

email を歴史資料にするために：

OSS・ePADD のススメ

堀内 暢行

はじめに

「20XX 年、世界の日本現代史研究者は、史料の真贋問題から研究することができなくなった。」

これを読んだ多くの人は、「単なる冗談だ」と受け取るだろう。または、「そんなことになるはずが無い」と目にとめることもしないかもしれない。しかし、多くのアーカイブズ学研究者はそう遠くない未来にそれが起こる可能性を危惧し、それを回避するために研究を続けている。

社会生活において、IT（Information Technology）はかけがえのない技術となった。それにともない、記録は紙ベースから電子情報に取って代わられた。こうした状況は、ICT（Information and Communication Technology）や IoT（Internet of Things）といった用語が社会で常用されるようになった現在を鑑みれば、ますます進化していくであろうことについて疑念を挟む余地はない。

*本論は 2021 年度「デジタルアーカイブ学会第 6 回研究大会：一般研究発表セッション 3」にて筆者と橋本陽（京都大学大学文書館特定助教）による共同報告「電子メールの長期保存の課題：ePADD の有用性と問題点」を元としている。その際 email の長期保存の必要性やアーカイブズ学的方法の詳細については、橋本が担当した。この点については、本論をすすめる上で必要な箇所を同氏の論を参照し、助言を得たが文責は筆者にある。なお、同報告ファイルは https://www.jstage.jst.go.jp/article/jsda/5/s1/5_s55/_article/-char/ja/ を参照のこと。

email を歴史資料にするために：OSS・ePADD のススメ

本論の目的は、こうした状況に適応したアーカイピングのシステムを紹介することにある。とはいえ、電子記録情報全般をその射程におくことは、筆者の知識では到底対応することはできない。そこで本論では電子記録のうち、email に限定して考えてみたい。また、当該問題は世界的に共有しているといえるが、筆者が関係してきた経験と紙幅の関係から日本国内に限定する。

email に限定する理由の一つは、社会活動において公私ともに email は欠かせないコミュニケーション・ツールになっていることがあげられる。日本国内の企業活動においては対面や電話よりも email を望む傾向が強いようだ⁽¹⁾。さらに COVID-19 まん延防止策の一環としてリモートワークが中心となったことから、その利用率は増加傾向にある⁽²⁾。email が日本社会で主流となったのは、インターネットが普及した 2000 年以降であった。企業だけではなく、インターネットに接続できる携帯電話を個人が持つことで、その利用率は右肩上がりが増えていった。こうした状況にありながら、企業・個人間でやりとりされた email データの保存システムについて、日本社会、特に企業では等閑視されてきたように思われる。というのも、筆者はこれまで 20 社以上の企業で資料整理活動に従事してきたが、多くの企業で email がその対象になることはなかった。もちろん、資料整理従事者とはいえデータ・サーバーに外部の人間がアクセスすることは許容されないであろう。しかし、その必要性を提案しても、その重要性は認識されるものの、それが適ったことは未だに無かったこともまた事実である。といっても、email を整理した経験は数回あった。しかしながらそのどれもが、email データをプリントアウトした「紙」であった。これでは、email 情報を記録として保存したとはいえない。電子記録としての email を保存管理するには、アーカイブズ学に即した手順が必須となる。email が「紙」となったことを考慮しても、それだけでは記録としての真贋を諮ることができないことはもちろん、証拠能力を保った状態での長期保存に適さないからである。付言すると、仮に筆者がデータ・サーバーへのアクセス権を付与

されたところで、そのノウハウが無いことからアーカイブズ学的長期保存に適応した整理はできなかったであろう。そもそも、日本国内で email を適切に保存している機関は管見のところ皆無である。しかし、それはあくまで日本国内に限定した話であり、世界に目を向けるとすでに多くの研究蓄積があり、それは日々進展している。政府機関はもちろん、図書館やアーカイブズでは電子記録の長期保存や管理が実践されている。

こうした取組の遅れや、日本の問題を打開するために、本論では email のアーカイブズ学的整理・長期保存を目的に開発されたアプリケーション・ソフトウェアである ePADD の有用性を紹介するとともに、それを運用する際の現段階における課題も提示することとしたい。

1. 国際基準における記録——email の位置づけ

そもそも email とは、アーカイブズ学においてどのような位置づけにあるのだろうか。

2016 年 4 月、国際標準化機構（International Organization for Standardization: ISO）は記録管理に関する国際標準として規定された ISO 15489-1 の第 2 版（Information and documentation-Records management-Part 1: Concepts and Principles. 「情報及びドキュメンテーション—記録管理—第 1 部：概念及び原理」：以下、第 2 版）を発行した⁽³⁾。同標準の「3.14 記録」は、「法的な義務の遂行において又は業務の処理において、組織又は個人によって証拠及び資産として作成、受領及び維持された情報」（「3. 用語及び定義」）と定義している⁽⁴⁾。また、記録の管理は、

