2017a)。ソフトウェアは、ある固有のハードウェア向けに設計されており、その実行を妨げる何らかの変化が生じると、旧式化してしまう。旧式化したソフトウェア自体や、旧式化したソフトウェアにより閲覧可能なデジタルコンテンツが利用できるよう、エミュレータにより、元々動いていた古

³¹ ハーバード大学では、数年後、EAS をオープンソースのアプリケーションとして公開予定だという。本報告書執筆時点では、利用はハーバード大学内のみに限られている。正規化については、デジタル保存連合によるハンドブック内の「ファイル形式と標準」の箇所を参照されたい。「分野によっては、旧式化よりも、増殖化が問題となっているのは間違いない。ファイル形式が正規化されていなければ、多様なファイル形式とその多数のバージョン違いに組織は悩まされることになる。PDF、文書、画像ファイル形式の異なるバージョンに、である。特有のデータ形式がものすごい速さで生み出されていく領域になると、この問題はさらに悪化しよう。どのファイル形式がリスクに晒されているのか、あるファイル形式に用いられるツールはどれか。こういったファイル形式のすべてを把握し管理できるのだろうか。深刻な課題である。」 (Digital Preservation Coalition 2015)。

いコンピュータ環境を再現できるかもしれない。電子メールは、ソフトウェアとの関係性という意味で特徴がある。具体的には、他のデジタルオブジェクトよりも、利用者側と管理者側のソフトウェア間で頻繁にやり取りが交わされるという点である（The Document Foundation Wiki 2018）。電子メールコレクションにエミュレーション環境を導入することとした場合、そのコンピューティング環境（電子メールのクライアント／アプリケーションのバージョン情報など）に関して記したメタデータは不可欠となる。

電子メールを扱う（開封、編集、削除、送信などを行う）アプリケーションは多数あり、また、電子メールの標準は（他のデジタルオブジェクトよりも）比較的単純である一方、エミュレーションのソフトウェアでは、実際の動作体験や、添付画像や HTML メールなど、利用者側に表示される内容が変更されてしまう可能性もある（Archives New Zealand 2018）。

いいかえれば、エミュレーションとは、元々利用されていたアプリケーションでの実際の動きを体感できる仕組みといえる（OS 全体がエミュレートされるのであれば、あたかも元の OS で動いているかのように、添付ファイルを実行し、閲覧もできる）。長期にわたりエミュレーション戦略を可能とするには、未来の研究者へ元のコンテンツを提示できるよう、電子メールのアプリケーションが保存されていなければならない。さらにいえば、検索機能の限界や、「全員に返信」機能の使いやすさなどを未来の利用者が理解できるようにしておくには、古い電子メールソフトウェアが再現できなければならない。

ある文脈内での電子メールの利用：

電子メールが収集されるのと同時に、元のデスクトップ環境も受け入れ対象となる場合がありうる。この元のデスクトップ環境を動かすエミュレータが実現できれば、作成者の設定や好みなどの文脈をも未来の研究者は体験することが可能となろう。個人の電子メールコレクションの場合、デジタルフォレンジックのツールにより、所有者のハードディスクから電子メールのイメージファイルが取得されるかもしれない。さらに、そのイメージファイルはエミュレーション環境へ移され、実際に再現も可能となろう。企業や政府機関内で集められる電子メールの場合、より統一された形でデスクトップ環境のディスクイメージを収集できるかもしれない。これにより、必要なソフトウェアも収集でき、将来、エミュレータを介して利用できるようにすることも可能となろう。

エミュレーションの課題：

他のアプローチと同じく、エミュレーションもまた課題を孕む。旧式ソフトウェアの入手や保存、知的財産権で保護されているソフトウェアやコードの扱いなどについてである。加えて、添付ファイルなどのソフトウェアに依存した資料へのエミュレーション環境が必要になるかもしれない。そうするとますます、単一の機関がすべてのソフトウェアを収集できるわけではない。こういったソフトウェアに依存した文化遺産資料へのアクセス機会を将来的にも保証するには、ソフトウェアを識別・収集・記述・保存することを目的とした国際的な取り組みが欠かせないが、ある程度調整されつつ、分散的に行われる形が望ましい。なお、エミュレーションに関する技術的な情報、関連プロジェクトの詳細については、本タスクフォースのウェブサイトで公開している（Task Force on Technical Approaches for Email Archives

2018a)。

5.2 柔軟なワークフロー設計を支える相互運用性

いずれの保存アプローチでも、複数のシステムやツールが組み合わせられる。端的にいえば、相互運用性が重要といえる。オンライン版のオクスフォード英語辞典によれば、相互運用性とは、「2 つ以上のコンピュータシステムまたはソフトウェアの機能が、データを交換し、利用すること」だという。

5.2.1 複数ツール間での処理

電子メールの保存を相互運用的な観点から捉えると、複数システム間で（アカウント情報、メッセージ、ヘッダなどの）データ交換が行われている点を指摘できる。電子メールの処理は、反復的に実施されていく。利用可能な資源の範囲内で、あらかじめ計画された作業として実施される場合もあれば、対象コレクションについて何か新たな情報がもたらされた場合に実施される場合もある。また、技術が進展し、対象データを新しいやり方で扱えるようになったことで、整理済みのコレクションの扱いを見直す場合もありえる。したがって、電子メールのライフサイクル全体を通して（未来を含む）、データが扱えるツールでなければならない。それにより、アーキビストは、コンテンツ（メッセージを削除する、あるいは、機微データを加工するなど）またはメタデータ（公開猶予期間終了に伴い権利情報を更新する、あるいは、ファイル形式の変換などの保存活動を記録する）を確認し、編集することが可能となる。

あるツールでは解析・分析・理解できないコンテンツやメタデータでも、アクセス可能なのであれば、保存の対象となろう。また、電子メールのツールによりメタデータの扱いが異なったとしても（例えば、あるツールでは個々のメッセージ単位でメタデータを管理している一方、他のツールではアカウント単位で管理している場合など）、メタデータとコンテンツを関連づけて処理・表示し、ツール間でメタデータを交換することも可能だろう。あるいは、あるツールでは対象コンテンツやメタデータをうまく処理できないとしても、他のツールでは可能かもしれないため、データ自体を保存しておくことも想定されよう。

2016 年、こういった相互運用性に絡む課題について、ハーバード大学図書館のワークショップで議論が交わされた。電子メールの保存に取り組む実務者が集まり、管理ツールの要件について話し合われるとともに、共同して課題に取り組む方向性が確認された（Harvard Library 2016; Murray and Engle 2015）。参加者間で共有されたのは、単一のシステムやワークフローでは問題の解決に結びつかないという見解である。これは、各機関において、それぞれ固有の方針や目的に沿った解決策が求められるためである。より現実的なアプローチについても議論されており、そこでは、（個別のニーズに応じて）多数のワークフローに対応可能な、柔軟かつ互換的に各種ツールを使い分けられる環境が構築できないか、意見が交わされている。技術や要件が変わったとしても、ツールを組み合わせうまく適用させていくことができれば、組織の柔軟性を高めることにつながろうし、異なる組織同士の取組をも促すことになろう。

このような柔軟なワークフローを設計するに当たり、電子メールのデータ（コンテンツと

メタデータ) を利用・交換するのに必要な最低限の要件について、実務者とツール開発者との間で合意がなされていなければならない。何が必要かコミュニティ内で共通理解が進むほど、必要な機能が多様なツールで実現可能となるだろう。いいかえれば、相互運用性の高いツールを構築するには、個々のニーズを満たしつつ、コミュニティ全体で共有されたニーズを満たす機能要件が必要であり、そのためにはバランスの取れた柔軟なアプローチが望ましい。

相互運用性を探る一つの手段として、ハーバード大学のワークショップでは、既存ツールの評価も実施された。ワークショップ後も、ハーバード大学図書館は、Artefactual Systems 社と手を組み、各種ツールに関するデータを収集し続け、相互運用性を高める上での課題と機会について調査している。最終報告書『電子メール保存システムの相互運用性』では、相互運用性に関する汎用的な要件をまとめており、これは、今後議論を展開していく上での土台となろう (Simpson 2016)。

プロプライエタリなシステムから電子メールデータの収集手法を開発することを目的とした TAMES プロジェクトにも触れておこう。初期のデモを確認した経験からも、タスクフォースのメンバーらは、組織アーカイブズが Capstone 型の収集を行い、集められたデータを整理していく上で、この TAMES プロジェクトは非常に役立つとと考えている。もしこのプロジェクトが追加の外部資金を獲得できるのであれば、プレディクティブ コーディング機能の開発が予定されているという。また、プロプライエタリなシステムから電子メールのデータを収集する予定の図書館と共同で取り組む想定という (Simpson 2016)。

5.2.2 コミュニティで共有可能なデータモデルの策定

ハーバード大学の報告書で示された汎用的な要件の中には、メタデータとコンテンツともに、その基本的なファイル形式と構造に関する共通理解が必要との指摘も含まれていた。その対応策の一つとしては、データのモデル化作業が考えられる。

電子メールの保存を目的とした抽象データモデルが共有されていれば、ファイル形式やメタデータに関する要件が後年変わったとしても、モデル自体は古びず、役に立ち続けることが期待される。コミュニティで共有可能なデータモデルがあれば、例えば METS や JSON が採用される場合など、パッケージ構造のファイル形式に対応する際にも有用と思われる。また、デジタル保存に固有の要件が加えられるにせよ、その核になるものとしてこのモデルは捉えられるだろう。保存機関で電子メールを整理する際に使用される多くのツールで、電子メールに関する RFC 標準が利用されているけれども、それらは、コミュニティで共有可能なデータモデルを考える上での起点になるかもしれない³²。

電子メール向けのデータモデルでは、データ交換用のパッケージ形式、コンテンツ、メタデータ、それらの関係性を表す構造情報が含まれることになるだろう。またデータモデル内では、構成やその要素において必須になるものと任意のものとが定義される。相互運用性を高める目的から、ファイルのパッケージ形式もコミュニティ内で選択することになるかもし

³² これは、2017 年 10 月 31 日に、本タスクフォースのメンバーの一人であるウェンディ・ゴーゲルとの会話において、カリフォルニア電子図書館 [CDL] の UC キュレーションセンター [UC3] 副所長であるステファン・アブラムス (訳注: 2018 年に退任) から推奨された内容である。

れない。それにより、コンテンツ自体、コンテンツへのポインタを組織化できる、あるいは、ツール同士でのデータ移行・交換が可能となろう。また電子メールのデータを処理するツールは、対象となるメタデータやコンテンツの種類、所在情報など、データモデルが示唆するところを活用して構築することができよう。それにより、ツールの機能を最大化することにもつながる。

コミュニティ内で構築された一例として、保存システムへ電子メールのデータを投入し保管するためのコンテンツモデルを策定したハーバード大学図書館の例を挙げておこう (Harvard Wiki 2018, 57)。

5.2.3 ファイル形式の要件定義

電子メールの保管・交換向けファイル形式を定義する作業は、保存コミュニティが主体的に取り組めるものといえる。いくつかの試みがこの領域で個別に開始されたしたが、うまく調整していけば、よりよいツールの実現化につながるだろう。

電子メールのメッセージと添付データ：

前述した通り、電子メールのメッセージを保管・管理する上で、デファクト標準のファイル形式は MBOX と EML である (Library of Congress 2016; Murray 2014; Prom 2011)。MBOX の入出力に対応しているツールは既に多数あるため、ツール構築時の最低要件を検討する上で、MBOX を起点にするのは理にかなっていよう。EML から MBOX への変換、あるいはその逆の変換を行うツールがあれば、より柔軟性が増す。

メッセージとは別に添付データが存在し、MIME 形式で表されう。MIME タイプに従いデータは処理・保存される、あるいは、バイナリデータとしてツール間でやり取りされるが、保存という観点では、固定性のチェックと外部識別子による確認が伴う。

メタデータ：

電子メールを保存する実務者の中には、一つの選択肢として、METS に PREMIS を組み込むことを検討しているものもいるかもしれない。いずれも、アーカイブズ領域やデジタル保存コミュニティで広く受け入れられているものである。

PREMIS『保存メタデータのためのデータ辞書』は、国際的に普及している標準の一つだ (PREMIS Editorial Committee 2008)。「長期にわたり、デジタルオブジェクトの利用可能性、同一性、持続性、実行可能性、理解容易性、真正性」を保証できるよう、デジタル保存を行う上で核となるメタデータが定義されている (Lavoie and Gartner 2013, 2)。この PREMIS を採用するとしても、次に、どのように実装するかを決めねばならない。PREMIS を実装するコンテナ形式として広く普及しているものとしては METS が挙げられる。METS は、他の形式のメタデータを包むことができ、あるいは、参照する形でまとめることができる (Zierau and Peyrard 2016)。

PREMIS は、アーカイブズやデジタル保存のコミュニティにおいて重視された機能に対応できるようモデル化・設計されている。このことは、PREMIS を採用する一つの理由となる。このデータモデルでは、5 つの実体が定義されている。例えば、イベントや権利という

実体が挙げられる (Caplan 2009; Dappert et al. 2013)。「イベント」という実体は、ファイル形式の変換や削除など、処理過程でコンテンツに加えられた変更点を記録するために用いられる。「権利」という実体は、権利情報を記録するために用いられ、その多くは実際に動作するよう設計される。電子メールの権利情報が多数のツールで認識できるようになれば、非常に有益であることは間違いない。多数のツール上で、(手動または自動で) その権利情報が表現され、管理され、そして動作するのであれば、アクセス制限を課することが容易になるだろう。

PREMIS と METS は、種類を問わずデジタルコンテンツの保存に役立つが、電子メールに特化したメタデータプロファイルが普及すれば、その有用性はさらに高まるだろう (Caplan 2009; Zierau and Peyrard 2016)。メタデータプロファイルをモデル化する上で、METS 内に PREMIS を組み込むのは一つの方法となりうる。

交換：

交換パッケージ用のファイル形式について合意が進むと、ツール間での移行が促進されるよう。パッケージ化されたファイルは、使用される各ツールで認識・処理できるのが望ましい。しかし、パッケージ化された中のコンテンツ自体へアクセスしたり、展開し (パッケージを解い) たりする必要は必ずしもない。仲介的なツールがこの役割を担い、後続の処理過程において用いられるツールで対応可能な形式へコンテンツを変換してしまうこともできるだろう。

これら交換パッケージは、あるツールでは出力形式となり、また、別のツールでは入力形式となるわけだが、OAIS 参照モデルの語で表現すると、それぞれ、配布用情報パッケージ (Dissemination Information Package : DIP)、提出用情報パッケージ (Submission Information Package : SIP) に呼応する (OAIS 2012)。米国議会図書館とカリフォルニア電子図書館が開発した BagIt は、コンテナ型のファイル形式で、多くのツールで生成・展開が可能である。電子メールを交換するパッケージ形式として、BagIt は有力な候補となりうる (Library of Congress 2018)。

5.2.4 API と相互運用性

ここでの API (アプリケーション プログラミング インタフェース) とは、データ交換時に共通の方式でコンテンツとメタデータをソフトウェアツールが処理できるように構成されたプロトコルのことをいう。

API により、多数の点で相互運用性が高まる。主には、機能等の詳細を把握することなく、異なるシステム同士を一体化させられる点が挙げられる。システム間の違いに縛られることなく、環境を柔軟かつ反復的に変えていくことが可能となる。これは、システム間の依存関係が固定的でないためであり、より小さな単位でサービスをつくりあげていくことが現実味を増している。ニーズの変化に呼応するため、既存のアプリケーションを捨て、新しいツールの開発を行う上で、API の技術は役に立つ。

機能的な観点からいえば、様々なデータのシリアルライゼーションやファイル形式 (XML、HTML、JSON) に API は対応でき、様々なセキュリティ技術 (トランスポート層のセキュ

リティ、暗号化など）を活用できる。また、ウェブ環境において、大半の API は REST (Representational State Transfer) の原則に従って動いている。

API がもたらす明らかなメリットは、設計が十分になされた場合ではあるものの、技術的な複雑さから開発者を解放してくれる点である。利用者は、統合したい複数システムで用いられたプログラミング言語をすべて理解する必要はない。ただ、API コールに必要な言語を分かってさえすればよい。特に小規模機関のアーキビストからすると、セキュリティ、完全性、真正性が担保できるけれども、過度な技術スキルやインフラを不要とするツールが欲しい。こういった要求を満たす上で、REST API は有用な技術となるかもしれない。エンドポイントまたは URL プレフィックスに、パラメータと返戻データの形式を指定してやり取りすることができる。

電子メール特有の REST API :

グーグルやマイクロソフトなど、電子メールサービスの提供企業でも REST API が利用できるようになってきた (Google 2017; Microsoft Developer Network 2018)。これは、機能が不足している電子メールソフトウェアを補うことに主眼が置かれているようだが、電子メールの保存という点からも有用な技術といえるかもしれない。他にも REST API のサービスは提供されており、異なる電子メールサービスを統合的に利用することが可能となっている (Context.io 2018; FWD:Everyone 2018; Nylas 2018)。

可能性 :

- API により、電子メールのやり取りが抽象化可能となる。電子メールへアクセスする際、特定のクライアントソフトウェアを使う必要はないため、API ライブラリをコミュニティで開発していく流れが促進されよう。
- 電子メールの出力または移行ツールの検証やバックアップの目的からも API は機能する。例えば、API を介してメッセージ件数をえることができ、Google Takeout で電子メールを出力した結果が問題ないか確認することもできる。
- 一般的に、これらの API は、細かい粒度でリアルタイムに電子メールのデータへアクセスできるようになっている。API を介せば、ある時点のダンプデータを取得するというより、ある期間の増分データを収集していくアプローチも可能にしてくれる。そのようなニーズが今時点の文化遺産機関にあるかという微妙ではあるが、組織アーカイブズや企業アーカイブズにおいては、大いに可能性を秘めた機能といえるだろう。

デメリット :

- API は、単一のエンドポイントからアカウント全体を移行することには対応していない。アカウント全体を移行するには、(少なくとも) いくつかの API コールが必要であるが、往々にしてそれは膨大なものとなろう (特に Gmail においては、個々のメッセージ単位でコールが必要となるかもしれない)。
- API の返戻の多くは、広く受け入れられ、文書化もされている MBOX や Maildir などの

形式ではなく、メッセージ単位で生成される JSON ファイルである³³。通常、返戻内のメッセージは RFC 2822 にもとづき生成されているはずであるが、すべてのヘッダ情報が提供されているかどうか、別途解析されるのが望ましい。

- こういった課題に対する一つの解決策として挙げられるのが、JSON Mail Application Protocol (JMAP) かもしれない。JSON ベースの標準で、電子メールのクライアントアプリケーションとサーバ間のやり取りをより効率化することが目指されている。IETF の JMAP ワーキンググループは、データモデルとプロトコルに関するドラフト文書の策定に取り組んでおり、2018 年 5 月には新版が公開された。なお、このプロジェクトは GitHub 上で進められている。

制約やリスク：

電子メールを処理する上で、API は逆に実装上の制約を生む可能性もある。特に、大半の電子メールの処理は、アカウントレベルで行われ、後続ツールに引き継がれるまで、大量のメッセージを処理するには長時間かかってしまう。電子メールの処理過程では、あるオブジェクトがまだ前工程で処理されているところに、あるツールでは別のデータを次工程へ引き渡すことが行われるが、ここにマイクロサービスのアプローチを組み込むのは容易でない。加えて、仮に電子メールの保存アプリケーションで API をより広く活用するのであれば、次の課題についても検討が必要である。

- すべてのアプリケーションが API に対応しているわけではなく、システム間連携が不可能な場合もある。
- 文書化が杜撰または不完全なことから、API を十分に活用できない場合もある。
- 個別のアプリケーションに関する詳細な知識が求められるというより、開発者は、API のエンドポイントと動作パターンについて学ぶ必要に迫られるかもしれない。
- すべてのサービスで大規模な運用ができるわけではないため、特に対象の量が増すほどに、拡張可能性が課題となってくるかもしれない。
- 大規模データになるほどネットワーク帯域に負荷がかかり、エラー検知時の複雑さや対応コストが増すこともありうる。
- API により、システム機能を抽象化できる。しかし、抽象化は常に有用とは限らず、時に問題の解決を遅らせてしまう要因となりうる。

API のワークフロー シナリオ：

仮に、追加のツール開発に必要な共通の交換形式が策定可能だとすれば、公開 API が電子メールの保存管理に役立ちうる領域は次の 2 つにまとめられる。

1. ライフサイクルの各段階をつなぐ機能として役立ちうる。例えば、電子メールの評価選

³³ Maildir は、電子メールの格納ファイル形式の一つである。各メッセージは個別のファイルとして扱われ、一意の名称が付けられて管理される。そのため、他のメッセージに対する作業の影響は受けない。ダニエル・バーンスタインが qmail 向けに作成したものであるが、他のプログラムでも実装されている。Maildir は、「ロック機能を扱う必要があるプログラムコードがなくなるよう」に考案されたものである (Maildir 2018)。

別に対応したツールで API が提供されれば、受入や保存管理向けのツールと連携が可能となる。大半のツールには入力／出力機能が十分に備わっているが（IMAP や SMTP などのプロトコルにもとづくものが多い）、その多くには汎用的な API が欠けている。あるいは、電子メールの整理・開示ツールから、別に存在する記述用システムへメタデータを出力可能となるかもしれない。いいかえれば、文化遺産コミュニティで開発されたツールとコミュニティ外で開発されたツールとを含め、JMAP などの API を活用することにより、ワークフローの統合が実現できるかもしれない。

2. 大量データの処理時に役立ちうる。電子メールの保存時に行われる作業量は大規模となるのが目に見えている以上、公開 API を活用し、作業の自動化と効率化が実現できれば有益だろう。例えば、選別評価、ファイル形式の変換、固定性の確認時にトピックモデルを応用することも含まれよう。

5.3 ワークフローと実装シナリオ

本節では、共通のワークフローについて、シナリオベースで紹介する。規範的なものではなく、現状なされている実践例のイメージを提示したい。あわせて、相互運用性が高まることにより、作業効率や成果が向上することを期待できる領域についても触れたい。

5.3.1 ビット列保存のワークフロー

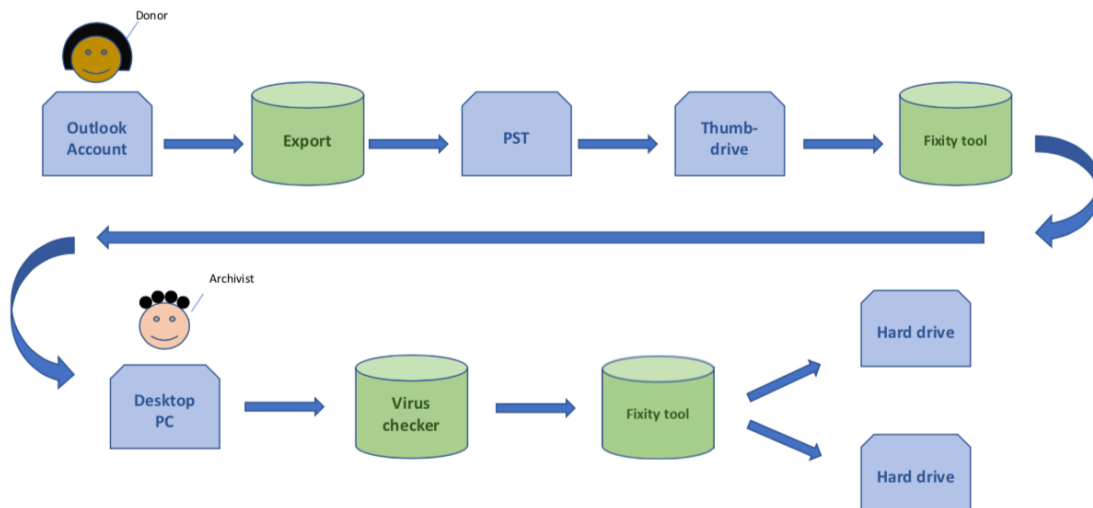
作成者との関係と受入前の事前評価

サンチェス女史は、ある地域で著名な作家の一人であり、電子的記録も含まれる調査資料を、その地域にある小規模な歴史資料館へ寄贈したいと考えていた。その歴史資料館の所蔵資料の大半は紙の記録であったが、デジタル資料も受け入れ始めたところである。資源が許す限り、デジタル資料にも対応していく姿勢を資料館は示していた。デジタル技術に詳しい一人のアーキビストの手により、デジタルコレクションが構築され始めた段階に過ぎず、これまでの関心は、文書ファイルと絵画資料の保存・提供に絞られていた。しかし、サンチェス女史は、寄贈対象に電子メールのコレクションも含めたい意向を示していた。資料館側としては、どのようにすれば提供できるかまだ知見がないものの、将来的に保存・提供が可能となることを期待して、女史の意向を受け入れたいと考えている。

寄贈時の合意書では、サンチェス女史は複数のアカウントを保有しているが、そのうち、仕事上の調査で使用していた一つのアカунツ（マイクロソフトの Office 365 家庭向けライセンスに含まれる Outlook アカウンツ）のみを歴史資料館へ寄贈することが記されている。私的目的で使用している複数の Gmail アカウンツについては、引き続き彼女自身が管理することとなる。歴史資料館のアーキビストであるミシェルは、対象コンテンツやアカウンツの文脈についてサンチェス女史にインタビューを行い、このアカウンツには、個人を特定できる情報や他の私的な機微情報がほとんど含まれておらず、したがって、事前の選別評価作業もほぼ不要であることを確認している。

電子メールの収集

ミシエルの助言に従い、サンチェス女史は Outlook の出力機能を使い、対象アカウントのコンテンツを PST ファイルではき出す。さらに女史は、歴史資料館へ引き渡す予定の他のデジタル記録とあわせて、PST ファイルも USB メモリへ格納する（第 3 図）。あわせて女史は、歴史資料館の指示通り、USB ドライブ内の各ファイルについて、オープンソースのツールを用いて、チェックサム値を生成する（AV Preserve 2018）。



第 3 図. ビット列保存の基本ワークフロー

評価選別、整理、保管

ミシエルは、データ（PST ファイルも含む）が含まれた USB メモリを受領し、そのデータを業務用 PC のローカルドライブへコピーする。ミシエルは、ウィルスチェックを行った後、真正性に問題がないか、ツールで固定性情報を確認する。何も問題がなければ、当該データを保管用の外付けハードディスクへコピーする。1 つはコレクション保管部屋に置き、もう 1 つは後続処理で使用する。保存システムにデータが投入されるまで、この 2 つのハードディスクにより、ある程度の冗長性が担保される。

5.3.2 マイグレーションのワークフロー

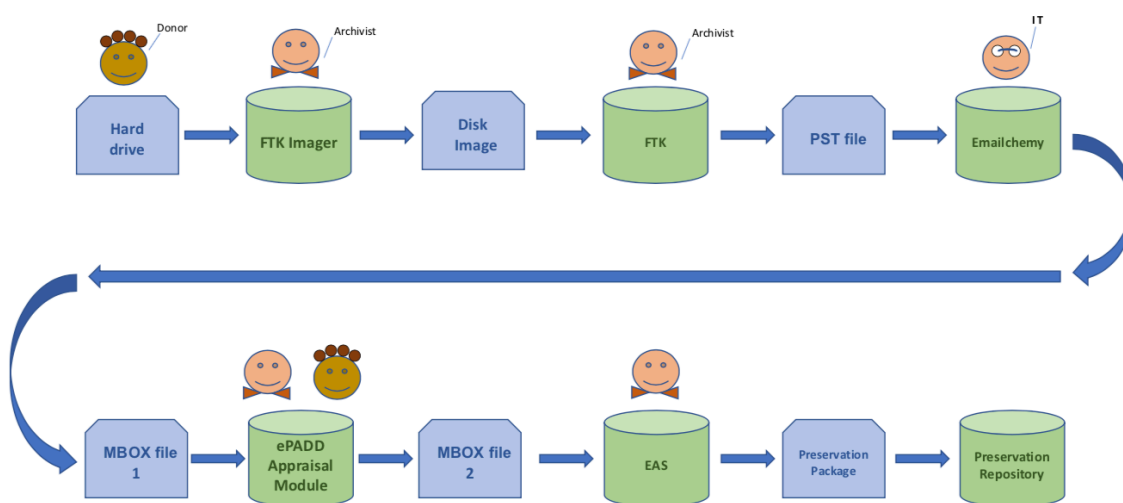
マイグレーションのワークフロー（1）：ハーバード大学図書館

作成者との関係と受入前の事前評価

1. キースは、ハーバード ビジネススクールにおけるベイカー図書館特殊コレクション部門のデジタルアーキビストである。退官予定の F 博士から、ハードディスクを受領した。
2. キースは、FTK Imager を使い、対象ディスクのディスクイメージを作成する。さらに、FTK を使い、コンテンツを分析した結果、事前に把握していなかったが、Outlook を含む、F 博士の PST ファイルをキースは発見した。
3. キースは F 博士に連絡を取ったところ、博士は電子メールがハードディスク上に含まれていたことに気づいておらず、寄贈することにためらいをみせた。キースは、まず電子

メールの内容確認を行ってはどうかと博士に提案するとともに、この電子メールがいかに重要なものであるかを繰り返し説明した。

4. 博士は、電子メールの確認を行うことについて同意した際、当該アカウントは先月閉鎖したのでもう使用していないとも述べた。キースは、閲覧室に博士が作業できるスペースを用意する。キースは博士のために、ePADD の選別評価モジュールが動くマシンの準備を進める。ePADD を利用するためには、PST ファイルを MBOX に変換しなければならない。キースは、セキュアな環境で変換してもらうよう図書館の IT 部門に依頼する。IT 部門では変換ツール Emailchemy のライセンスを所有していたため、そのツールを使い、PST から MBOX への変換を実施した。
5. キースは、F 博士が確認できるように、MBOX ファイルを ePADD の選別評価モジュールに投入する。
6. F 博士が閲覧室に到着後、ePADD の選別評価モジュールでの検索方法と、ビジネススクールへ寄贈してもよいものに対するフラグの付与方法について、キースは F 博士に教える。F 博士は、彼女と夫や息子とのやり取りを特定・除外できることに満足し、また、電子メールが含まれることで、大学教員アーカイブズ資料 (faculty archives) としての価値が高まると考えるようになった。F 博士による選別評価作業が問題なく終わられた後、ベーカー図書館特殊コレクション部門で受け入れられるように、キースは MBOX ファイルを出力する (第 4 図)。



第 4 図. ハーバード大学図書館におけるマイグレーションのワークフロー

受入、整理、保存システムへの投入

7. ハーバード大学図書館のアーキビストとして、キースは、電子メールを整理し、そのデータを大学の保存システムへ投入する際、EAS (電子メールアーカイビングシステム) を用いる。キースは、EAS 上のセキュアな格納場所へ、ePADD から出力された MBOX ファイルを投入する。また、EAS 上で、受入 ID とコレクションとの紐付け情報などの小規模なメタデータを付与する。データ投入の過程で、EAS は Emailchemy を使い、電子メールのメッセージを EML 形式に変換し、各メッセージと添付ファイルにメタデー

タを付与していく。またキースは、F 博士が次期大統領候補にアドバイザーとして仕えた経歴があることを知っていたため、関連メールを特定する検索を実行し、集められたものにシリーズタイトル名を付した（将来、そのような観点から扱いやすいように）。キースは同様に、学生、職員、経営管理上の記録を特定する検索を実行し、該当データには大学のガイドラインに従い公開猶予期間を設定した。あわせて、アクセス制限が必要なメッセージ、添付ファイルにはフラグを立てた。次いでキースは、コンテンツ（メッセージと添付）とメタデータを自動的にパッケージ化するためのボタンを押下し、長期保管のため、パッケージ化したファイルをハーバード大学の保存システムへ投入する。対象のメッセージと添付データは、大学の方針に沿って、すべて公開猶予期間に当てはまるものであるため、この時点で検索や開示についてキースが対応すべきことはない。

マイグレーションのワークフロー（2）：スタンフォード大学図書館

作成者との関係と受入前の事前評価

ePADD は、保存対象となる電子メールの内容を熟知している作成者本人を支援するためにつくられたものだ。ePADD を使えば、電子メールが保存機関へ引き渡される前に、機微情報の確認を行い、対象データにフラグを立てておくことが可能である。このプロセスがどのように進められるか、アッシュ夫妻の例をシナリオとして取り上げて見てみよう。いずれも、スタンフォード大学図書館の特殊コレクション部門へ寄贈されるものとする（第 5 図）。

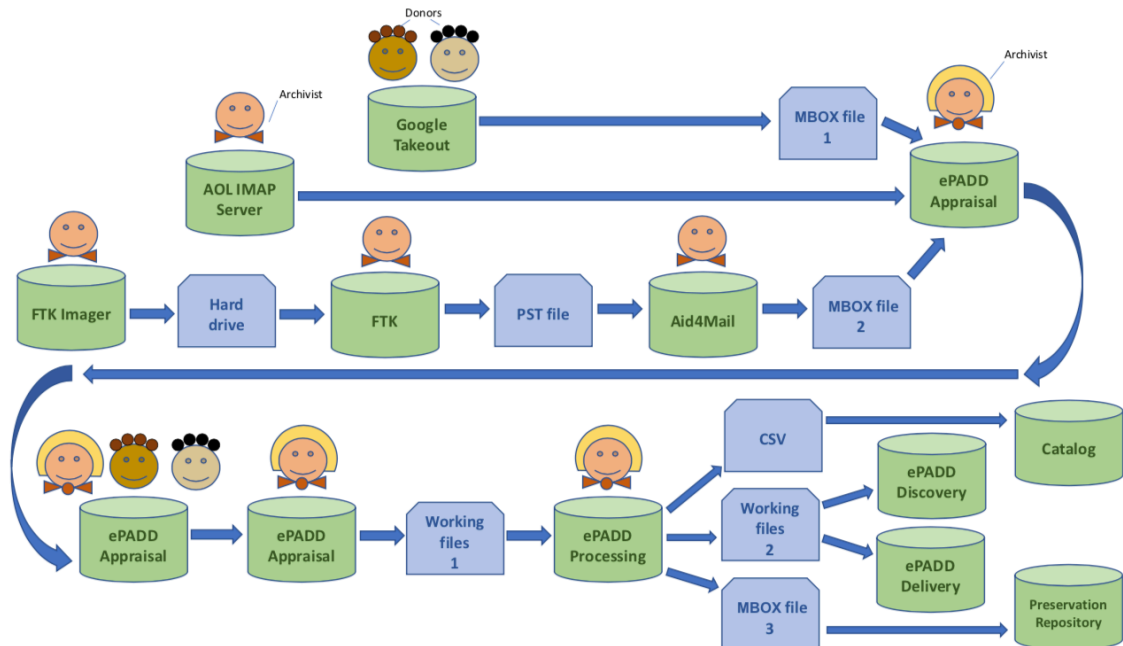
合意書が取り交わされた後、プロジェクト担当のアーキビストまたはキュレーターは、可能であればデジタルアーキビストと協力の上、対象コレクションの作成者らを訪問する。保存機関の担当チームは、対象となるデジタル資料、想定されるワークフローなどについて、夫妻にインタビューを行う。その結果、コレクション内には多数のファイル形式が含まれること、その中には電子メールも含まれることが確認された。具体的には、アカウントが 3 件分あり、それぞれ別のオンライン電子メールサービスのものである。

アッシュ夫妻が保有するアカウントは、1) Gmail アカウント（使用中）、2) もはや使用していない AOL のアカウント、3) 古い Outlook のアカウント（もはや使用されておらず、外付けハードディスク上で保管されていた）である。

電子メールの収集：

1. Gmail アカウントについて、アッシュ女史は Google Takeout を使い、対象アカウントのデータを選択し、ダウンロードする。対象アカウントの MBOX ファイルがダウンロード可能になったら、Google から彼女へ通知がいく。彼女は MBOX ファイル（第 5 図中の file 1）をダウンロードし、FTP を介してスタンフォード大学側の担当者フレイヤへデータを送る。
2. スタンフォード大学において、フレイヤは、AOL の IMAP サーバから直接、アッシュ夫妻の古い AOL アカウントのデータを ePADD 選別評価モジュール内へ投入する。
3. 使用されていない Outlook のメッセージについては、スタンフォード大学のデジタルアーキビストであるピーターが、ボーンデジタルラボにある FTK Imager を使い、ハード

ドライブのディスクイメージを作成する。次いで、FTK を使い、コンテンツ自体の分析を行う。ピーターは、Outlook を含む PST ファイルを発見したため、FTK を使って対象データを抽出し、Aid4Mail を用いて、MBOX ファイル（第 5 図中の file 2）へ変換する。



第 5 図. スタンフォード大学図書館におけるマイグレーションのワークフロー

選別評価

4. アッシュ夫妻の訪問に備えて、フレイヤは、スタンフォード大学図書館に導入された ePADD 選別評価モジュール内に MBOX ファイル（file 1 と file 2）を投入する。AOL アカウントのデータ投入先と同じである。
5. アッシュ夫妻が到着後、フレイヤは両氏に、引き渡し対象から除外したい、あるいは、公開猶予期間を設けたいメッセージと添付を特定する方法について教える。
家族間の私的なやり取り、病気・体調に絡むメッセージはすぐに特定されたので、アッシュ夫妻はそれらを移管対象から外すことにする。ePADD には機微情報が含まれるメッセージを特定する機能（文字列照合、語句検索、固有表現抽出、添付画像確認などを含む）があるものの、アッシュ夫妻は当該機能を利用していない。夫妻は、もっとも私的な情報と思われるものが除外されればそれでよく、残りはすべて移管対象としてよいと判断した。

整理

6. アッシュ夫妻退出後、フレイヤは、ePADD 選別評価モジュールでアッシュ夫妻が移管対象外と判断したものを除くメッセージすべてを出力する。この出力作業により、作業ファイル 1 が生成されるので、フレイヤはこの作業ファイル 1 を ePADD の整理モジュールへ投入する。
7. 整理モジュールでフレイヤは次の作業を行う：
 - a. 保存用コピーの MBOX ファイル（file 3）を出力する。このファイルには、アッシュ

夫妻から寄贈され、デジタル保存システムへ投入される電子メールが含まれる。

- b. 研究目的での利用に供するため、次の後続処理を進める：
 - i. 機微・機密情報、法的に保護されたコンテンツを特定し、除外した上で、残りのメッセージを作業用ファイル 2（提供用コピー）として出力する。
 - ii. データクリーニング、典拠コントロールや名称の統一などを実施した上で、目録レコード投入用に CSV ファイルをはき出す。

受入、保存、利用

8. フレイヤは、出力された CSV ファイルをスタンフォード大学の目録レコードに投入する。
9. 研究者から当該コレクションの利用が申し込まれたら、フレイヤは作業用ファイル 2（提供用コピー）を ePADD の開示・提供モジュールに投入する。
 - a. 開示モジュールは、スタンフォード大学で管理された独立型のウェブアプリケーションとして動く。実際に閲覧室を訪れて全文を見る前に、まずは遠隔から、編集加工された電子メールコレクションを研究者は閲覧・検索できる。
 - b. 提供モジュールは、スタンフォード大学の閲覧室内に設置されたワークステーション上で動く。そこでは制限なく、添付を含む提供用データすべてに研究者はアクセスできる。
10. フレイヤは、MBOX ファイル（file 3）をデジタル保存システムへ格納する。

マイグレーションのワークフロー（3）：スミソニアン協会アーカイブズ

選別評価と受入

1. スミソニアン協会の展示責任者カールは、退職の予定である。収集担当アーキビストのジェニファーは、彼の電子メールを選別評価し、受入作業を進めるところである。
2. スミソニアン協会の電子記録アーキビストであるリンダは、電子メールへのアクセス許可を IT 部門へ公式に依頼する。IT 部門の電子メールシステム管理者と共同で、カールの現アカウント、旧アカウントのデータ収集を行う。
3. リンダは、カールが保有する 2 つのアカウントから収集した電子メールファイルを正規化するために MessageSave ツールを用い、MBOX ファイル（set 1）を生成する。
4. 生成された MBOX ファイル（set 1）がリンダにより DArchMail へ投入されると、機械的にメタデータが生成される。

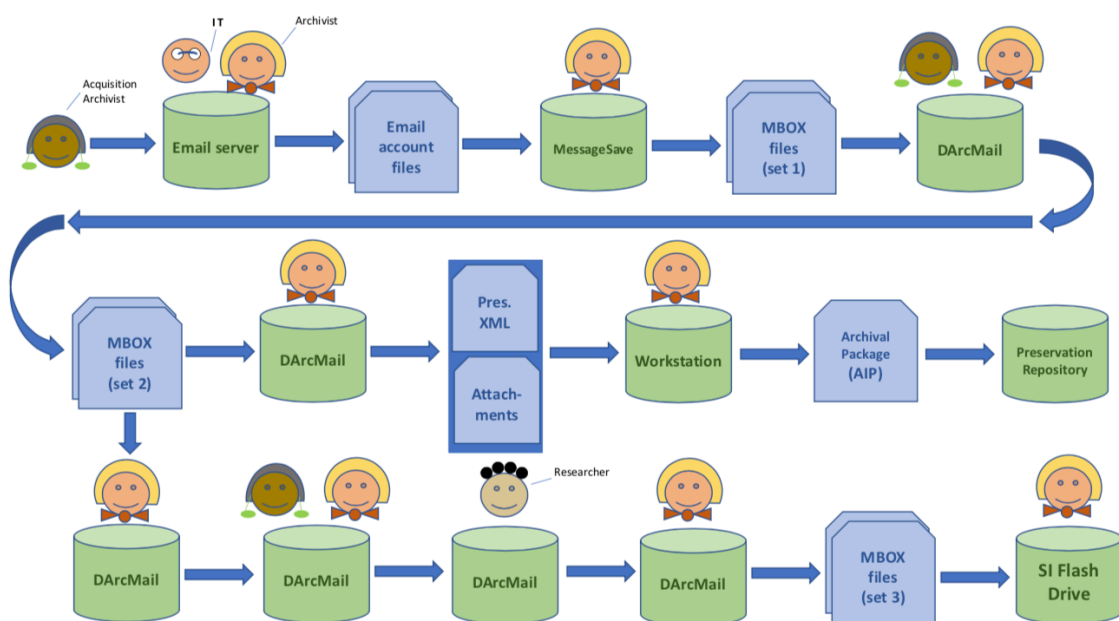
整理、保存、保存システムへの投入

5. DArchMail 上で、ジェニファーとリンダは対象電子メールの整理を共同で行う：
 - a. ジェニファーは、メタデータの確認を行い、最終受入前に整理が必要か判断する。
 - b. ジェニファーとリンダはコンテンツを分析した上で、最終的な受入対象とするメッセージを取捨選択するため、メタデータ項目、本文、添付ファイル名などを対象に詳細な検索式を考える。

6. 取捨選択の結果、最終的に選ばれた受入対象を、2 つ目の MBOX ファイル (set 2) として、DArchMail から出力する。
7. DArchMail を使い、リンダは受入対象の MBOX ファイルを選択し、保存用の XML 形式へデータ変換する。この作業過程において、DArchMail では次も行われる：
 - a. 受入対象内の各メッセージ、かつ、受入対象全体の固定性に関する値を生成する
 - b. 受入対象コンテンツのメタデータを生成する
 - c. 添付データを関連ディレクトリ上に切り離す
8. ローカルのワークステーション上で、リンダは、保管用情報パッケージ (AIP) をフォルダ内に作成する。AIP には、元の電子メールファイル、正規化された MBOX ファイル (set 1)、DArchMail からの出力ファイル (XML と添付データ) が含まれる。
この API フォルダは、長期保存用のシステムへ SFTP を介して移行される。

利用

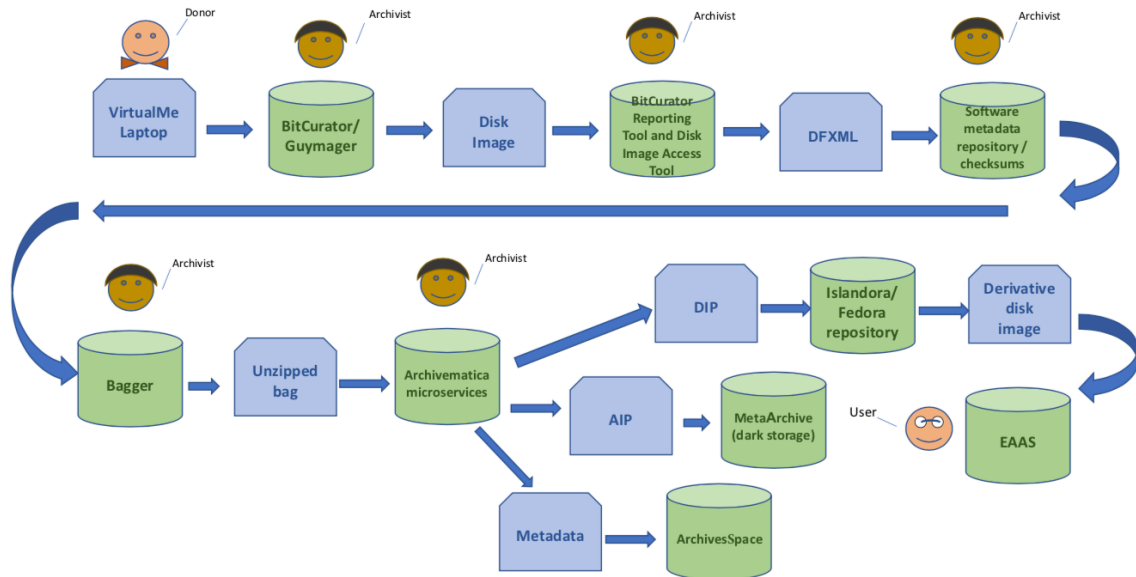
9. 登録済の研究者メイジーから、当該電子メールの利用が申し込まれる。
10. リンダは、スタンドアローンのワークステーションにある DArchMail へ MBOX ファイル (set 2) を投入する。
11. リンダとジェニファーは、対象アカウントのメッセージを確認し、機微情報または個人情報が見つければ、それらに加工を施す。
12. メイジーは、当該電子メールの確認作業を進め、研究目的からコピーしておきたいメッセージを選択する。
13. DArchMail を使い、リンダは、メイジーが選択したメッセージを対象として、MBOX ファイル (set 3) を出力することにより、配布用パッケージ (DIP) を作成する。次にリンダは、メイジーのために、(セキュリティの観点からスミソニアン協会が用意した) 空のフラッシュドライブに DIP を格納する (第 6 図)。



第 6 図. スミソニアン協会における XML マイグレーションのワークフロー

5.3.3 エミュレーションのワークフロー

電子メールの保存アプローチの一つとして、ソフトウェアの保存やエミュレーションは重要な手段である。ここでは、受信箱や関連アプリケーションからメッセージを抽出し、エミュレーション環境で実行する作業過程を概観したい（第7図）。



第7図. エミュレーションのワークフロー例：
電子メールと添付のディスクイメージ利用

作成者との関係と受入前の事前評価

1. 機関 X のデジタルアーキビストであるジェーンは上司とともに、引退した情報学研究者アルカディ・イヴァノフのもとを訪れる。彼は、資料の寄贈について検討している。
2. 保存と利用に関する機関の方針や作業工程について話し合った後、イヴァノフ氏は興味がそそられ、ジェーンに「仮想の自分 (VirtualMe)」プロジェクトを紹介する。これは、2000 年代後半からイヴァノフ氏が進めているものだ。生身の人間としてではなく、デジタルの記録として残すため、仮想空間でのみ存在するデジタルペルソナ（デジタル上の人物像）を、イヴァノフ氏は購入したラップトップ上で作成してきた。
3. イヴァノフ氏は、古いコンピュータと個人文書を寄贈することに合意する。イヴァノフ氏は、私的なデジタル資料について公開猶予期間が設けられることを望む一方、「仮想の自分」データ（デスクトップ上にある Outlook のデータを含む）については、すべて公開してもかまわないと考えている。

受入と整理

4. アーカイブズ機関へ戻った後、ジェーンは、アーカイブズ資料管理システムの一つである ArchivesSpace で受入記録を作成し、新しく収集する資料について記述作業を始める。
5. ジェーンは、「仮想の自分」データを BitCurator が搭載されたワークステーションまでも

っていき、Guymager を使い、「仮想の自分」のディスクイメージを作成する。あわせて、Info ファイルに受入番号と記述対象の資料 ID を格納する。

- a. ジェーンは、BitCurator の報告ツールとディスクイメージ利用ツールを動かし、コンテンツの分析を行う。
- b. ジェーンは、BitCurator から DFXML (デジタルフォレンジック用の XML) を出力し、対象ドライブ上にあるファイルすべてをリスト化する。リストには、実行ファイルやファイルの依存関係も含まれる。
- c. ジェーンは、電子メールのメッセージと添付データが利用可能なエミュレーション環境を準備する。ドライブ上にあるファイルとそれらのチェックサム値を、ソフトウェアのメタデータリポジトリ上にあるチェックサム値と比較するスクリプトを彼女は動かす。なお、ソフトウェアのメタデータリポジトリには、収集済みの既知のソフトウェアに関する記述・技術メタデータと、そのチェックサム値が格納されている。
- d. ジェーンはさらに、(チェックサム値で) 検証済みのソフトウェアのリストと、ファイル拡張子のリストとを比較し、ディスクイメージに含まれるファイルすべてを開くのに必要な追加のソフトウェアを特定する。

投入と利用

6. ジェーンが属する機関 X では、投入から開示までのワークフローが、複数のオープンソースツールを組み合わせることにより実現されている。
 - a. ジェーンは、BitCurator から関連データ (「仮想の自分」データのディスクイメージ、BitCurator がはき出すレポートデータ、[デジタルフォレンジックのツールの一つである] bulk_extractor からの出力データ、DFXML ファイルを含む) を出力の上、Bagger と呼ぶツール (BagIt の仕様に従ったパッケージファイルを生成する GUI ツール) を使い、未圧縮のパッケージファイル (bag) を作成する。これにより、デジタル保存システムである Archivematica へ投入することが可能となる。なお、作業過程で数度、追加のメタデータが METS ファイルに足されていく。
 - b. Archivematica では配布用情報パッケージ (DIP)、いいかえれば未圧縮のパッケージファイル (bag) が生成される。この DIP は、保管・開示目的から、デジタル保存システムである Islandora/ Fedora に投入されていく。
 - c. 適切なコレクション単位で DIP が投入された後、「仮想の自分」パッケージファイル (bag) を解析していくことになる。ここで、新規のオブジェクトと処理対象データに作業 ID が付与される。例えば、受入番号と資料 ID を付された METS ファイルが対象となる。また、DFXML は Solr によりインデキシングされ、検索や機械的なダウンロードが可能となる。
 - d. Archivematica では保管用情報パッケージ (AIP)、いいかえれば未圧縮のパッケージファイル (bag) も生成されるが、この AIP は、非公開のダークアーカイブ目的から、分散型保存システムのネットワークである MetaArchive で管理される。
 - e. 受入番号と ArchivesSpace 上で付された資料 ID にもとづき、ディスクイメージのデ

ータが ArchivesSpace 上で関連づけられる。これにより、ArchivesSpace 内のメタデータに、Islandora 上にあるオブジェクトへのリンクと、METS ファイル上に記録されたディスクイメージに関する最小限の記述データが追加される。

7. Islandora/ Fedora の遠隔利用者は、
 - a. Islandora の開示機能により、ディスクイメージ上での選択やクリックといった行為が可能になる。これにより、ディスクイメージ内のファイルすべて（電子メールの添付を含む）を利用するのに必要な追加のソフトウェアとあわせて、対象ディスクイメージの派生データが生成される。
 - b. サービスとしてのエミュレーション（Emulation as a Service : EAAS）機能をブラウザ上で動かし、対象ディスクイメージの派生データを利用することが可能になる。
 - c. コンテンツ全体を閲覧することが可能になり、閲覧時点でのマシンの状態を表す DOI も生成することができる。

6. 今後の方向性：提言と次なる段階

電子メールは、重要な歴史的記録の一つとなっている。この重要な記録を保存し、アクセスを保証することは、アーカイブズ機関やアーキビストにとって大切な役割といえ、またその価値がある（Society of American Archivists 2011）。電子メールの保存という課題に向き合わなければ、管理・説明責任を放棄することになり、ひいては、歴史的な記録が失われてしまうことになる。この問題は放っておいてどうにかなるものではない。行動を起こすべきは、まさに今しかない。

前章で概観したワークフローのシナリオは、いずれも実現可能ではあるものの、現状の保存機関が電子メールのコレクションを選別評価し、受け入れ、整理・保存し、利用に供する上では、依然として困難を伴う。保存機関の担当者は、多様なツールから適切なものを選び出し、それらを複雑な作業工程に組み込んでいかなければならない。技術に詳しい職員がいて、比較的経営書資源が潤沢な機関であれば、これも現実的な話になりえようが、大半の機関では無理であろう。既存のツールが電子メールを保存できないからではない。問題が複雑なためである。コミュニティもツールも生まれきてはいるが、まだ熟していない。基礎調査や方針策定が必要な面もある。

課題は明確だ。良質な事例も確認できる。電子メールを保存すること自体は可能だが、今後は、より積極的な取り組みと幅広いステークホルダーの関与が欠かせない。したがって、本タスクフォースでは、互いに補い合う2領域に絞り、提言を行いたい。ここでは、1) コミュニティの形成とアドボカシー、2) ツールのサポート・検証・開発、を想定する。各領域とも、望ましい取り組みについて列挙した。すぐにでも対応を開始できそうな障壁が低い課題もあれば、一方、綿密な計画と資金を要す課題も含まれる。

6.1 コミュニティ形成とアドボカシー

電子メール保存を進める上でなされなければならないもっとも重要な仕事は、単純だが、その取り組みを進める、または、より深く関与したいと望むアーカイブズ機関や図書館を育て、発展させていくことにある。理論的には、現代の記録資料を収集・保存しているアーカイブズ機関すべてにおいて、電子メールも収集・保存対象となりえる。それを実現するためには、知識を蓄え、情報を共有し、共同で課題に取り組んでいかねばならない。こういった活動は、米国アーキビスト協会（Society of American Archivists）、英国デジタル保存連合（Digital Preservation Coalition）など、既存の組織や体制でもある程度は実現可能かもしれない。障壁の低い課題であれば、それら既存の団体による取り組みで解決可能なものもある。しかし、外部からの支援や関与がなければ、関連ツールやサービスを保存コミュニティ内で醸成していき、電子メール保存の技術開発を長期的に持続可能な形で維持するのも難しいと思われる。それゆえ、本タスクフォースでは、文化遺産コミュニティを支える組織と協力して取り組むべきより高次の活動についても、いくつか提言を行うこととしたい。

6.1.1 障壁の低い／短期的な課題

電子メールコレクションに関して組織的な取り組みの準備ができているか評価する：

保存機関は、電子メールの受入、整理、保存、利用提供といった作業を組織的に行いうるか、判断を迫られる。保存コミュニティでは、その評価方法が必要とされている。現状の機能、職員、ツールで強みとなるのはどこか、見直しが必要なところはどこかが把握できれば、デジタル保存システムやワークフローの組織的な改善に取り組みやすくなるだろう。

- ・ **活動：**全米デジタル管理連盟（NDSA）が定めたデジタル保存の推奨レベルについて、その電子メールに特化した版を策定し、ウェブサイトで公開する（National Digital Stewardship Alliance 2013）。
- ・ **予定：**本タスクフォースのメンバーらは、参加者を募集の上、2018 年夏にワーキンググループを開く予定である。組織的な関与がどのくらいになるか分からないが、外部からの支援をえられるかもしれない。

研修を実施し、スキルの向上を図る：

電子メール保存の論点に関する啓蒙活動やスキル向上の研修が保存コミュニティ内では求められている。多数の保存機関では、まだ電子メールを収集していない。鶏が先か卵が先かという問題である。アーキビストらは、自身が技術的に対応できるようになるまで、そして、寄贈者などの提供者や当該機関の協力者（記録管理者や顧問弁護士）から要望が出されない限り、電子メールを集めようとはしないかもしれない。端的に言えば、アーキビストとキュレーターは、提供者からの信頼、組織内での信頼を勝ち取る必要がある。いいかえれば、責任をもって電子メールの管理ができるということを示さねばならない。同様に、組織の経営層は、電子メールの保存が当該組織に価値をもたらすことを理解しておかなければならない。電子メールのコレクションが手元にはあるが次の段階に進めない保存機関もあれば、電子メールコレクションの受入準備に取り組まねばならない機関もあるかもしれない。

保存機関は、大規模な電子メールコレクションを扱うことのできる人材を見つけ、育てていくのが望ましい。ある程度の専門性が求められるけれども、既存の研修体系を活用し、今いる職員を育てていくこともできるだろう。あるいは、History Lab やウオータールー大学における歴史研究グループのウェブアーカイブといった学際的なプロジェクトも、情報共有や専門スキルの向上といった点で、有用なモデルとなりえるかもしれない。そのような団体であれば、いったん研修の骨格が固まると、持続して実施できるように思える。

- ・ **活動：**主な課題を概観し、利用可能なオープンソースの（時にプロプライエタリな）ツールの実演を行うなど、電子メール保存の基本事項を押さえた、拡張性のある研修・ワークショップのカリキュラムを策定する。入門編は半日コースとなるだろう。本報告書が活用できると思われる。1 日コースでは、ツールの実演（動画音声などを記録する）、参加型の学習も含められることになるだろう。一度形が固まれば、その研修内容は、国際的にも活用できるものとなるだろう。
- ・ **予定：**本タスクフォースのメンバーらは、ボストンで行われる iPres 2018 において、電子メール保存のチュートリアルを実施予定である。会議後には研修資料を公開予定で

もある。また、そこでえられた知見を反映した改定版を、2019 年の国際アーカイブズ評議会（International Council on Archives : ICA）に提出する想定である。さらに、米国アーキビスト協会、米国・州文書館長評議会の電子記録プロジェクト、米国政府アーカイブズ・記録管理者協会（National Association of Government Archives and Records Administrators）などからの支援をえながら、既存のカリキュラムに落とし込んでいくため、研修内容を体系立て、より持続可能性のあるものにしていく。

提供者向けに電子メール保存について分かりやすく説明する：

個人デジタルコレクションの提供者の側からすると、保存対象資料に電子メールが含まれることがいかに重要か、プライバシーやセキュリティがどのように担保されるかといった点で、困惑することが多々あろう。

- ・ **活動：**個別に修正可能な合意書の雛形を策定する。雛形では、提供者と保存機関双方の役割と責任について詳細を記す。具体的には、ワークフロー、機微情報の確認、編集加工の精度、公開猶予期間の想定、検索や利用提供に関する情報が含まれる。
- ・ **予定：**本タスクフォースのメンバーらは、参加者を募集の上、2018 年夏にワーキンググループを開く予定である。
- ・ **活動：**ePADD の選別評価モジュール機能について、アーキビストと提供者が理解しやすいように、研修用の動画を作成する。
- ・ **予定：**ePADD／スタンフォード大学は、ユーザコミュニティとワーキンググループを開き、電子メール作成者や提供者向けに、文書や動画教材を作成予定である。

COPTR での電子メールツール評価の維持管理：

本タスクフォースでは、電子メール保存のツールの洗い出しと分析も行った。ソフトウェアの世界は変化が激しく、随時、機能が追加・削除されていく。そのため、この種の情報は、広く公開し、柔軟な環境で管理していくのが望ましい。

- ・ **活動：**関連ツールの棚卸しリストを、公開ウィキ上で管理している Community Owned digital Preservation Tool Registry（COPTR）へ移行する。
- ・ **予定：**本報告書を刊行後、本タスクフォースのメンバーらは、COPTR に連絡を取り、棚卸しリストを COPTR へ移行する手筈を整えるとともに、継続して更新していけるよう、持続可能な計画を策定する。

ファイル形式の比較表を策定する：

ファイル形式の変換作業が保存ワークフローの一部に組み込まれている場合、ある特定の形式を選ぶことになるわけだが、その利点と影響はどのようなものになるか。MBOX と EML は、採用されたツールに依存する面もあるが、電子メール保存において事実上の標準扱いとなっている。だが、他にも選択肢はあり、例えば、XML ベースの Email Account XML Schema（EAXS）が挙げられる。ファイル形式の比較表があれば、保存機関において今後の

計画やワークフローが検討される際、形式変換に伴うリスクについて理解しやすくなる。比較表の作成作業は、既存の取り組みからも知見をえられる。例えば、連邦機関デジタルガイドライン事業（Federal Agencies Digital Guidelines Initiative : FADGI）では、静止画や動画のファイル形式が比較検証されているが³⁴、有用な情報源となろう。えられた知見がいったん形としてまとめれば、NDSA や英国デジタル保存連合（DPC）などにおける取り組みの一環として、作成された表の維持管理が行われていくことも期待される。

- ・ **活動**：ファイル形式の評価表を策定する。表には、ファイル形式の構造、標準文書、技術メタデータ、ヘッダ項目、想定される挙動などの情報が含まれる。
- ・ **予定**：この作業は、Test Existing Tools for Data Impact and Data Loss プロジェクトの成果にもとづき、まとめられることになる想定である。また、当該プロジェクトに従事したものが継続して関与していく予定である。

6.1.2 影響の大きい／長期的な課題

電子メールの保存コミュニティを持続させる

電子メールの保存ツールを扱うコミュニティを、長期的な観点から、どのように醸成していくか考えていく上で、本報告書はよい契機となりうる。アーカイブズ機関、博物館、図書館にとって有用なオープンソースのツールの中には、コンソーシアムの形態で維持管理されているものがある。また、外部資金にもとづくプロジェクトも出始めている。よく知られた例としては、BitCurator、CollectionsSpace、そして ArchivesSpace が挙げられる³⁵。それらのソフトウェアは自由に利用できるものの、開発に携わりたい、あるいはサポートを受けたい場合には、当該コンソーシアムに組織は参加せねばならない。

しかし、少なくとも現在のところ、電子メールの保存コミュニティにおいて、このモデルが模索されているようには見えない。オープンソースの電子メール保存ツールは既にいくつか存在するが、いずれも異なるニーズに応えようとするもので、それぞれ異なる機能をもつ。例えば ePADD は、特に収集アーカイブズにおいて、電子メールを保存する際に広く利用されており、そのコードはオープンソースとして GitHub 上で公開されている。あるいは、TOMES は、ePADD の少し後から開発が始められたもので、組織アーカイブズにより適しているとされる（Gibson 2018）。これも、ソースコードが GitHub 上で公開されている。ただし、いずれのプロジェクトも、継続的に開発できるかどうかは外部資金に依存している。スミソニアン協会アーカイブズは、DarcMail のソースコードを自身のウェブサイト上で公開するとともに、GitHub でも入手できるようにしている（ハーバード大学の EAS もいずれ公開されるだろう）。本タスクフォースでの話し合いでは、これらの機関や他のプロジェクトでえられた情報や知見を共有すべきことが確認された。付録では、様々なツールや実践プロジェクトを紹介しているが、それらで抜け落ちている側面を今後検討していく上で、本報告

³⁴ FADGI, “Guidelines: File Format Comparison Projects – Still Image and Audio-Visual Working Groups,” December 2, 2014, http://www.digitizationguidelines.gov/guidelines/File_format_compare.html.

³⁵ BitCurator コンソーシアムの運営管理は、非営利団体の Educopia が担っている。コンソーシアムへの参加費用は、1 機関につき年間 2,000 ドルである。ArchivesSpace と CollectionsSpace は、非営利団体の Lyris が運営管理している。コンソーシアムの参加費用は段階的で、参加機関の規模と予算に応じて決められる。年間費用は、460 ドルから 1,725 ドルの 4 段階で設定されるが、運営により関与するカテゴリ Leaders Circle への参加の場合は年間 2,500 ドルとなる。

書が一つのきっかけになればと願う。

- ・ **活動**：主なオープンソースツールを開発しているプロジェクトの責任者らが集まり、より統合的なツール開発に必要な高次の機能要件を策定するプロジェクトを立ち上げる。このプロジェクトでは、個別のツールやサービスに必要な支援策がまとめられることで、短期的な資金調達モデルの提案が可能となろう。あわせて、長期的な支援を行う組織をつくる第一歩ともなりえよう。
- ・ **予定**：プロジェクトの計画を策定するとともに、資金源を探る。

電子メール保存ツールの要件を検討する：

繰り返しになるが、州政府や大規模組織の多くは、企業向けに開発された電子メール保存ツールを利用している。あるいは、クラウドベースの企業用システムの一機能として、電子メール保存ツールを利用している場合もある。州政府のアーキビストによれば、そのようなツールを改修し、機能を追加することで、対象記録を収集・管理するのがより便利になるだろうという。これは、経営管理層のアカウントや、特定案件に関連した電子メールを収集する場合についても当てはまる。

- ・ **活動**：保存コミュニティ主催で、短期的なプロジェクトやイベントを開くことが考えられる。おそらく、米国州政府最高情報責任者協会（National Association of State Chief Information Officers : NASCIO）、米国州政府文書館長協議会（Council of State Archivists : CoSA）、米国国立公文書記録管理局（NARA）、学術コミュニティと連携して進められることになる。共同して取り組むことで、機能要件が洗い出される。それにより、リスク管理、情報開示、他の業務システムとの統合、そして、経営・法規・歴史的な観点から保存すべき記録の収集を進める上で、電子メールの保存ツールが役立てられることになる。
- ・ **予定**：このアイディアを追求することに関心が寄せられるのであれば、本タスクフォースのメンバーらは、CoSA、NARA、NASCIO と連携して、後続プロジェクトの立ち上げに着手する想定である。運営会議を実施するに当たっては、外部機関からの支援が欠かせないだろう。

電子メールの真正性に関する判断規準を策定する：

一連の保存作業の過程で、電子メールのデータは移行・変換・処理されるわけだが、その際、データ自体には変更がなく、欠損がないことを担保するためには、さらなる調査検証が必要である。そのような取り組みの主たる目標は、ツール環境の改善である。メッセージやアカウントのプロパティが真正なものであるか検証する際に用いられるプロファイル情報やスキーマを策定することもおそらく含まれる。

保存コミュニティにおいては、その前提として、何をもって「真正な電子メール」といえるのか、その判断基準に対する共通理解が求められている。例えば、電子メールのヘッダには、署名や真正性を検証することに関連した項目が多数含まれ、それらは伝送時点の真正性を証明するために利用されている。だが、そういった項目を長期にわたり保持しておく必要はどのくらいあるのだろうか。配布リストや bcc 宛先が欠落したメッセージは、真正なもの

かもしれないが、不完全なものではないのか、あるいは、別扱いものとなるのか。元々の文脈で再現されたメッセージの方が、より真正なものと見なされるのか。何をもって真正とみなすかの要因となりうるものを、どのように残しておくことができるのだろうか。このメッセージは変更されていない可能性が高いとみなせるかもしれない証拠が積み重なっていれば、利用者側もその存在を把握し、解釈できる場合もあるだろう。

- ・ **活動**：2012 年に、InSPECT プロジェクトでは、電子メールのメッセージのうち重要なプロパティを判定し、Thunderbird や Outlook などのテスト対象システムからデータを出力または移行する際に、それらプロパティを残しておくべきか判断するための検証手順を策定した (Knight 2010)。基本的に妥当なやり方ではあるが、ツールや電子メールの形式が更新されていく (例えばヘッダ項目の新規追加) に従って、手順も更新していかなければならないだろう。
- ・ **予定**：この作業は、本タスクフォースが推奨する別の活動、つまりは既存ツールが処理時にデータへ影響を及ぼす可能性 (データ欠損など) の検証と並行して実施できるものの、真正性の問題は、それに集中し、研究課題として長期にわたり取り組むべきものだ。そして可能であれば、外部資金をえて、実務家を巻き込みつつ、iSchool のプログラムと共同して実施されるのが望ましい。

電子メールの研究データとしての価値を示す：

エンロンの電子メールコーパスなど、公開された電子メールのデータを用いた研究プロジェクトは既にある。だが、歴史研究者などにとって、電子メールが豊富な情報源であることを示す事例を増やすには、さらなる作業が欠かせない。デジタルヒューマニティーズ分野の研究者、歴史家、データサイエンティストらが、電子メールのコレクションを存分に利用できる環境を整えることができれば、(個人や組織の) 電子メールという媒体ならではの分析方法による、学際的な研究が期待できよう。歴史研究のコミュニティで理解が広まれば、Jisc の研究データ検索サービス (Research Data Discovery) などのプロジェクトと電子メールの研究とを融合させることにつながるかもしれない。

- ・ **活動**：研究目的で電子メールのデータを利用したい機関を対象として、歴史家などが「エンベディッド研究者」となってくれる (実践領域にも深く関与してくれる) ような、データ関連の研究プロジェクトを立ち上げる。関与した研究者には、研究成果とともに、電子メールのデータなくしてこの成果はえられなかった理由についても公表してもらいたい。候補機関としては、HILT Institute (Humanities Intensive Learning and Teaching)、米国歴史学会 (American Historical Association) が挙げられる。

MBOX に関する IETF RFC 標準を修正する：

MBOX について規定された IETF RFC 4155 の現行版では、MBOX の派生形式までは十分に扱えていない。少なくとも MBOX には 4 つの下位タイプがあり (MBOXO、MBOXRD、MBOXCL、MBOXCL2)、これらは共通の MBOX 構造にもとづき、つくられている。ある派生形式用のツールは、別の派生形式用ツールと互換性があるとは必ずしも限らない。RFC 標準文書の内容を明確に書き換えることにより、当該ファイル形式の標準化の程度がより確

かなものになろう。また、当該ファイル形式を識別・分類する精度が向上し、ツールの相互運用性を高めることにもつながる。

- ・ **活動**：IETF と連絡を取り、RFC の改定手順について確認する。RFC 4155 の原作者や他の寄与者にも連絡を取り、仕様改定に関するワーキンググループを立ち上げる。

EML に関する標準を修正する：

EML 形式は、インターネットメッセージフォーマット (IMF) を規定した IETF RFC 5322 の中で定められている。IMF では、メッセージはすべてアスキーコードの文字列で表現される。EML は、IMF の下位形式であり、マイクロソフトの Outlook Exchange やアップルの Apple Mail などの他の電子メールプログラムで利用されている。EML は、電子メール保存において広く使用されている形式であるものの（例：ハーバード大学の EAS システム）、その標準文書自体は公開されていない。

- ・ **活動**：IETF と連絡を取り、RFC の新規策定手順について確認する。マイクロソフトを含め、関係者と連絡を取り、公開仕様の策定に向けたワーキンググループを立ち上げる。

電子メールの保存ワークフローにおいて PDF の位置付けを見直す：

電子メールのメッセージを PDF として出力するという機能は、多くの電子メールアプリケーションに既に備わっている。しかし、重要なヘッダ項目や他の技術メタデータまでは出力されないことが多い。あるいは、形式の変換時にデータが隠されてしまうことが多い。さらに、スレッドの情報や添付データとの関係も失われてしまう。PDF ソフトウェアの機能、特に電子メールアプリケーションに埋め込まれた機能を改善することにより、電子メール保存のワークフローを大幅に単純化できるかもしれない。

- ・ **活動**：ベンダー中立の国際的組織である PDF Association と協力し、電子メール保存において PDF 形式に求める要件を整理する。
- ・ **予定**：本タスクフォースのメンバーらは、PDF Association に連絡を取り、2018 年秋にはこのプロジェクトを開始する。

6.2 ツールのサポート・検証・開発

例えば ePADD、EAS、DarcMail、TOMES などのツールは相互補完的な関係にあり、それぞれ、収集アーカイブズ、機関アーカイブズ、政府機関などのニーズを満たしている。この意味で、各ツールはそれぞれの役割を果たしている。しかしいずれも、親機関（そして、比較的規模は小さいだろうが、協力機関）による支援、あるいは政府系助成機関からの資金に大きく依存しており、この先どうなるかは不確かである。加えて、多くの機関では、電子メールを扱うのに商用のツールが利用されている。それら企業向けツールには、アーキビストにとって有用と思われる機能が含まれている。大きくは2つあり、ジャーナリング機能と法令遵守対応機能であるが、いずれも費用が高すぎるため、アーキビストには手が届かない。しかし、もしそれらがより入手しやすくなったら、あるいは、そのオープンソース版が開発されることになったら、データの収集と機微情報の確認という困難な作業は大幅に改善されることだろう。

コレクションの規模が小さかろうが大きかろうが（あるいは、多数のアカウントが含まれたコレクションだろうが、単一のアカウントに多数のメッセージや添付データ含まれているものだろうが）、問題なく動くツールが望ましい。突き詰めれば、今時点で大規模なものが、明日にはそうでなくなりうるため、アーキビストにとっては、必要に応じて性能や容量を変更可能なツールが必要といえる。次に示す提言は、資金提供者とソフトウェア開発コミュニティ双方に向けたものである。

6.2.1 障壁の低い／短期的な課題

データの欠損や変更に関する既存ツールの検証：

電子メールを保存するワークフローにおいては、基本的に一連のツールが欠かせない。オープンソースのものもあれば、プロプライエタリなものもある。特にファイル形式の変換時に、それらツールが及ぼす影響については、評価も文書化もまだなされていない。例えば、技術メタデータやヘッダ項目は、追加・削除・変更されてしまわないか。ツールを利用する順番によって何か影響が及ぼされないか。特定のファイル形式に関しては、あるツールの方が他のツールよりうまく処理できる場合はないか。あるファイル形式の方が他よりうまく処理できることはないか。どのくらいの量のエンベロープが維持されるか。手始めに電子メールファイルの形式変換時に焦点を当てた調査を行い、次に、データの読み込み時の影響調査を実施する。この基礎的な作業成果により、次の目標が明確になっていくだろう。例えば、真正性の定義を策定すること、電子メールのデータモデルを作成すること、ツール開発の要件を明確にすることなどが挙げられる。

- ・ **活動：**様々な保存機関から多様な電子メールを集める。標準的なヘッダ項目、拡張ヘッダ項目が使用されている電子メールを含む。調査対象とするメッセージを様々なツールで処理後、ヘッダ情報に変化がないか比較を行う。データの欠損や変更がないか確認の上、個々のツールを評価し、ワークフローの改善案やツールの改修案について提言する。この作業を進めるに当たっては、InSPECT プロジェクトの成果が参考となろう。
- ・ **予定：**本タスクフォースのメンバーの一部がプロジェクトの計画を立てる。また、異なるアカウントから集めた多様な電子メールのデータセットを対象とし、ファイル形式を変換するツールがもたらす影響について探るため、外部資金の申請を行う。

電子メールのファイル形式を識別・分類・検証するツールの改善を行う：

保存コミュニティは、電子メールのファイル形式を高精度で識別・分類・検証するため、より精緻で柔軟なツールを求めている。相互運用性を高めるためには、識別・分類・検証といった各側面に特化したツールというよりも、既存ツールや関連アプリケーションに含まれたそれら機能を統合していく方が肝要である。この作業は、（前述した）データの欠損や変更に関する既存ツールの検証プロジェクトに続いて実施されることになる。

- ・ **活動：**ファイル形式の分析や真正性に関するプロジェクトが完了後、ファイル形式の識別・分類機能の向上を目的とした作業として実施される。JHOVE、Siegfried、Apache Tikaなどの広く使用されているツールが具体的な対象となろう。

6.2.2 影響の大きい／長期的な課題

機微情報の確認ツールを改善する：

保存機関が直面する喫緊の課題の一つに、機微情報の確認作業が挙げられる。個人情報や機微情報を機械的に特定・削除・編集加工・制限できる、オープンソースのより高度なツールが求められている。現状のツールは、社会保障番号や電話番号などの構造化された個人情報を対象としているが、学歴（「家庭教育の権利とプライバシーに関する法」における保護対象³⁶）や医療記録（「医療保険の相互運用性と説明責任に関する法」における保護対象³⁷）などの構造化されていない情報までは扱えていない。

電子メールの機微情報確認で用いられる自然言語処理ツールは、機械学習にもとづく性能向上が期待されている。それにより、機微情報に該当するか微妙なものについても、識別・抽出することが可能となろう。現状のワークフローでは、自然言語処理にもとづき、人名、団体名、地名などが識別されているが、これは固有表現抽出の技術に依存している。Wikipedia のカテゴリと照合される場合もある（ePADD では、複数アカウントを対象に照合が行われる。Wikipedia カテゴリとの統合とその拡張が次の課題となるかもしれない）。あるいは、ePADD や TOMES においても確認できるが、辞書を作成することにより、人間が介在した作業効果を高めることができよう。機械学習は完全なものといえないものの、複雑で大規模なデータセットを対象とした実験において、学習機能を組み込んだ顔認証システムの認識精度は高いことが証明されている。この技術を電子メールに適用できない理由はない（Phillips 2018）。機微情報を含むメッセージの特定には、人による目視確認の方がよいとする仮定はあくまで憶測に過ぎず、検証も実証もなされていない。それゆえ、機械で確認する可能性を排除すべきでない。

機密解除の広がりや法曹・法務業界の動きは、機械学習の活用を促すものとなろう。米国の公益的機密解除委員会（Public Interest Declassification Board : PIDB）は、2012 年に報告書を刊行しており、米国大統領に対し、「機密指定および機密指定解除のあり方を刷新するため、共同作業を促すこと、既存技術の活用方針を決めること、新手順の策定・検証を行うこと」を要望している。具体的には、「機密解除や情報管理の作業を支援するとともに、歴史家や関心をもつ市民が利用できるにするために、タグ付与、インデキシング、検索、大容量ストレージ、検索インタフェイス、他のルールベースのアプリケーション」について検討することを求めている（Public Interest Declassification Board, 2012, 26）。同様に、EU 一般データ保護規則の第 17 条では、消去の権利または忘れられる権利が規定されている。そこでは、データ主体がデータ管理者に自身の個人データを消去させる権利、データ主体が自身の個人データの配布を停止できる権利、また、データ主体が自身の個人データを第三者により取り扱われないようにする権利、が定められている。この規則が電子メールの保存にどのような影響を及ぼすかは、まだ十分に整理されていないが、定められた条項を見る限り、機微情報の確認やその編集加工といった面で重大な制約が課されていることは疑いえない。このような課題についても保存コミュニティは注視していく必要があるだろうし、逆にいえば、

³⁶ 家庭教育の権利とプライバシーに関する法（Family Educational Rights and Privacy Act : FERPA）
<https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>.

³⁷ 医療保険の相互運用性と説明責任に関する法（Health Insurance Portability and Accountability Act : HIPAA）
<https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html>.

個人情報や機微情報を機械的に特定・削除・編集加工・制限できる、オープンソースのより高度なツールを実現させていく契機ともなりえよう。

こういった機密解除の広がりや法曹・法務業界の動きの中で、既存ツールの検証を持続的に実施していくことが重要と思われる。次いでその成果が、保存コミュニティでも活用するオープンソースのツールに反映されていくことことになるだろう。とはいえ、どのような機能が求められるのか。機械支援型レビューについて、法曹・法務業界で蓄積された知見とどのように比較検討すればよいか。どのツールがもっともよいのか。どのような機能が実際の利用に耐えられるのか、それはどのような文脈においてか。このような作業にもとづき、不足事項と優先項目が洗い出され、要件が明確化されていく。

- ・ **活動：**ノースカロライナ州立文書館では、TOMES 2.0 での分類・確認作業を支援するため、機械学習ツール（Google が開発したソフトウェアライブラリ TensorFlow）の活用を検討している
- ・ **活動：**イリノイ大学では、電子メールを分類し、制限対象資料を特定するため、企業向けプレディクティブ コーディング用ツールの評価を行なっている。
- ・ **予定：**本タスクフォースメンバーらは、既存ツールを評価し、オープンソースの機械学習分類ツールの要件を策定することを目的とし、プロジェクト計画の作成と外部資金の獲得に関心を寄せている。これは、自然言語処理など、他のコミュニティと共同で取り組むことのできる領域かもしれない。

既存ツールを維持し、統合する：

電子メールを保管・収集するやり方は多数ある。同様に、組織単位やコレクション単位での分析ニーズも様々であろう（電子メールの提供方法もまた多数考えられる）。そうすると、多様なワークフローやツールが導入されていることは、何も驚くべきことでない。だが、上述した通り、共通項もいくらか見出され始めている。あるいは、相互補完的なアプローチがあると捉えることもできよう。一方では、私的な個人の電子メールを収集・保存することに関心をもつところもあれば、他方では、政府、大学、企業といった組織の記録として電子メールを位置付けているところもある。それら両部門で採用されるツールの違いは、保存対象となる記録がもつ性質や歴史的な経緯を反映するものだろう。組織アーカイブズは、収集アーカイブズと異なることが常である。いずれにせよ、ツールを完全に一本化してしまうようなことは、理想的でもなければ、その必要性もない。

DarcMail、ePADD、EAS、TOMES は、将来的な統合や連携を見据えた取り組みが行われるにふさわしいツールといえる。もちろん、支援対象にふさわしいツールかどうかは、コミュニティ形成の状況次第であるが、そのことと持続可能性とは必ずしも結びつかない面もある。現状のワークフローの違いを埋め、ツール間の相互運用性をより高めるには、複数機関、関連ソフトウェア企業、資金提供機関などによる計画的な取り組みが欠かせない。これらのツールが、既存のコンソーシアムやプロジェクトに統合されていくことは可能なのだろうか。もしできないのであれば、各組織がそれぞれ違うツールを維持管理する状況が続くことになる。しかし、各組織で異なるツールを管理し、各々がそれらに資金を投下していく妥当性をどこまで主張し続けられるだろうか。似たような話として、本タスクフォースが関

連企業と一時的に情報交換した中では、別ツールの統合や開発が進むことにより、複数機関の関与を促す汎用的なサービスが生まれてくるのではないかと、という見方も示された。

- ・ **活動**：スタンフォード大学において、開発課題が挙げられた際、集約された検索サイトの必要性について言及した実務家がいた。そのような検索サイトでは、整理された膨大なデータを研究目的で利用できるように公開するため、複数機関がアップロードできる形が望ましい。ポータルサイトが構築されれば、オンライン上でのデータ公開に関心はあるものの、自身らでサーバ環境を用意できない機関が抱える課題の一つを解決することにつながるかもしれない。スタンフォード大学の担当者らは、そのようなサイトを内部で運営できないか検討を始めている。先に挙げた、電子メールの研究データとしての価値を示す活動の成果は、電子メールコレクションを広く公開し、認知度を高めていく上で、基盤的な役割をもつことになろう。このような取り組みを進めるに当たっては、外部からの支援や協力が重要となってくる。
- ・ **活動**：ノースカロライナ州立文書館では、TOMES の改修を進める上で、追加の外部資金獲得を計画している。次の段階では、政府資料の保存に焦点が当てられるとともに、個人資料についても検討がなされる予定である³⁸。TOMES では、他の州立文書館や大学アーカイブズ、収集アーカイブズとの連携も模索されている。
- ・ **活動**：2017 年の夏以来、ハーバード大学図書館では、オープンソースのソフトウェアとして、EAS の再構築作業が進められている。本報告書執筆時点では、ロードマップと技術面での計画が立てられ、大学の他のシステムとの統合可能性に関する技術的な検証作業が実施されている。現状の資金が確保できれば、ソースコードの公開は数年後となる想定である。もし追加の資金がえられれば、開発日程を前倒しすることは可能だろう。また、ハーバード大学では、ツール間の互換性が担保できるよう、既存のオープンソースプロジェクトと協力していきたいと考えている。
- ・ **活動**：スミソニアン協会アーカイブズでは、DArcMail の改修を行い、2017 年 12 月に公開した。整理・保存用ツールでは、MySQL に加えて、SQLite にも対応した。この改修により、小規模な組織での運用改善が見込まれる。ツールと文書ともに、スミソニアン協会アーカイブズのウェブサイト上で、オープンソースとして公開されている。
- ・ **予定**：本報告書で触れたツールの開発計画や、相互運用性や連携向上などのその他活動が継続して実施されていけば、ツール開発機関としても恩恵を受けることだろう。おそらく 3～5 年程度あれば、親機関、協力機関、外部機関からの支援を受ける形で、電子メール保存ツールのコンソーシアムを立ち上げることも可能かもしれない。

セルフアーカイブ用のツールを開発する：

本報告書で言及した電子メールの収集・整理・保存ツールの大半は、記録管理者、法務担当、あるいはアーキビストのニーズを満たすためにつくられてきた。一方で多くの人が、自身のデジタル履歴を自身で残したいと考えている。以前とあまり変わらないのかもしれない

³⁸ TOMES 2.0 プロジェクトでは、機械学習を組み込み、フラグ付与機能の改善が予定されている（保存対象外の資料を分離し、アーキビストが提供用ファイルを作成できるツールを構築するため）。当プロジェクトでは、ePADD の提供機能との連携が模索されている。

い。手紙はいったん引き出しにしまいこまれ、保存機関へすぐに寄贈されることはない。同じように、職員は自身の電子メールのコピーを取っておきたいと考えているかもしれない。そうすれば、残されたアカウントに埋め込まれた組織の記憶が後任者により検索され、そこから何かをえることにつながるかもしれない。いいかえれば、自身で電子メールを収集・保存しておくといった、これまでにないサービスが成り立つかもしれない。これは、自身が有用と思うウェブサイトや他の情報資源を収集できるインターネットアーカイブの取り組みと同様だ。ローカルで動くソフトウェアとして実現される可能性もあれば、あるいは、こちらの方がよりありえそうだが、ウェブ上のサービスとして実現される可能性もあろう。収集されたデータは、提供者の管理下に置かれるとともに、他者と共有することもできよう。また、出力機能や他の手段を介して、最終的には保存機関へ寄贈できるような設計にしておくこともありうる。

- ・ **活動と予定:** このアイディアに関心が寄せられるのであれば、本タスクフォースのメンバーの一部は、既存ツールの評価とサービスの構築を行うため、計画書を作成し、外部資金の獲得を狙う。この取り組みを進めるに当たっては、企業や研究機関から支援をえられる可能性もあり、あるいは、ePADDのようなツールを流用する可能性もありうる。

ツールの相互運用性に関する標準を策定する（参考実装としても提供する）:

電子メールを保存する上では多数のシステムとツールが用いられている。しかし、それらシステム・ツール間の相互運用性は体系化されていない。そのため、次の取り組みを進める上で必要な標準が、保存コミュニティにおいてはまだ合意されてない。具体的には、1) セキュアかつデータの完全性を担保した仕組みで電子メールのコレクションを交換する場合、2) 他のツールでも扱えるよう、メタデータを電子メールコレクションに追加する場合、3) 複数ツールで処理されたコレクションの管理記録全体を残す場合、である。

この問題を解消するために利用されている、あるいは策定された標準は既に多数ある。例えば PREMIS は、一連の管理過程を記録するためのメタデータ標準で、(相対的に) 広く利用されているものだ。研究データ同盟 (Research Data Alliance) には、機関間でデータコレクションを交換するための、標準 API 仕様を策定するワーキンググループが設けられている。他にも、メタデータのアプリケーションプロファイルや、ファイルのパッケージ化に関する標準が挙げられる。

標準は、広く実装され、維持管理されない限り、うまくいかない。それゆえ本タスクフォースでは、標準の策定とともに、既存ツールに策定した標準を参考実装として組み込むプロジェクトも提案したい。実装に協力してくれる機関とともに標準を策定することによって、より実用的で実行可能性のある成果をえられるだろう。参考実装のアプローチを取ることによって、それらツールを使って連携したい利用者同士が即座に検証することができる。また、ここでえられた知見を標準に反映していけば、コミュニティ全体へ還元することにもつながろう。

- ・ **活動:** 既存の標準を検証し、課題を洗い出す。API に対するニーズなども含む。電子メール保存ツールの相互運用性を担保する上で、必要なデータモデルや主な (望ましい) 標準について合意をえる。電子メール固有のニーズに対応するため、(必要があれば)

それら標準の拡張を図る。既存ツールの開発者と協力し、それら標準の実装を進め、相互運用性の高いワークフローを実現する。

- **予定:** 本タスクフォースのメンバーの一部は、電子メール保存ツール間の相互運用性の向上に必要な標準を策定し、その標準をツールに実装することを目的としたプロジェクトを立ち上げるため、計画書を作成し、外部資金の獲得を狙う。

■ 付録 A : システム処理の自動化

ハーバード大学図書館では、出所来歴の情報を記録しておくため、作業プロセスのデータを自動的に収集している。記録されたデータは、PREMIS データモデルのイベントという実体にマッピングして管理される。メタデータは、アイテム／メッセージ単位、パケット単位で記録される。次に示す通り、ここでのパケットとは、ハーバード大学の電子メールアーカイビングシステム（EAS）への投入単位であるが、おおむねコレクションに相当する単位である。

Events in EAS

Associated docs	g:\LTS\LDI Systems\EAS\specs\deletions_wg_20120416_PM.docx
-----------------	--

Events in EAS are a means of tracking the processing history of items in EAS.

Current events created in EAS are:

Event Level	Event Content Type	Event Type	Event Value	Agent
Item (1)	email message	normalization	extracted from a file in {0} format and converted it to {1} format e.g. <i>extracted from a file in Outlook for Windows/unknown format and converted it to RFC-2822 eml format</i>	Emailchemy
Item	email message	normalization	rewrote line breaks	EAS
Item (2)	email message	normalization	removed embedded files	EAS
Item	email message	association	associated external attachment [{0}]	EAS
Item	email message	delete component	deleted attachment [{0}]	person
Item (3)	email message	metadata update	AdminFlag removed: [{0}]	person
Item	email message attachment	normalization	extracted from a file in {0} format and converted it to {1} format e.g. <i>extracted from a file in Outlook for Windows/unknown format and decoded</i>	Emailchemy
Item	email message attachment	delete component	deleted email message [{0}]	person
Packet	email message	delete component	deleted email message [{0}]	person
Packet	email message attachment	delete component	deleted attachment [{0}]	person

Notes:

1. This event modifies the email message by adding a header e.g.:

X-Converted-By: Emailchemy 11.2.2 Embedded Edition; licensedTo="Harvard_Library_1"

2. This event modifies the email message by adding a header:

X-Converted-By: EAS 0.9.2; conversionDate=2013-01-09T19:20:07 UTC

It also removes the embedded attachment and modifies the embedded content header and content e.g.:

-----07000608000030409010704
Content-Type: text/plain; charset=UTF-8
X-EAS-INLINE-ID: 970492
X-EAS-MESSAGE-DIGEST: type=md5;0b330856810e35dd98289225bceffed1

Start EAS original body part headers
Content-Type: image/png; name="pre_processing_flow.png"
Content-Transfer-Encoding: base64
Content-Disposition: inline; filename="pre_processing_flow.png"
End EAS original body part headers

EAS Inline Converted: extracted to external location; pre_processing_flow.png; cid:

3. This event is not written to DRS as an event but is written to huiDrsAdmin/adminFlag/flagID

■ 付録 B : 図書館、アーカイブズ機関、博物館向けの電子メールツール

電子メールというツールがこれまで発展してきた原動力は、民間部門と、記録を管理したいというニーズから生まれてきたといえるが、過去 10 年を振り返ると、図書館、文書館、博物館（LAM）での取り組みも確認できる。電子メールの管理ライフサイクル上の課題に対応できるよう、

ワークフローが整備され、新たなツール類が作成されてきた。このような取り組みの目的は、経営管理上の価値または歴史的価値のある電子的やり取りを保存し、長期的にアクセス可能な状態にしておくことにある。デジタル保存のワークフロー上に電子メールツールを再定義・統合することに加え、文化遺産機関においては、デジタル保存特有の要件を満たす新しいツールが開発されてきた。あわせて、電子メール保存の手法を改善することが目的とされたプロジェクトも多数立ち上げられ、多くの参考とすべき事例が生み出されてきた。LAM コミュニティの開発者は、保存要件を満たす形で、メッセージやメッセージフォルダをパースしたり、変換したりするため、プログラミング言語の電子メール向けライブラリ（例えば、Python の email や mailbox モジュール）を活用している。近年では、Preservica や Archivematica などのデジタル保存システムも登場してきた。コンテンツの受入から保存システムへの格納まで、全体のワークフローを網羅的にカバーしている。いずれもマイクロサービス設計となっており、ワークフローの各段階に特定の機能（と関連ツール）が組み込まれている。LAM においてはプロプライエタリなツール（や Preservica のようなシステム）も利用されているが、本文で触れたように、オープンソースのツールも活用されだしており、それらはデジタル保存特有の要件にあうよう修正・改修され、全体のワークフローやシステム内に組み込まれている。電子メールのツールに関して、文化遺産の領域以外で利用されているものを含む、より包括的なリストについては、次の URL を参照されたい。

Archivematica

基本情報

- ・ 開発元 : Artefactual Systems
- ・ リンク先 :
<https://www.archivematica.org/en/>
https://wiki.archivematica.org/Main_Page
- ・ 入手条件 : システムは自由に利用でき、オープンソースである。Archivematica のソースコードはすべて、GNU Affero General Public License (A- GPL 3.0) のもと配布されている。
- ・ システム要件 :
 - ・ OS :
 - ・ Ubuntu 14.04.5 64 bit Server Edition
 - ・ CentOS 7.3.1611 64 bit
 - ・ Ubuntu 16.04 への対応予定

- ・ ハードウェア要件（最低要件）
 - ・ プロセッサ：2 CPU コア
 - ・ メモリ：4 GB +
 - ・ ディスク容量：20 GB + 処理対象となるコレクションの3～4 倍の容量が必要
（例：50 GB のデータ移行時には 200 GB が必要）
- ・ 現状：Artefactual Systems により開発・維持管理されており、ユーザコミュニティも拡大中。2018 年 5 月にバージョン 1.7 が公開された。

ツールで何ができるか？

- ・ 目的：Archivematica はマイクロサービス設計となっており、「ISO の OAIS 機能モデルに準拠し、データ投入から利用提供まで、デジタルオブジェクトに関する一連の処理を行うことが可能な統合ソフトウェア」である。さらに、METS や PREMIS に対応しており、記述・管理・権利メタデータを扱うことが可能となっている。受入段階で、コンテンツは Archivematica の管理領域（backlog）に配置されることになるが、その前に次の処理が実行される：ファイル単位で UUID とチェックサム値を生成、ウィルスチェックと隔離、ファイル形式の特定・分類・検証、技術メタデータの抽出、機微情報・私的情報を識別するためのフォレンジックツール（bulk_extractor）のデプロイ。選別評価の段階では、データの中身や技術的側面について確認・分析が実施できる。ArchiveSpace と連携させることにより、記述作業とコンテンツ自体との紐付けも実行できる。データ投入段階では、適切なプラットフォームヘデータを引き渡すことができるよう、保存用情報パッケージと配布用情報パッケージが作成される。あわせて、保存・提供のため、ファイルの正規化もこの段階で行われる。
- ・ 保存管理のライフサイクル上の位置付け | 整理・保存：Archivematica は、受入時ににおいて、ネイティブ形式からの正規化（少量の場合）、あるいは一括処理が実行される。これにより、ファイル形式の特定・分類・検証、技術的メタデータの抽出、個人情報や機微情報の特定が可能となる。アーキビストによる手動での編成作業にも対応している（MBOX や EML ファイルを保存メタデータと関連づけるなど）。また、保存システムへ投入するためにコンテンツをパッケージ化する作業にも対応している（Storage Service 機能により、格納場所を追跡可能）。ただし、Archivematica 自体には、オンライン上での開示・検索機能が組み込まれていない。
- ・ 強み・弱み・課題：強みは、マイクロサービス設計となっている点、パイプライン処理に似たアーキテクチャとなっている点が挙げられる。これにより、標準化された手順を介して、作業過程ごとにコンテンツを移動させながら処理していくワークフローを構築できている。また、コンテンツの真正性と完全性を長期的に担保できるように、基本的な保存機能を多数備えている（かつ、作業記録を網羅的に保存メタデータとして記録している）。ただし、電子メールの処理という面ではいくつか課題を抱えている。例えば、添付データをメッセージと切り離し、別途保存することができない。また、電子メ

ールのファイル形式を正規化する際、変換対象が限定的という制約もある。この意味で、Archivematica は他のツール（例えば ePADD）と組み合わせるのがよいかもしれない。そうすれば、より高度な、電子メールに特化した運用が可能となり、さらに、コンテンツをパイプライン処理に乗せていけば、他のツールで保存作業が補完されることになる。加えて、Archivematica には高度な検索やデータ分析機能も用意されていない。

ファイル形式

- ・ 投入時：

- ・ MBOX
- ・ PST
- ・ Maildir

他のファイル形式も投入可能だが、デフォルトの保存機能しか利用できない。

- ・ 出力時：

- ・ MBOX
- ・ 元のファイル形式

DArcMail (Digital Archive Mail System)

基本情報

- ・ 開発元：スミソニアン協会アーカイブズ（Smithsonian Institution Archives：SIA）；当アーカイブズとユーザコミュニティで維持管理
- ・ リンク先：<http://siarchives.si.edu/blog/yes-we%E2%80%99re-still-talking-about-email>
- ・ 入手条件：ユーザガイドとソースコードのダウンロードは次の URL から可能
<https://siarchives.si.edu/what-we-do/digital-curation/email-preservation-cerp>.
- ・ システム要件：
 - ・ OS：Windows、Mac、Linux
 - ・ Python
 - ・ リレーショナルデータベース：MySQL、SQLite
 - ・ ワークステーション／クライアント版とサーバ版あり
- ・ 現状：現行バージョンは 1 である。スミソニアン協会が今時点では維持管理しているが、第三者も開発に参加して欲しいと考えている。

ツールで何ができるか？

- ・ 目的：初期の選別評価、保存（AIP）、利用提供（DIP）を実行する目的で設計されたツールである。AIP と DIP 双方の形で元のアカウントデータを保持できる。とはいえ、柔軟な対応が可能で、データのサブセットを生成し、AIP と DIP をそれぞれ修正することもできる（Simpson 2016）。
- ・ 保存管理のライフサイクル上の位置付け | 整理・保存・開示・提供：

DArcMail では、電子メールの正規化、アイテム単位での整理、一括処理、手動での編成、検索、パッケージ化、提供が可能となっている。

- ・ **強み・弱み・課題：**

利用者は、電子メールを様々な単位で処理できる（個々のメッセージ単位、そのグループ単位、アカウント単位）。一括処理時の最大件数は 10 万件である。当ツールでは、MBOX ファイルを、EMA (EMail Account XML) 形式 (RFC 5322 の要件を満たす保存用スキーマ) へ変換する。EMA は、個別のメッセージからアカウントまですべてに適用可能である。EMA 保存形式のメッセージやアカウントのデータは、SHA-1 で照合される。ただし、DArcMail では、EML 形式へ正規化されない。また、EMA スキーマは普及していない。さらに、追加のメタデータ項目（例えば権利メタデータ）が必要であり、それは別システムに記録されねばならない。

ファイル形式

- ・ **投入時：**MBOX
- ・ **出力時：**
 - ・ XML email schema (AIP：マスターコピー)
 - ・ MBOX (DIP：提供用コピー)
 - ・ 添付は埋め込まれるか、あるいは別構造とされる

EAS (Electronic Archiving System)

基本情報

- ・ **開発元：**ハーバード大学図書館
- ・ **リンク先：**http://nrs.harvard.edu/urn-3:hul.eother:eas_overview
- ・ **入手条件：**当システムは、ハーバード大学のインフラ環境に強く依存しており、現状はハーバード大学内での利用にとどまる。EAS をオープンソース化する作業は、2017 年夏に開始された。
- ・ **システム要件：**管理用 UI である EASi はウェブベースのアプリケーションで、現状、ハーバード大学内の認証ユーザのみ (ID と PIN を入力後) 利用可能となっている。EAS では機微データを管理する可能性があるため、セキュアな VPN トンネルを設定する必要がある。対応ブラウザは、Firefox、Google Chrome、Safari である、
- ・ **現状：**維持管理されているが、現状、サポートはハーバード大学のコミュニティに限定されている。バージョン 1.0.5 は、2017 年 4 月に公開された。

ツールで何ができるか？

- ・ **目的：**EAS は、電子メールのコンテンツ（メッセージと添付）をアーキビストが選別評価・整理の上、ハーバード大学の保存システムへ投入するために利用するツールである。EAS では、電子メールコンテンツを一括または個別に処理可能である。メタデータの

追加・編集・削除、個々のメッセージや添付の削除、事後確認が必要なコンテンツのマーキング、機微データとしてセキュアな管理が必要なコンテンツのマーキング、PREMIS 権利メタデータの付与とコンテンツ単位での関連付けなどが行える。あるいは、EAS では、ファイル形式の変換、データ削除、添付の処理などの実行履歴や、技術メタデータを自動的に記録できる。メタデータはすべて（メッセージや添付といった）アイテムレベルで保管される。

- ・ **保存管理のライフサイクル上の位置付け | 整理・保存：**

EAS は、データ投入、選別評価、整理、保存システム投入時のファイル変換を行うための処理ツールである。

- ・ **強み・弱み・課題：**EAS は、複数段階の処理をまたいだ対応が可能であり、この柔軟性が一つの強みとなっている。例えば、最小限のメタデータしか付与されていない状態でもコレクション全体を長期保存用のストレージシステムへ直接データ投入ができる。あるいは、メタデータを個々のメッセージや添付に付加していくことも可能となっている。検索インタフェースも充実しており、検索条件を容易に変更でき、また、検索結果をある条件で絞り込んでいくことも行える。ただし、EAS では、受入前にアーキビストが選別評価することはできない。また、コンテンツとメタデータがいったんハーバード大学の保存システムへ投入された後は、認証ユーザしかそれらへアクセスできない。公開猶予期間や機微データの関係から、現時点では、一般ユーザへの開示・提供機能も用意されていない。受入前の選別評価、開示、提供機能が必要であれば、他のツール（例：ePADD）と組み合わせることが欠かせない。ハーバード大学でこれらを実現するためには、あるいは、EAS を利用する他のコミュニティにとっても、EAS はオープンソースソフトウェアとして配布されるのが望ましく、そのようなプロジェクトが2017年夏に開始された。

- ・ **ファイル形式**

- ・ **投入時：**OLM、MBOX、PST、EMLX

当ツールの現行の仕様では、必要に応じて1件追加したい場合でも、特定のクライアントソフトウェアからしか電子メールコンテンツ（メッセージと添付）を投入できない。対応しているものは次の通り：

- ・ Eudora for Windows/6.2
 - ・ Eudora for Windows/version unknown
 - ・ Mac OS X Mail/2.x
 - ・ Mailman/2.0.5
 - ・ Mailman/2.1.15
 - ・ Outlook for Mac (OLM only/version unknown)
 - ・ Outlook for Windows/version unknown
 - ・ Thunderbird/2.0.0.23

- Thunderbird/version unknown

整備・保管時には、すべて EML 形式へ変換される。EML 内の改行データは EAS 上で正規化される。添付はデコードされ、EAS 内で別ファイルとして保管され、関連する電子メールとはメタデータで紐付けされる。元々のクライアントソフトでメッセージと添付が別々に分けられて保管されているデータの場合は、EAS 上で、メタデータにもとづきメッセージと添付の関連付けが機械的に行われる。

- 出力時：EAS からの出力機能はない（ハーバード大学のシステム内部での出力は除く）。

ePADD (Email: Process, Appraise, Discover, Deliver)

基本情報

- 開発元：スタンフォード大学図書館（アショカ大学の Sudheendra Hangal 教授から技術支援を受けている）
- リンク先：
 - <http://epadd.stanford.edu/epadd/collections>
 - <https://github.com/ePADD/epadd/releases>
- 入手条件：Apache ライセンス 2.0 のもと配布されているオープンソースソフトウェアである。最新バージョンは GitHub 上で入手できる。
- システム要件：
 - OS: 64-bit, Windows 7 SP1 / 10, Mac OS X 10.11 / 10.12
 - Windows のインストール環境：Java Runtime Environment 64-bit, 8u101 以上
 - 確認環境：Windows 7 and OSX 10.9/10.10 のマシン（Java 7 or 8）
 - ブラウザベースのソフトウェアは次に対応：Chrome、Firefox.
 - メモリ：8GB RAM
 - ePADD は Java と Javascript で書かれており、Apache Tomcat [v7.0] 上で動く（Java EE Servlet API [v3.x]、Java Mail [v1.4.2]）。Apache Lucene [v4.7] と Apache Tika [v1.8] により、テキスト・メタデータ抽出、インデキシング、検索が可能となっている。グラフ描画や可視化には C3.js [D3-based reusable chart library v0.4.10] が用いられている。ファイルのパッケージ化には、オラクルの Java Application Bundler (Mac 用) と Launch4J (Windows 用) が利用されている。他にも Java ライブラリ（例：Lang、commons、CLI、IO、logging）が採用されている。JSON データを扱うために、org.json や Gson ライブラリも用いられている。
 - ePADD では、独自の自然言語処理ツールキットも実装されており、固有名表現抽出、名寄せなどの処理も実行できる。このツールキットは、ePADD の初期ベータ版で使用されていた Apache OpenNLP を置きかえたものである。スタンフォード大学では、ePADD 内のライブラリとして Muse も使用し続けていたが、Apache OpenNLP では（少なくとも固有名表現抽出の点において）不十分であることが確認された。改修が何度か試みられた後、この独自の固有名表現抽出ツールが開発さ

れることとなった。当ツールキットでは、Wikipedia や DBpedia、Freebase、Geonames、OCLC FAST、米国議会図書館件名標目表や LC の名称典拠ファイルなどの外部データが活用されている。

・ **現状：**

- ・ スタンフォード大学の特殊コレクション・大学アーカイブズ部門で開発・管理されている。協力機関は、ハーバード大学、メトロポリタン・ニューヨーク図書館カウンシル (METRO)、イリノイ大学アーバナ シャンペーン校、カリフォルニア大学アーバイン校である。
- ・ 現状、米国博物館・図書館サービス振興機構 (IMLS) の図書館分野における先導事業助成プログラム (National Leadership Grant for Libraries) の支援を受けて開発が進められている。当プログラムは、図書館やアーカイブズ領域で直面する課題に対応するプロジェクト、先進的な取り組みを行っているプロジェクトを支援するものである。なお、2015 年に初期版が開発された際には、主に全米歴史出版物記録委員会 (NHPRC) からの助成を受けた。
- ・ 利用者向け資料はスタンフォード大学図書館で管理されている。
<https://library.stanford.edu/projects/epadd/documentation>.
- ・ ユーザコミュニティの掲示板、メーリングリスト、フォーカスグループ会議、辞書作成作業部会に関する情報は次の URL から確認できる。
<https://library.stanford.edu/projects/epadd/community>.

ツールで何ができるか？

- ・ **目的：**ePADD は、スタンフォード大学の特殊コレクション・大学アーカイブズ部門で開発されたソフトウェアパッケージである。当ツールにより、提供者やキュレーターによる選別評価、整理、開示（オンライン上でのメタデータの公開）、閲覧室内でのメッセージ全文や添付の提供が可能となる。
- ・ **保存管理のライフサイクル上の位置付け：**ePADD は、選別評価、整理、開示、提供という 4 つのモジュールから構成される。これらは、電子メールを保存・管理するライフサイクルの各段階に対応している。
 - ・ 選別評価：当モジュールにより、保存機関へ電子メールのデータが引き渡される前に、作成者、資料提供の仲介者、キュレーターがそのデータを容易に確認できる。
 - ・ 整理：当モジュールにより、電子メールのデータを整理し、それらに関する記述作業を行うことができる。
 - ・ 開示・提供：開示モジュールにより、編集加工した電子メールアーカイブズ資料をウェブ上で利用できるようにすることができる。提供モジュールにより、電子メールの全文を閲覧室内で確認することが可能となる。
- ・ **強み・弱み・課題：**ePADD は開発中であり、2018 年末に追加機能が公開される予定で

ある。

- ・ **強み：**
 - ・ 非構造化データを自動的に構造化データに変換可能
 - ・ 抽出されたエンティティに永続的識別子を自動的に付与可能
 - ・ 典拠レコードを提供可能（データ間の関係性を確認できる）
 - ・ メッセージに注釈を付与可能
 - ・ 添付画像を一度に確認可能
 - ・ 同一人物の異なる複数アドレスをグループ化可能
 - ・ 複数エンティティをグループ化可能
 - ・ ユーザ向けに検索式を生成可能
 - ・ 辞書を利用した複雑な検索条件を生成可能（テンプレートが用意）
 - ・ メーリングリストからメッセージを分離可能
- ・ **課題：**
 - ・ 開示機能において、コレクションをまたぐ検索ができない。
 - ・ 提供機能において、対応している添付ファイルの形式は、Quick View Plus で実行できるものに限られている。また、受信者名などのメタデータに注釈を付けたり、修正したりすることもできない。
 - ・ GraphML 形式での出力にまだ対応していない。
 - ・ 現時点で、添付データは変換されず、商用ソフトウェア（Quick View Plus など）で対応している約 300 種類の形式のものしか閲覧することができない。

ファイル形式

- ・ **投入時：**選別評価モジュールのデフォルト設定では、投入画面がまず開く。対象電子メールアカウントの所有者に関する情報（氏名と電子メールのアドレス）の入力が求められる。さらに、利用したい IMAP アカウントの情報を入力したり、MBOX ファイルを指定したりすることで、データが投入されていく。なお、アカウント、MBOX ファイルともに複数選択可能である。
- ・ **出力時：**現状、保存用に MBOX 形式のデータが出力できる。出力画面上でファイル形式などの条件を指定して、出力対象データを選択可能である。Apache Tika で認識されていない（つまり、ePADD 上ではインデックスが作成されていない）データを出力対象とすることもできる。

Preservica Standard Edition

基本情報

- ・ **開発元：**Preservica 社（Tessella の子会社）
- ・ **リンク先：**
<http://preservica.com/>
 ウェビナー（要登録）：

<https://preservica.com/events/webinars-live-demos/15/07/2015/email-archiving-and-preservation>

ツールで何ができるか？

- ・ **目的：**ウェブサイト上では以前、「Preservica により、データの投入、管理、保管、提供、保存といった OAIS 準拠のワークフローを包括的に実現できる。また、Universal Access モジュールにより、コンテンツを安全に公開できる」と紹介されていた。データの投入は完全に自動化されている（10 TB 以上のコレクションも一括登録可能）。SIP の作成ツールも用意されている。マイクロサービス設計となっており、ウイルススキャン、チェックサムの確認、ファイルの分類、技術メタデータの抽出、ファイル形式の正規化など、ワークフローの各段階でコンテンツを処理することが可能となっている。アーキビストはメタデータをデジタルコンテンツに付加でき、管理画面や Universal Access モジュールからも検索できるようになる。
- ・ **保存管理のライフサイクル上の位置付け**
 - ・ **整理：**受入前の選別評価には対応していないが、Preservica では、EML 形式への正規化を行い、後続の整理作業を効率化させている。手動による整理、メッセージ本文のインデキシングもできるよう、アイテム単位での処理、一括処理、「会話のような関係性（conversational relationships）」の維持管理が可能となっている。
 - ・ **保存：**Preservica は、コンテンツのパッケージ化や、パッケージ化したファイルの格納にも対応している。
 - ・ **開示・提供：**Universal Access モジュールを介して、オンライン上での開示、利用提供が可能である。
- ・ **強み・弱み・課題：**電子メールに特化したツールでないが、Preservica は、そのようなコンテンツの長期保存・提供に資する機能も有す。ワークフローが自動化されており、大半の保存機能（ファイル形式の正規化を含む）が含まれている点が主たる特徴となっている。また、メタデータ（メッセージとスレッドやフォルダとの関係性についての情報を含む）を抽出・追加できる点も、保存という面で重要な機能といえる。統合的なプラットフォームである点もまた重要である。特に、電子メールの保存と提供の両面をサポートしている点が挙げられる。とはいえ、Preservica では、機微情報・個人情報の特定や高度な検索（自然言語処理や固有表現抽出）まではできないようである。

ファイル形式：

- ・ **投入時：**
 - ・ MS Outlook（PST、MSG）
 - ・ Lotus Notes
 - ・ MBox
 - ・ Gmail（Google Takeout で MBOX に出力後）
- ・ **出力時：**

- EML
- 添付は保存用形式に正規化されるかもしれない。
- メタデータは、Axiell Calm、Adlib、ArchivesSpace などのシステムへの出力、EAD、MODS、Dublin Core での出力に対応

TOMES Tool (Transforming Online Mail with Embedded Semantics)

基本情報

- **開発元:** ノースカロライナ州立文書館 (ユタ州立文書館、カンザス州立歴史協会の支援)
- **リンク先:**
<https://www.ncdcr.gov/resources/records-management/tomes>
<https://github.com/StateArchivesOfNorthCarolina>
- **入手条件:** オープンソースで、最新版は GitHub 上で入手可能
- **システム要件:**
 - 64-bit OS (Docker が動く環境 <https://www.docker.com/community-edition>)
 - 4 GB RAM (最低要件)
 - 最新ブラウザ (Chrome、Firefox、IE、Safari)
- **現状:** 開発中。ノースカロライナ州立文書館で管理されている。開発資金は、全米歴史出版物記録委員会からの助成による。

ツールで何ができるか?

- **目的:** TOMES ツールは、個人情報、機密情報、固有表現を識別するため、自然言語処理にもとづくタグ付与機能が用いられ、アカウント単位で電子メールをより早く処理することが可能となる。州政府向けに整備された XML 形式の辞書が使用されている。まだ開発中の段階であり、2018 年 9 月に公開予定である。
- **保存管理のライフサイクル上の位置付け | 受入・整理**
- **強み・弱み・課題:**
 - **強み:**
 - 政府資料など、大規模な電子メールコレクションでも処理が可能である。
 - 反復処理が可能で、処理困難な対象アカウントでもより早く利用可能な状態にすることができるかもしれない。
 - Docker を利用しており、多くのコンピュータ環境で実装可能となっている。
 - **課題:**
 - 自然言語処理によるアシスト機能を効果的に利用するには、専用の知識データが欠かせない。
 - EAXS は保存に特化したフレームワークであり、見直しされるのが望ましい。

ファイル形式

- ・ 投入時 : PST、MBOX、EML
- ・ 出力時 : タグ付きの EAXS XML

■ 付録 C : 電子メール保存の調査プロジェクト

電子メール保存に関するシンポジウム (Archiving Email Symposium)

調査主体：米国議会図書館、米国国立公文書記録管理局

概要：「2015 年 6 月 2 日、米国議会図書館と米国国立公文書記録管理局は、電子メールのメッセージや添付データの受入・保存の現状に関する情報共有を目的として、電子メール保存に関するシンポジウムを共催した（場所は米国議会図書館）。参加者は約 150 名で、技術者、ソフトウェア開発者、図書館員、キュレーター、記録管理者、小規模機関のアーキビスト、研究者など、多彩な実務家が集まった。他にも、全米人文科学基金（National Endowment for Humanities）、米国博物館・図書館サービス振興機構（Institute of Museum and Library Services）、全米歴史出版物記録委員会（National Historical Publications and Records Commission）など、数千規模の職員がいる大規模な政府系機関や資金提供機関からの参加も見られた。当イベントでは、インフォーマルなワークショップも 6 月 3 日に開かれ、ツール、作業手順、方針の改善点をより明確にするため、シンポジウムで確認された諸課題について議論が交わされた」（Murray and Engle 2015）。

実施年：2015 年

報告書および参照情報源：

- <http://www.digitalpreservation.gov/meetings/archivingemailsymposium.html>
- We Welcome Our Email Overlords: Highlights from the Archiving Email Symposium

カーカネットプレス (Carcanet Press) の電子メール保存プロジェクト

調査主体：マンチェスター大学図書館

概要：「マンチェスター大学図書館（UML）が所蔵する現代アーカイブズ資料の中で、もっとも重要なものは、詩を専門とした英国の出版社、カーカネットプレスのそれである。著名な詩人、批評家、編集者、翻訳者、アーティストとのやり取りの記録が、このアーカイブズ資料の重要な要素となっている。これまでは年に一度、やり取りの記録が大量の紙資料として UML へ送られてきていたが、現在ではやり取りの大半が電子メールで行われるようになっており、結果、それらが無くなってしまった。そこで、UML のような図書館が、デジタル形式の電子メールを保存できるようになることが重要となってきた。この価値ある研究資料が失われないことを担保しなければならない」（Baker, Butler, and Green 2012, 3）。カーカネットプレスの電子メール保存プロジェクトでは、「従来型の技法とデジタル保存の標準双方」が採用された。当プロジェクトにより、多数の成果が生まれたが、主たるものを次に挙げておく。

- メタデータ抽出やマイグレーション結果の自動検証用のソフトウェアコード
- 保管用情報パッケージのデータモデルとメタデータプロファイル
- 新規ワークフローに関する詳細文書

- ・ デジタル保存専用のハードウェア、デジタル資料の処理を実行するためのセキュアなネットワークストレージ

実施年：2012 年

報告書および参照情報源：

- ・ <https://www.escholar.manchester.ac.uk/api/datastream?publicationPid=uk-ac-man-scw:165096&datastreamId=FULL-TEXT.PDF>
- ・ <https://www.escholar.manchester.ac.uk/api/datastream?publicationPid=uk-ac-man-scw:226625&datastreamId=FULL-TEXT.PDF>

CERP (Collaborative Electronic Records Project)

調査主体：スミソニアン協会アーカイブズ、ロックフェラーのアーカイブセンター

概要：当プロジェクト開始時に目指されたのは、「対象コレクションにあるボーンデジタル資料を管理・保存する上での手法や技術を開発することである。主な目的は、運営ガイドラインを策定すること、技術的知見をえることであり、それにより、未来の研究者が電子情報を利用できるようになり、また、資料提供者、協力機関、他の関連非営利団体と成果を共有することも可能となる」。開始後まもなく、プロジェクトの範囲は、電子メールに絞られることとなった。当プロジェクトチームは、電子メールコレクション・保存 (Electronic Mail Collection and Preservation) プロジェクトと共同して、アカウント単位の XML スキーマを策定した。20008 年の最終報告時点で、「CERP は、最良事例のガイドライン、ワークフローの概要、ソフトウェアの評価、SIP/AIP/DIP のモデル、メッセージと合わせアカウント単位で保存するソフトウェアツール、DSpace 用のデータ投入モジュールを作成した。また、89,000 件以上のメッセージを 99%の成功率で処理した」とまとめられている (Adgent and Fuhrig 2009, 3-4)。

実施年：2005～2008 年

報告書および参照情報源：

- ・ <http://siarchives.si.edu/cerp/>
- ・ http://siarchives.si.edu/cerp/CERP_project_summary_122008_CC.pdf

DAVID (Digital Archiving in Flemish Institutions and Administrations)

調査主体：フィリップ・ブドレ (アントウェルペン市立文書館)

概要：「この取り組みは、マックス・ワイルディアース財団の科学研究基金によるプロジェクトの一つであり、アントウェルペン市立文書館とルーヴェン・カトリック大学 法と情報学学際センターとの共同で進められた。当プロジェクトの目標は、電子的アーカイビングに関するマニュアルを策定することであった……DAVID プロジェクトでは、電子メール保存に関する法的な、保存上の要件を検証し、とりうる戦略について提言した。これにもとづき、モデルとなるソリューションが策定された。このような理論的な作業に加え、報告書では、電子メールおよび関連する電子的記録を管理・保

管する手順の実装に向けた整理も行なった」(Boudrez 2006, 2)。

実施年：1999～2003 年

報告書および参照情報源：<http://www.imaginar.org/taller/dppd/DPPD/179%20pp%20DAVID.pdf>

Kaine Email Project@LVA

調査主体：バージニア図書館 (Library of Virginia)

概要：2010 年 1 月、州知事を退任するティム・ケイン政権は、バージニア図書館に対し、200 件以上の電子メールアカウントに含まれた 130 万件にのぼるメッセージを移管した。法律上、図書館へ移管された州知事の記録は、「目録が整い次第、公衆の利用に供されるもの」とされている (Va. Code § 2.2-126)。州知事の電子メールを移管されたバージニア図書館側としては、新しくツールを開発し、受入、整理、利用提供に関する手順を整備しなければならない。

実施年：2010～(継続中)

報告書および参照情報源：<http://www.viriniamemory.com/collections/kaine/>

MeMail (Email Preservation at the University of Michigan)

調査主体：ミシガン大学ベントレー ヒストリカル図書館

概要：アンドリュー・メロン財団の資金をえて 2 年間実施されたプロジェクト MeMail では、ミシガン大学において、多様な電子メールアプリケーションが使用されることから、また、電子メールの管理方法が個人により様々であることから生じる保存の問題を解消していくことが目指された。長期的な価値をもつ電子メールの選別評価を作成者とともに行う作業、アーカイブズ機関へ電子メールを移管する上で適切なツールを特定する作業が実施された。実験の参加者らは、価値あるメッセージを「保存箱」へドラッグ&ドロップ、転送またはコピーすることが求められ、次にアーキビストが対象となる電子メールを出力し、保存のワークフロー上へ流していく。当調査によれば、記録作成者が価値ある電子メールを評価選別することは必ずしもできず、結果、大学の電子メールサービスとして Gmail を採用した際、作成者が保存箱へ転送するというやり方ではうまくいかないことが示された。当プロジェクトの成果は、ベントレー図書館が、より頑健なデジタル保存の手順とワークフローを構築するのにも役立てられたという。

実施年：2010～2011 年

報告書および参照情報源：

- SAA Campus Case Study #14: Partnering with IT to Identify a Commercial Tool for Capturing Archival Email of University Executives at the University of Michigan
<http://files.archivists.org/pubs/CampusCaseStudies/CASE-14-FINAL.pdf>
- SAA Campus Case Study #15: Will They Populate the Boxes? Piloting a Low-Tech Method for Capturing Executive E-mail and a Workflow for Preserving It at the University of Michigan

[http:// files.archivists.org/pubs/CampusCaseStudies/CASE-15-FINAL.pdf](http://files.archivists.org/pubs/CampusCaseStudies/CASE-15-FINAL.pdf)

PeDALS (Persistent Digital Archives and Library System)

調査主体：

- ・ アリゾナ州立図書館・文書館・公的記録局（主担当）
- ・ アラバマ州立文書館・歴史資料館（Alabama Department of Archives and History）
- ・ フロリダ州立図書館・文書館
- ・ ニューメキシコ州立記録局・文書館
- ・ ニューヨーク州立文書館、ニューヨーク州立図書館
- ・ サウスカロライナ州立文書館・歴史資料館（South Carolina Department of Archives and History）、サウスカロライナ州立図書館
- ・ ウィスコンシン歴史協会

概要：当プロジェクトは「2008年1月から2012年3月にかけて実施された調査プロジェクトで、2つの実用的な目標が設定された。一つ目は、電子刊行物や記録資料を処理するワークフローの自動化・統合化に関する考え方を整理することである。二つ目は、対象コレクションの真正性と完全性を担保できるよう、安価なストレージネットワークを用い、『デジタル保管所（digital stacks）』を実装することである。これらの実用的な目標に加え、PeDALSでは、同じような課題に取り組む機関が情報共有できるコミュニティの構築が目指された。それにより、多様な保存機関のニーズを満たすシステムが実現され、継続的に改修することも、最良事例の蓄積を促していくことも可能となるかもしれない。さらにその先を見据え、できるだけ低コストで諸課題を解消していくことを試みた」（PeDALS 2013）。

実施年：2008～2012年

報告書および参照情報源：

- ・ <http://web.archive.org/web/20130306060835/http://www.pedal-spreservation.org/>
- ・ <http://azmemory.azlibrary.gov/cdm/ref/collection/statepubs/id/15540>

TOMES (Transforming Online Mail with Embedded Semantics)

調査主体：ノースカロライナ州立文書館、ユタ州立文書館・記録局、カンザス歴史協会

概要：当プロジェクトでは、「全米歴史出版物記録委員会からの助成を受け、重要な政府の記録を収集・保存し、利用に供することができるよう、永続的な価値をもつ公務員の電子メールを対象とした調査が実施された。TOMESの調査に協力いただいた州は、カンザス、ユタ、ノースカロライナである。オンライン上の電子メールサービス（マイクロソフトのOffice 365やGmail）から電子メールのデータを移行し、持続可能性の高いオープンソース言語へ変換する手順を開発することに主眼が置かれた。加えて、アーキビストがデータを効率的に処理し、利用に供することができるよう、ePADDの成果をもとに、自然言語処理や州政府向けの辞書を組み込んだ選別評価ツールの

開発も予定されている」(NCDNCR 2018)。

実施年：2015～2018 年

報告書および参照情報源：<https://www.ncdcr.gov/resources/records-management/tomes>

■ 付録 D : 参照情報源

URL はすべて 2018 年 5 月 31 日現在 (訳注 : 翻訳作業においてはすべて未確認)

- Adgent, Nancy, and Lynda Schmitz Fuhrig. 2009. The Collaborative Electronic Records Project Summary. Sleepy Hollow, NY, and Washington, DC: The Collaborative Electronic Records Project. http://siar-chives.si.edu/cerp/CERP_project_summary_122008_CC.pdf.
- AIIM. 2018. Glossary. Accessed May 31. <http://www.aiim.org/Resources/Glossary/Glossary-List-Page>.
- Alderman, Liz. 2017. “Bell Pottinger, British P.R. Firm for Question-able Clients, Collapses.” New York Times, September 12 (Business Day). <https://www.nytimes.com/2017/09/12/business/bell-pottinger-administration.html>.
- Animal Adventure Park. 2018. Official Animal Adventure Park April the Giraffe Page. Accessed May 31. <http://www.aprilthegiraffe.com/>.
- Ankerson, Megan. 2012. “Writing Web Histories with an Eye on the Analog Past.” New Media and Society 14 (3): 384–400. <https://doi.org/10.1177/1461444811414834>.
- Apple. 2018. “iCloud: Add an Email Attachment in iCloud Mail.” Posted March 30, 2018. https://support.apple.com/kb/ph2629?locale=en_US.
- Archives New Zealand. 2018. “Visual Rendering Matters.” Accessed May 31. <http://archives.govt.nz/resources/information-management-research/rendering-matters-report-results-research-digital-object-0>.
- Attfield, Simon, and Larry Chapin. 2018. “The Reconstruction of Narrative in E-Discovery Investigations.” Presentation at Email Preservation: How Hard Can it Be? January 24, 2018, Woburn House, London. Digital Preservation Coalition. <https://www.dpconline.org/docs/miscellaneous/events/2018-events/1766-dpc-email-ii-attfield-chapin/file>.
- AV Preserve. 2018. “Fixity.” Accessed May 31. <https://www.avpre-serve.com/products/fixity/>.
- Baker, Fran. 2014. Carcanet Press Email Preservation Project Phases 2-3: Final Report. Carcanet Press Email Preservation Project. United Kingdom: Carcanet Press. <https://www.escholar.manchester.ac.uk/api/datastream?publicationPid=uk-ac-man-scw:226625&datastreamId=FULL-TEXT.PDF>.
- _____. 2015. “E-Mails to an Editor: Safeguarding the Literary Correspondence of the Twenty-First Century at The University of Manchester Library.” Special Collections in a Digital Age 21 (2): 216–224. <https://doi.org/10.1080/13614533.2015.1040925>.
- Baker, Fran, Phil Butler, and Ben Green. 2012. Carcanet Press Email Preservation Project: JISC Final Report. <https://www.escholar.manchester.ac.uk/api/datastream?publicationPid=uk-ac-man-scw:165096&datastreamId=FULL-TEXT.PDF>.
- Banday, M. T. 2011. “Technology Corner: Analysing E-Mail Headers for Forensic Investigation.” Journal of Digital Forensics, Security and Law 6 (2): 49–64. <https://doi.org/10.15394/jdfsl.2011.1095>.
- Bearman, David. 2017. “Review of ‘Office of the Secretary: Evaluation of Email Records Management

- and Cybersecurity Requirements, ESP-16-03.” *The American Archivist* 80 (2): 459–462.
<https://doi.org/10.17723/0360-9081-80.2.459>.
- Ben-David, Anat, and Hugo Huurdeman. 2014. “Web Archive Search as Research: Methodological and Theoretical Implications.” *Alexandria: The Journal of National and International Library and Information Issues* 25 (1/2): 93–111. <https://doi.org/10.7227/ALX.0022>.
- BITS Security Program. 2013. Email Authentication Policy and Deployment Strategy for Financial Services Firms. Washington, DC: BITS Security Working Group. <http://www.fsroundtable.org/wp-content/uploads/2015/05/BITSEmailAuthenticationFeb2013.pdf>.
- Boudrez, Filip. 2006. Filing and Archiving Email.
http://www.expertisecentrumdavid.be/docs/filingArchiving_email.pdf.
- Bradner, S. 1996. “The Internet Standards Process—Revision 3. RFC— Best Current Practice.” Harvard University. <https://datatracker.ietf.org/doc/rfc2026/>.
- Brunton, Finn. 2013. *Spam: A Shadow History of the Internet*. Cambridge, MA, and London: The MIT Press. <https://mitpress.mit.edu/books/spam>.
- BSG Web Group. 2017. “Wrench.” September 5, 2017. <https://nasa3d.arc.nasa.gov/detail/wrench-mis>.
- Bunn, Jenny, Sara Brimble, Selene Obolensky, and Nicola Wood. 2015. Team Europe EU28 Project 2015–16: Perceptions of Born Digital Authenticity. InterPARES Report. InterPARES Trust.
https://interpar-estrust.org/assets/public/dissemination/EU28_20160718_UserPerceptionsOfAuthenticity_FinalReport.pdf.
- Caplan, Priscilla. 2009. Understanding PREMIS. Library of Congress Network Development and MARC Standards Office. Washington, DC: Library of Congress.
<https://www.loc.gov/standards/premis/understanding-premis.pdf>.
- Casserly, Martyn. 2017. “The Best Free Email Services for 2017.” Tech Advisor. Internet Feature (blog), September 11, 2017. <https://www.techadvisor.co.uk/feature/internet/best-free-email-services-for-2017-3613837/>.
- Cerbain, Jose. 2016. “Reports of the Death of Email Are Greatly Exaggerated.” Huffington Post (blog), July 2, 2016. https://www.huffingtonpost.com/advertising-week/reports-of-the-death-of-email_11114786.html.
- Cheney, Kyle. 2017. “GOP Chairmen Seek to Interview Top FBI Officials on Clinton, Trump.” Politico, December 19. <https://www.politico.com/news/hillary-clinton-emails>.
- Cocciolo, Anthony. 2016. “Email as Cultural Heritage Resource: Appraisal Solutions from an Art Museum Context.” *Records Management Journal* 26 (1): 68–82. <http://dx.doi.org/10.1108/RMJ-04-2015-0014>.
- Cohen, Jordan. 2015. “Looking at the EMAIL MARKETINGscape” (blog), January 26, 2015.
<https://www.emailvendorselection.com/author/jordan-cohen/>.
- Cohen, William W. 2015. Enron Email Dataset. May 8, 2015. <https://www.cs.cmu.edu/~enron>.
- Context.io. 2018. Accessed May 31. <https://docs.context.io/>.
- Cormack, Gordon V., and Maura R. Grossman. 2017. “Navigating Imprecision in Relevance Assessments on the Road to Total Recall: Roger and Me.” In *SIGIR ’17 Proceedings of the 40th International ACM*

- SIGIR Conference on Research and Development in Information Retrieval, 5–14. ACM Press.
<https://doi.org/10.1145/3077136.3080812>.
- Crispin, Mark R. 2003. “Internet Message Access Protocol–Version 4 Rev1. RFC 3501.” Network Working Group, Internet Engineering Task Force. <https://tools.ietf.org/html/rfc3501>.
- Crocker, D. 2008. “Internet Mail Architecture.” Network Working Group, Internet Engineering Task Force. <http://www.bbiw.net/specifications/draft-crocker-email-arch-11.html>.
- Daintith, John, and Edmund Wright. 2008. *A Dictionary of Computing*. 6th ed. New York: Oxford University Press, Inc.
- Dappert, Angela, Sébastien Peyrard, Carol C. H. Chou, and Janet Delve. 2013. “Describing and Preserving Digital Object Environments.” *New Review of Information Networking* 18 (2): 106–173.
<https://doi.org/10.1080/13614576.2013.842494>.
- Dayley, Alan, Julian Tirsu, Garth Landers, and Shane Harris. 2016. *Magic Quadrant for Enterprise Information Archiving*. ID: G00294240. Stamford, CT: Gartner, Inc.
<https://www.gartner.com/doc/3535317>.
- Digital Curation Centre. 2018. “DCC Curation Lifecycle Model.” Accessed May 31.
<http://www.dcc.ac.uk/resources/curation-lifecycle-model>.
- Digital Preservation Coalition. 2015. “File Formats and Standards.” *Digital Preservation Handbook*, 2nd ed. <https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards>.
- dmarc.org. 2018. *Domain Message Authentication Reporting & Conformance (DMARC)*. Accessed May 31. <https://dmarc.org/>.
- Ducheneaut, Nicolas, and Victoria Bellotti. 2001. “E-Mail as Habitat: An Exploration of Embedded Personal Information Management.” *Interactions* 8 (5) 30–38.
<https://dl.acm.org/citation.cfm?id=383305>.
- Duke Law Center for Judicial Studies. 2018. “New EDRM Enron Email Data Set.” Accessed May 31.
<https://www.edrm.net/resources/data-sets/edrm-enron-email-data-set/>.
- Dyer, Jessica. 2017a. “UNM’s Krebs Encouraged Staffers to Purge Emails.” *Albuquerque Journal*, October 29. <https://www.abqjournal.com/1085003/unms-krebs-encouraged-staffers-to-purge-emails.html>.
- _____. 2017b. “Krebs Said He Deleted UNM Emails.” *Albuquerque Journal*, November 23.
<https://www.abqjournal.com/1097004/krebs-said-he-deleted-unm-emails.html>.
- Email Design Reference. 2018. “Responsive Email.” MailChimp. Accessed May 31.
<https://templates.mailchimp.com/development/responsive-email/>.
- Exterro. 2018. “Predictive Coding (Technology Assisted Review)”. *The Basics of E-Discovery*, Chap. 7B. Accessed May 31. <https://www.exterro.com/basics-of-e-discovery/predictive-coding/>.
- Ferguson, Rob. 2017. “Civil Servants Kept Emails in Case They Shed Light on McGuinty’s Gas Plant Cancellations.” *The Toronto Star*, October 20 (Queen’s Park).
<https://www.thestar.com/news/queenspark/2017/10/20/emails-kept-in-case-they-shed-light-on-mcguintys-gas-plant-cancellations.html>.
- Ferrante, Ricc. 2015. “Archiving Email: Institutional Approaches to Processing and Archiving Email

- [902].” Presentation at Archiving Email Symposium, June 2, 2015, Washington, DC, co-hosted by the Library of Congress and the National Archives and Records Administration.
<https://www.youtube.com/watch?v=4xTVnkqsOF0&list=PLEA69BE43AA9F7E68&t=0s&index=5>.
 Transcript available at <https://stream-media.loc.gov/webcasts/captions/2015/150602osi0902.txt>.
- Firstbrook, Peter, and Neil Wynne. 2015. Magic Quadrant for Secure Email Gateways. ID: G00268427. Stamford, CT: Gartner, Inc. <https://www.gartner.com/doc/3084025/magic-quadrant-secure-email-gateways>.
- Fitzgerald, Neal. 2013. “Using Data Archiving Tools to Preserve Archival Records in Business Systems—A Case Study.”
http://purl.pt/24107/1/iPres2013_PDF/Using%20data%20archiving%20tools%20to%20preserve%20archival%20records%20in%20business%20sys-tems%20%E2%80%93%20a%20case%20study.pdf.
- Freed, N., and N. Borenstein. 1996. “Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies. RFC 2045.” Network Working Group, Internet Engineering Task Force. <https://tools.ietf.org/rfc/rfc2045>.
- FWD:Everyone. 2018. Accessed May 31. <https://www.fwdeveryone.com/>.
- Gearan, Anne, and Philip Rucker. 2017. “Trump Criticizes How Mueller Obtained Transition Emails, Says No Plans to Fire Special Counsel.” Washington Post, December 17 (Politics).
https://www.washingtonpost.com/politics/mueller-unlawfully-obtained-emails-trump-transition-team-says/2017/12/16/6162f350-e2cc-11e7-8679-a9728984779c_story.html.
- Gibson, Jeremy. 2018. “State Archives of North Carolina.” GitHub.
 Accessed May 31. <https://github.com/StateArchivesOfNorthCarolina>.
- Gmail Help. 2018. “Send Google Drive Attachments in Gmail-Computer.” Gmail Help. Accessed May 31. <https://support.google.com/mail/answer/2487407?co=GENIE.Platform%3DDesktop&hl=en>.
- Google. 2017. “Gmail API Overview.” Last updated May 11, 2017.
<https://developers.google.com/gmail/api/guides/>.
- Grace, Stephan, Gareth Knight, and Lynn Montague. 2009. Investigating the Significant Properties of Electronic Content Over Time. Final Report.
https://web.archive.org/web/20151024064901if_/http://www.significantproperties.org.uk/inspect-finalreport.pdf.
- Graham, Robert. 2016. “Yes, We Can Validate the Wikileaks Emails.” Errata Security (blog), October 21, 2016. <http://blog.erratasec.com/2016/10/yes-we-can-validate-wikileaks-emails.html>.
- Grammer, Geoff. 2017. “Krebs’ Emails Show His Desire to Protect Donors, Acknowledges Mistakes Made.” Albuquerque Journal, September 18. <https://www.abqjournal.com/1065745/krebs-emails-show-his-desire-to-protect-donors-acknowledges-mistakes-made.html>.
- Grossman, Maura R., and Gordon V. Cormack. 2011. “Technology- Assisted Review in E-Discovery Can Be More Effective and More Efficient than Exhaustive Manual Review.” Richmond Journal of Law and Technology 17 (3): 1–48. <http://jolt.richmond.edu/jolt-archive/v17i3/article11.pdf>.
- _____. 2014. “Comments on ‘The Implications of Rule 26(g) on the Use of Technology-Assisted Review.’” The Federal Courts Law Review 7 (1): 285–313.

- <http://www.fclr.org/fclr/articles/pdf/comments-implications-rule26g-tar-62314.pdf>.
- Harvard Library. 2016. "Email Archiving Stewardship Workshop." Published March 16, 2016.
<http://library.harvard.edu/03092016-1642/email-archiving-stewardship-workshop>.
- Harvard University Archives. 2018. "University Records Policies." Accessed May 31.
<http://library.harvard.edu/university-archives/managing-university-records/policies>.
- Harvard Wiki. 2018. DRS Content Guide to In-Production Content Models. Last modified January 26, 2018.
https://wiki.harvard.edu/confluence/pages/viewpage.action?pageId=204385879&preview=/204385879/218248076/public_drs_content_guide.pdf.
- History Lab. 2018. "History as Data Science (Freedom of Information Archive)." Accessed May 31.
<http://www.history-lab.org>.
- hMailServer. 2018. Storing the Message in the Database. Accessed January 10.
https://www.hmailserver.com/documentation/atest/?page=faq_storing_message_in_database.
- Hopkins, Julie, and Adam Sarnier. 2015. Market Guide for Email Marketing. ID: G00276045. Stamford, CT: Gartner, Inc.
<https://www.gartner.com/doc/3173144?ref=SiteSearch&stkw=email%20marketing&fml=search&srcId=1-3478922254>.
- Huth, Geof. 2016. "Appraising Digital Records." In *Appraisal and Acquisition Strategies*, edited by Michael Shallcross and Christopher Prom, 7–68. Chicago: Society of American Archivists.
- Illinois State Archives, and Records and Information Management Services, University of Illinois at Urbana-Champaign. 2017. "Processing Capstone Email Using Predictive Coding." University of Illinois, Urbana Champaign.
https://www.uillinois.edu/cio/services/rims/about_rims/projects/processing_capstone_email_using_predictive_coding/.
- Informatica. 2018. "Email Verification." Accessed May 31. <https://www.informatica.com/products/data-quality/data-as-a-service/email-verification.html#fbid=q2fSFYMPuEz>.
- Jääskeläinen, Anssi, Miia Kosonen, and Lissa Uosukainen. 2017. "Developing a Citizen Archive." BloggERS! (blog), February 1, 2017. <https://saaers.wordpress.com/2017/02/01/developing-a-citizen-archive/>.
- John, Jeremy Leighton, Ian Rowlands, Peter Williams, and Katrina Dean. 2010. *Digital Lives: Personal Digital Archives for the 21st Century: An Initial Synthesis*. A Digital Lives Research Paper, beta version 0.2 (March 3). British Library. <http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf>.
- Johnston, David. 1990. "5,000 Files Erased From Poindexter's Computer." *New York Times*, March 16.
<http://www.nytimes.com/1990/03/16/us/5000-files-erased-from-poindexter-s-computer.html>.
- Kekre, Anand. 2015. "Securing Email Attachments with Digital Rights Management." Vaultize's Secure Enterprise File Sharing Blog, April 16, 2015. <http://www.vaultize.com/blog/securing-email-attachments-with-digital-rights-management>.
- Klensin, J. 2008. "Simple Mail Transfer Protocol. RFC 5321." Network Working Group, Internet

- Engineering Task Force. <http://tools.ietf.org/html/rfc5321>.
- Knight, Gareth. 2010. Significant Properties Testing Report: Electronic Mail. Jisc, The National Archives, and Kings College London.
https://web.archive.org/web/20151024134638if_/http://www.significantproperties.org.uk/email-testingreport.pdf.
- Kucherawy, M. 2009. "Message Header Field for Indicating Message Authentication Status. RFC 5451." Network Working Group, Internet Engineering Task Force. <https://tools.ietf.org/html/rfc5451>.
- Landers, Garth, Shane Harris, and Jie Zhang. 2017. When to Use Microsoft's Native Capabilities for Archiving and E-Discovery. ID: G00315437. Stamford, CT: Gartner, Inc.
<https://www.gartner.com/doc/3658817>.
- Larramo, Mika. 2018. "SMTP Commands Reference." Accessed May 31.
<http://www.samlogic.net/articles/smtp-commands-reference.htm>.
- Lavoie, Brian, and Richard Gartner. 2013. Preservation Metadata, 2nd ed. DPC Technology Watch Report 13. Great Britain: Digital Preservation Coalition. <http://www.dpconline.org/docs/technology-watch-reports/894-dpctw13-03/file>.
- Leiba, Barry. 2013. "Change the Status of ADSP (RFC 5617) to Historic." Internet Engineering Task Force. <https://datatracker.ietf.org/doc/status-change-adsp-rfc5617-to-historic/>.
- Levinson, Edward, ed. 1998. "Content-ID and Message-ID Uniform Resource Locators. RFC 2392." Network Working Group, Internet Engineering Task Force. <https://tools.ietf.org/html/rfc2392>.
- Library of Congress. 2016. "MBOX Email Format." Last significant FDD update November 17, 2016.
<https://www.loc.gov/preservation/digital/formats/fdd/fdd000383.shtml>.
- Library of Congress. 2018. "Digital Content Transfer Tools." Digital Preservation. Accessed May 30.
<http://www.digitalpreservation.gov/series/challenge/data-transfer-tools.html>.
- Library of Virginia. 2016. "Kaine Email Project @ LVA." Virginia Memory.
<http://www.virginiamemory.com/collections/kaine/>.
- Light, Michelle, and Tom Hyry. 2002. "Colophons and Annotations: New Directions for the Finding Aid." *The American Archivist* 65 (2): 216–230. <https://doi.org/10.17723/aarc.65.2.13h27j5x8716586q>.
- Lin, Tom C. W. 2016. "Compliance, Technology, and Modern Finance." Temple University Legal Studies Research Paper No. 2017-06. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2904664.
- Loftus, Mary J. 2010. "The Author's Desktop." *Emory Magazine* (Winter).
http://www.emory.edu/EMORY_MAGAZINE/2010/winter/authors.html.
- MacAskill, Ewen, and Owen Bowcott. 2017. "UK Prosecutors Admit Destroying Key Emails in Julian Assange Case." *The Guardian*, November 10 (Media).
<http://www.theguardian.com/media/2017/nov/10/uk-prosecutors-admit-destroying-key-emails-from-julian-assange-case>.
- Maemura, Emily, Christoph Becker, and Ian Milligan. 2016. "Understanding Computational Web Archives Research Methods Using Research Objects." *Proceedings of IEEE International Conference on Big Data*, December 5–8, 2016, Washington, DC. <https://doi.org/10.1109/BigData.2016.7840982>.
- Maildir. 2018. Wikipedia. <https://en.wikipedia.org/w/index.php?title=Maildir&oldid=836696378>.

- Manjoo, Farhad. 2017. "What We Lose When the World Moves On From Email." *New York Times*, July 12. https://www.nytimes.com/2017/07/12/technology/what-we-lose-when-the-world-moves-on-from-email.html?_r=1.
- Marshall, Catherine C. 2008a. "Rethinking Personal Digital Archiving, Part 1: Four Challenges from the Field." *D-Lib Magazine* 14 (3/4). <http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html>.
- _____. 2008b. "Rethinking Personal Digital Archiving Part 2: Implications for Services, Applications, and Institutions." *D-Lib Magazine* 14 (3/4). <http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html>.
- Microsoft. 2005. "Exchange Storage Architecture." Microsoft Exchange Server 2003. [https://technet.microsoft.com/en-us/library/bb124808\(v=exchg.65\).aspx](https://technet.microsoft.com/en-us/library/bb124808(v=exchg.65).aspx).
- _____. 2011. "Compliance Features in Exchange Online." Last modified December 19, 2011. [https://msdn.microsoft.com/en-us/library/hh147162\(v=exchsrvcs.149\).aspx](https://msdn.microsoft.com/en-us/library/hh147162(v=exchsrvcs.149).aspx).
- Microsoft Developer Network. 2011. "Set Up and Manage Information Rights Management in Exchange Online." Microsoft Exchange Server. [https://msdn.microsoft.com/en-us/library/gg597271\(v=exchsrvcs.149\).aspx](https://msdn.microsoft.com/en-us/library/gg597271(v=exchsrvcs.149).aspx).
- Microsoft Developer Network. 2018. "Outlook Mail REST API Reference." Last updated April 5, 2018. <https://msdn.microsoft.com/en-us/office/office365/api/mail-rest-operations>.
- Microsoft Exchange Online. 2016. "Journaling in Exchange Online." [https://technet.microsoft.com/en-us/library/jj898487\(v=exchg.150\).aspx](https://technet.microsoft.com/en-us/library/jj898487(v=exchg.150).aspx).
- _____. 2017. "In-Place Hold and Litigation Hold." [https://technet.microsoft.com/en-us/library/ff637980\(v=exchg.150\).aspx](https://technet.microsoft.com/en-us/library/ff637980(v=exchg.150).aspx).
- Mimecast. 2018. "E-Discovery & Compliance." Accessed January 11. <https://www.mimecast.com/solutions/archive/e-discovery-and-compliance/>.
- Morgan, Steve. 2015. "Cybersecurity Market Reaches \$75 Billion In 2015; Expected To Reach \$170 Billion By 2020." *Forbes*, December 20. <https://www.forbes.com/sites/stevemorgan/2015/12/20/cybersecurity%E2%80%8B-%E2%80%8Bmarket-reaches-75-billion-in-2015%E2%80%8B%E2%80%8B%E2%80%8Bexpected-to-reach-170-billion-by-2020/#1d2d65a930d6>.
- Murray, Kate. 2014. "Shaking the Email Format Family Tree." *The Signal* (blog), April 4, 2014. Library of Congress. <http://blogs.loc.gov/thesignal/2014/04/shaking-the-email-format-family-tree/>.
- Murray, Kate, and Erin Engle. 2015. "We Welcome Our Email Overlords: Highlights from the Archiving Email Symposium." *The Signal* (blog), July 9, 2015. Library of Congress. <https://blogs.loc.gov/thesignal/2015/07/we-welcome-our-email-overlords-highlights-from-the-archiving-email-symposium/?loclr=blogsig>.
- NARA (National Archives and Records Administration). 1997. "Resources—Publications: Disposition of Federal Records—Chapter 6." In *Disposition of Federal Records: A Records Management Handbook*, 2nd ed. Washington, DC: Office of Records Services. <https://www.archives.gov/records-mgmt/publications/disposition-of-federal-records/chapter-6.html#VI.PermanentRecords>.
- _____. 2010. "Processing the Presidential Records of Elena Kagan." *AOTUS Blog: The Blog of the*

- Archivist of the United States, June 22, 2010. <https://aotus.blogs.archives.gov/2010/06/22/processing-the-presidential-records-of-elena-kagan/>.
- _____. 2013a. "Guidance on a New Approach to Managing Email Records." NARA Bulletin 2013-02. <https://www.archives.gov/records-mgmt/bulletins/2013/2013-02.html>.
- _____. 2013b. User Guide: Managing NARA Email Records with Gmail and the ZL Unified Archive, version 1.0. <https://www.archives.gov/files/records-mgmt/email-management/sample-agency-user-guide-for-managing-email.pdf>.
- _____. 2015. White Paper on The Capstone Approach and Capstone GRS. <https://www.archives.gov/files/records-mgmt/email-management/final-capstone-white-paper.pdf>.
- _____. 2018. National Archives 2018–2022 Strategic Plan. <https://www.archives.gov/about/plans-reports/strategic-plan>.
- National Digital Stewardship Alliance. 2013. "Levels of Digital Preservation," version 1. <http://ndsa.org/activities/levels-of-digital-preservation/>.
- NCDNCR (North Carolina Department of Natural and Cultural Resources). 2018. "Transforming Online Mail with Embedded Semantics (TOMES)." Accessed May 31. <https://www.ncdcr.gov/resources/records-management/tomes>.
- Never Bounce. 2018. Accessed May 31. <https://neverbounce.com>.
- NHPRC (National Historical Publications and Records Commission). 2018. "Transforming Online Mail with Embedded Semantics (TOMES)." Accessed May 31. <https://www.ncdcr.gov/resources/records-management/tomes>.
- Novell Documentation. 2018. "Information Stored in the Post Office." GroupWise 2014 R2 Administration Guide. Accessed May 31. https://www.novell.com/documentation/groupwise2014r2/gw2014_guide_admin/data/adm_poa_understand_post_office_info.html.
- Nuix. 2018. "EDRM Enron Data Set." Accessed May 31. <https://www.nuix.com/edrm-enron-data-set/edrm-enron-data-set>.
- Nylas. 2018. "The Nylas APIs." Accessed May 31. <https://docs.nylas.com/reference>.
- OAIS. 2012. "Space Data and Information Transfer Systems—Open Archival Information System (OAIS)—Reference Model." <https://www.iso.org/obp/ui/#iso:std:iso:14721:ed-2:v1:en>.
- Oard, Douglas, William Webber, David Kirsch, and Sergey Golitsynskiy. 2015. "Avocado Research Email Collection—Linguistic Data Consortium." Release date: February 16, 2015. <https://catalog.ldc.upenn.edu/ldc2015t03>.
- Oberhaus, Daniel. 2016. "How to Use the Internet on the Summit of Everest." Motherboard (blog), July 31, 2016. https://motherboard.vice.com/en_us/article/4xa4zp/when-the-internet-came-to-everest.
- Onishi, Norimitsu. 2017. "South African Court Raises Pressure for Zuma to Go." New York Times, December 29 (Africa). <https://www.nytimes.com/2017/12/29/world/africa/south-africa-court-zuma-impeach.html>.
- Organized Crime and Corruption Reporting Project. 2017. "#GuptaLeaks to Be Released to Journalists Worldwide." Press release, November 10. <https://www.occrp.org/en/40-press-releases/presss->

- releases/7240-guptaleaks-to-be-released-to-journalists-worldwide.
- Owens, Trevor. 2014. "The EPADD Team on Processing and Accessing Email Archives." *The Signal* (blog), October 20, 2014. Library of Congress. <http://blogs.loc.gov/thesignal/2014/10/the-epadd-team-on-processing-and-accessing-email-archives/>.
- Pace, Nicholas, and Laura Zakaras. 2012. *Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery*. Santa Monica, CA: The RAND Corporation. http://www.rand.org/content/dam/rand/pubs/monographs/2012/RAND_MG1208.pdf.
- Palarchio, Joe. 2015. "Office 365—Is the 'Archive Mailbox' Still Relevant?" *Perficient* (blog), June 29, 2015. <https://blogs.perficient.com/microsoft/2015/06/office-365-is-the-archive-mailbox-still-relevant/>.
- Paradigm Project. 2008. *Workbook on Digital Private Papers*. <https://www.webarchive.org.uk/wayback/archive/20080701041124/http://www.paradigm.ac.uk/workbook/index.html>.
- Pearce-Moses, Richard. 2005. *A Glossary of Archival and Records Terminology*. Chicago: Society of American Archivists. <http://files.archivists.org/pubs/free/SAA-Glossary-2005.pdf>. See also <https://www2.archivists.org/glossary>.
- PeDALs. 2013. "About PeDALs." Accessed via Wayback Machine March 6, 2013. <http://web.archive.org/web/20130306060835/http://www.pedalspreservation.org/>.
- Pennock, Maureen. 2006. "Curating Emails: A Lifecycle Approach to the Management and Preservation of Email Messages." In *DCC Digital Curation Manual*, edited by Seamus Ross and Michael Day. <http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-emails/curating-e-mails.pdf>.
- Phillips, Tom. 2018. "China Testing Facial-Recognition Surveillance System in Xinjiang." *The Guardian*, January 18. <https://www.the-guardian.com/world/2018/jan/18/china-testing-facial-recognition-surveillance-system-in-xinjiang-report>.
- Pinpoint Labs. 2018. Accessed May 31. <http://www.pinpointlabs.com/productsservices/software/harvester>.
- Plante, Jeanette. 2015. "Challenges of Email as a Record Archiving Email Symposium." Presentation at Archiving Email Symposium, Department of Justice, June 2, 2015. http://www.digitalpreservation.gov/meetings/documents/aes15/9_Plante_2015-06-02%20Library%20of%20congress%20email%20management.pdf.
- Pogue, David. 2004. "State of the Art; Google Mail: Virtue Lies In the In-Box." *New York Times*, May 13. <http://www.nytimes.com/2004/05/13/technology/state-of-the-art-google-mail-virtue-lies-in-the-in-box.html>.
- Pontevoipe, Gioanfranco, and Silvia Salsa. 2009. *General Study 05— Keeping and Preserving Email*. InterPARES 3 Project. http://www.interpares.org/ip3/display_file.cfm?doc=ip3_italy_gs05a_final_report.pdf.
- Pratt, Kathryn Mary "Kary." n.d. "The Developing Standards for Authenticating Electronic Evidence." Accessed May 31. http://laws-docbox.com/Legal_Issues/70550752-The-developing-standards-for-authenticating-electronic-evidence-kathryn-mary-kary-pratt.html.

- PREMIS Editorial Committee. 2008. PREMIS Data Dictionary for Preservation of Metadata, version 2.0. <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>.
- Princeton University Department of Rare Books and Special Collections. 2018. "Access Policy for University Archives Collections." Accessed May 31. <https://rbsc.princeton.edu/policies/access-policy-university-archives-collections>.
- Prom, Christopher J. 2011. Preserving Email. Great Britain: Digital Preservation Coalition. <http://dx.doi.org/10.7207/twr11-01>.
- Public Interest Declassification Board. 2012. Transforming the Security Classification System. Report to the President from the PIDB. <https://www.archives.gov/files/declassification/pidb/recommendations/transforming-classification.pdf>.
- Radicati Group, Inc. 2016. "Email Market, 2016–2020—Executive Summary." London, UK: The Radicati Group, Inc. http://www.radicati.com/wp/wp-content/uploads/2016/01/Email_Market_2016-2020_Executive%20Summary.pdf.
- Reagan Library. 2017. "Reagan Library Topic Guide—Iran-Contra Scandal." Reagan Library. <https://www.reaganlibrary.gov/sites/default/files/archives/textual/topics/iran-contra.pdf>.
- Resnick, Peter W., ed. 2008. "Internet Message Format. RFC 5322." Network Working Group, Internet Engineering Task Force. <https://tools.ietf.org/html/rfc5322>.
- Return Path. 2015. Deliverability Benchmark Report: Analysis of In-box Placement Rates in 2015. <https://returnpath.com/wp-content/uploads/2015/10/2015-Deliverability-Benchmark-Report.pdf>.
- Rockefeller Archive Center. 2006. E-Mail Guidelines for Managers and Employees. The Collaborative Electronic Records Project. http://www.nypap.org/wp-content/uploads/2016/04/rockefeller_email_guidelines.pdf.
- Rockmore, Dan. 2014. "The Digital Life of Salman Rushdie." *The New Yorker*, July 29. <https://www.newyorker.com/tech/elements/digital-life-salman-rushdie>.
- Rodden, Kerry, and Michael Leggett. 2010. "Best of Both Worlds: Improving Gmail Labels with the Affordances of Folders." In CHI EA '10 CHI '10 Extended Abstracts on Human Factors in Computing Systems, 4587–4596. Atlanta: ACM. <https://doi.org/10.1145/1753846.1754199>.
- Rosica, Jim. 2017. "Despite 'Questions,' Grand Jury Clears Andrew Gillum in Email Controversy." Florida Politics (blog), August 8, 2017. <http://floridapolitics.com/archives/242684-grand-jury-andrew-gillum>.
- Rothenberg, Jeff. 1999. Ensuring the Longevity of Digital Information. <https://www.clir.org/wp-content/uploads/sites/6/ensuring.pdf>.
- Rothenberg, Jeff. 2000. "Preserving Authentic Digital Information." In *Authenticity in a Digital Environment*, 51–68. Washington, DC: Council on Library and Information Resources. <https://www.clir.org/pubs/reports/pub92/rothenberg/>.
- Rouse, Margaret. 2018. Definition: Electronic Discovery (e-discovery or ediscovery). Accessed May 31. <http://searchfinancialsecurity.tech-target.com/definition/electronic-discovery>.
- Sabato, Larry. 1998. "The Iran-Contra Affair—1986–1987." *Washingtonpost.com* Special Report: Clinton Accused. <http://www.washingtonpost.com/wp-srv/politics/special/clinton/frenzy/iran.htm>.

- Simpson, Joel. 2016. Email Archiving Systems Interoperability. Harvard Library Report.
https://dash.harvard.edu/bitstream/handle/1/28682572/HL_Email_Archiving_Systems_Interoperability_Report_2016.pdf?sequence=3.
- Society of American Archivists. 2011. "SAA Core Values Statement and Code of Ethics."
<https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics>.
- Spangler, Todd. 2017. "April the Giraffe's Live Birth Drew 14 Million YouTube Views on One Day."
 Variety, April 17. <http://variety.com/2017/digital/news/april-giraffe-live-birth-youtube-1202032345/>.
- SparkPost. 2015. Market Guide for Email Marketing. Stamford, CT: Gartner, Inc.
<https://pages.messagesystems.com/Gartner-WP-Down-load-Landing-Page.html?src=Blog&pc=BL-GD-GartnerEmailGuide>.
- Stanford University. 2018. "EPADD-Documentation." Stanford Libraries. Accessed May 31.
<https://library.stanford.edu/projects/epadd/documentation>.
- Sustainability of Digital Formats. 2013a. "Microsoft Outlook PST 97-2002 (ANSI)." Library of Congress Digital Preservation. Last significant FDD update November 25, 2013.
<http://www.digitalpreservation.gov/formats/fdd/fdd000377.shtml>.
- _____. 2013b. "Microsoft Outlook PST 2003 (Unicode)." Library of Congress Digital Preservation. Last significant FDD update November 25, 2013.
<http://www.digitalpreservation.gov/formats/fdd/fdd000378.shtml>.
- _____. 2015. "Lotus Notes Storage Facility." Library of Congress Digital Preservation. Last significant FDD update April 4, 2018. <https://www.loc.gov/preservation/digital/formats/fdd/fdd000433.shtml>.
- Task Force on Technical Approaches for Email Archives. 2018a. "Exploring Email Emulation."
<http://www.emailarchivestaskforce.org/documents/exploring-email-emulation/>.
- _____. 2018b. "Guide to Email Standards." <http://www.emailarchivestaskforce.org/documents/guide-to-email-standards/>.
- _____. 2018c. "Managing Email for Preservation."
<http://www.emailarchivestaskforce.org/documents/managing-email-for-preservation/>.
- _____. 2018d. "Email User Features." <http://www.emailarchivestaskforce.org/documents/email-user-features/>.
- _____. 2018e. "Email Archiving Tools." <http://www.emailarchivestaskforce.org/documents/email-tools/>.
- The Coalition of Technology Resources for Lawyers. 2016. 2016 Guidelines Regarding the Use of Technology-Assisted Review. <http://ctrlinitiative.com/wp-content/uploads/2014/07/2016-Guidelines-Regarding-the-Use-of-Technology-Assisted-Review.pdf>.
- The Document Foundation Wiki. 2018. "Feature Comparison: Mozilla Thunderbird–Microsoft Outlook." Last edited May 9, 2018.
https://wiki.documentfoundation.org/Feature_Comparison:_Mozilla_Thunderbird_-_Microsoft_Outlook.
- The National Archives. 2016. The Application of Technology-Assisted Review to Born-Digital Records Transfer, Inquiries and Beyond. United Kingdom: The National Archives.
<http://www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records->

transfer.pdf.

- Underwood, William, Marlit Hayslett, Sheila Isbell, Sandra Laib, Scott Sherrill, and Matthew Underwood. 2009. Advanced Decision Support for Archival Processing of Presidential Electronic Records: Final Scientific and Technical Report. Technical Report ITTL/CSITD 09-05. Atlanta, GA: Georgia Tech Research Institute. <http://perpos.gtri.gatech.edu/publications/TR%2009-05-Final%20Report.pdf>.
- U.S. Government Accountability Office. 2008. "Federal Records: Agencies Face Challenges in Managing E-Mail," GAO-08-699T (April 23). <https://www.gao.gov/products/GAO-08-699T>.
- U.S. Securities and Exchange Commission. 2002. "The Sarbanes-Oxley Act of 2002." <https://www.sec.gov/about/laws/soa2002.pdf>.
- Vaudreuil, Gregory M., and Glenn Parsons. 2004. "Voice Profile for Internet Mail Version 2 (VPIMv2). RFC 3801." Network Working Group, Internet Engineering Task Force. <https://tools.ietf.org/html/rfc3801>.
- Vogel, Karl, and Charles Cazabon. n.d. Database Back-End for Large Email Systems. Memoryhole. Accessed May 31, 2018. <http://www.memoryhole.net/~kyle/databaseemail.html>.
- Waters, Donald, and John Garrett. 1996. Preserving Digital Information: Report of the Task Force on Archiving of Digital Information. Washington, DC and Mountainview, CA: The Commission on Preservation and Access and The Research Libraries Group. <https://www.clir.org/pubs/reports/pub63/>.
- Weiss, Debra Cassens. 2017. "Law Firm's Automatic Deletion of Spam Emails Is Blamed for Failure to File Timely Appeal." ABA Journal, September 28. http://www.abajournal.com/news/article/law_firms_automatic_deletion_of_spam_emails_is_blamed_for_failure_to_file_t/.
- Whitt, Richard S. 2017. "'Through a Class Darkly' Technical, Policy, and Financial Actions to Avert the Coming Digital Dark Ages." Santa Clara High Technology Law Journal 33 (2): 117–229. <http://digitalcommons.law.scu.edu/chtlj/vol33/iss2/1>.
- Wikipedia. 2016. Lorraine v. Markel American Insurance Co. https://en.wikipedia.org/w/index.php?title=Lorraine_v._Markel_American_Insurance_Co.&oldid=709391592.
- _____. 2017a. "Emulator." <https://en.wikipedia.org/w/index.php?title=Emulator&oldid=815244333>.
- _____. 2017b. "Comparison of Webmail Providers." https://en.wikipedia.org/w/index.php?title=Comparison_of_webmail_providers&oldid=816032285.
- Zdziarski, Jonathan. 2008. "Email Database." In iPhone Forensics, 1st ed., 79. Sebastopol, CA: O'Reilly Media, Inc. https://books.google.com/books?id=R1XArTHPn9QC&pg=PA79&lpg=PA79&dq=how+are+email+messages+stored+in+database+tables?&source=bl&ots=_gzF-o6Fpsn&sig=NuJRndLZwmi7wAlvpuHTBd4oNbA&hl=en&sa=X&ved=0ahUKEwiH6-Dnr3UAhUGPT4KHei8A4I4ChDoAQghMAA#v=onepage&q=how%20are%20email%20messages%20stored%20in%20database%20tables%3F&f=false.
- Zhang, Jie, Debra Logan, and Garth Landers. 2014. Magic Quadrant for E-Discovery Software. ID: G00260499. Stamford, CT: Gartner, Inc.

<https://www.globanet.com/sites/default/files/resources/Gartner%20Magic%20Quadrant%20for%20eDiscovery%20Software%202014.pdf>.

Zierau, Eld, and Sébastien Peyrard. 2016. "Digital Preservation Meta- data in a Metadata Ecosystem." In *Digital Preservation Metadata for Practitioners*, 189–211. Springer. <https://doi.org/10.1007/978-3-319-43763-7>.