A) 業務活動の証拠となるための要求事項を満たすために、記録を作成及び捕捉すること。

B) 記録管理のための業務コンテキスト及び要求事項が時間とともに変化

することに伴い、記録の真正性、信頼性、完全性及び使用性を保護する適切な行動を取ること。

を包含するとした⁽⁵⁾。さらに「5.2.2 正式な記録の特性」では、真正性・信頼性・完全性・使用性という4つの特性についてそれぞれ規定している⁽⁶⁾。これらをまとめると、記録の適正な管理とは、これら4つの特性を備えた状態で記録を作成し維持していくということにほかならない。

email は何かしらの業務遂行上に作成されるものであり、まさしく「記録」にカテゴライズされる。よって、email を管理する際にも上記4特性の規定を遵守することが必須となる。ここで注意すべきことは、email が電子形態であるということである。「はじめに」で示したように、email をプリントアウトし、紙ベースで管理することも、一見すると上記4特性を維持しつつ行うことが可能のように思われる。しかし、ことはそう簡単ではない。

2003年にISOは記録管理と同様に、電子記録の長期保存についても推奨プラクティスを提示した。それが、OAIS参照モデル（ISO 14721:2003 Reference Model for an Open Archival Information System (OAIS)）である。これは2012年に改訂（ISO 14721:2012）された⁽⁷⁾。OAIS参照モデルは、ICTの使用を前提としたものではなく、アナログ情報を含めた情報システム（IS）の体系的な仕組みについて規定している⁽⁸⁾。OAISの開発経緯については紙幅の関係上多くを触れることはできないが、OAIS参照モデルが開発される以前からデジタル記録の長期的保存について宇宙データシステム諮問委員会（Consultative Committee for Space Data Systems: CCSDS）をはじめ、様々な団体がそれぞれに協議を重ねていた。彼らは後に自らの問題感心が多分野でも大きな範囲で共有されていることを認識し、結果、CCSDSが主導的立場となり、同参照モデルは開発された⁽⁹⁾。また、OAIS参照モデルの策定の契機が、時間が経過した電子情報の読み出しができなかったことにあり、電子情報の記録管理に適用される事例が散見される。こうした経緯からわかるように、

OAIS 参照モデルを利用することがデジタル記録の長期的保存に欠かせないことがわかっていく。

OAIS 参照モデルでは電子記録を保存管理する際、情報の再現性を保持する必要性が提示されている。これは、情報コンテンツを長期保存した際に、常に読むことが可能な状態を保ち続けることであり、且つそれは含まれているオブジェクトや構造をも包含したものでなければならないということを意味している。具体的には保存対象となる情報コンテンツに対する表現情報（Representation Information）を付し、さらに保存記述情報（Preservation Description Information）を追加して単一の情報パッケージ（Information Package）として保存することが求められる。表現情報には、情報コンテンツに関する内容情報やファイル構造等、保存記述情報には来歴やコンテキスト・不変性情報といった項目がある。以上から、email をプリントアウトした場合、保存記述情報など、再現性を担保するために必要な情報を付すことができず、電子記録を保存管理しているということにはならない。

しかし、email に限らず電子記録の長期保存を実現するために OAIS 参照モデルに準拠したアプリケーションを一から作ることは容易ではない。もちろん、可能ではあるものの、単なるクラウドを用いた電子記録の保存管理アプリケーションでさえも、莫大な開発・導入費用を要する状況にあって、それぞれの機関がそれらに予算を付けることは、現実的ではない⁽¹⁰⁾。また、一度システムを導入すると、他のシステムに変更することは容易ではなく、大概の場合、それまで構築した全てのデータを破棄し、新たにデータを構築し直す必要がある。

そうした問題を解決するために、OSS（Open Source Software）として無償で配付されている email の保存管理用アプリケーション・ePADD の利用を推奨したい。

OSS を導入する上での最大の問題は、多くの機関が利用することを忌避し

email を歴史資料にするために：OSS・ePADD のススメ

ていることにある。セキュリティの問題を考慮した結果であるが、実は OSS を完全に忌避できている機関は皆無だと思われる。すでに導入している多くのシステムには OSS が組み込まれた仕様になっているからである。例えばデータベースを構築する際には MySQL が組み込まれている場合が大半であろう⁽¹¹⁾。MySQL もまた、誰もが使用できるものであり、OSS の代表的な存在である。この例からわかるように、気づかいうちに OSS が組み込まれたシステムをすでに運用しているのである。したがって、OSS の運用の仕方によりセキュリティ問題を解決することが可能であり、過敏な対応をとる必要はない。

2. ePADD の機能と特徴

2.1 ePADD の開発経緯

ePADD は、アメリカ・カリフォルニア州にあるスタンフォード大学 (Stanford Univ.) 図書館が 2010 年に開発を始め、National Historical Publications & Records Commission からの資金を得てフルバージョンを配付した (2015 年)。その後も、the Institute of Museum and Library Services やアンドリュー・W・メロン財団、そしてその助成を受けたイリノイ大学のプログラム (Email Archiving : Building Capacity and Community) などからの資金を得て現在も開発を継続している⁽¹²⁾。

2.2 基本的特徴

繰り返しになるが、ePADD はスタンフォード大学図書館が OSS として開発・公開していることが最大の特徴である。民間企業が開発・販売しているアプリケーションとは異なり、OSS であることから、誰もが無料で利用できることは大きな利点である。同様に、OSS で公開されていることは、ユーザー

のニーズに適応したアプリケーションに改変することが可能である。これは、仮に開発者であるスタンフォード大学図書館が開発を停止しても利用を継続し、アップデートすることができることを意味している。システム面から確認すると、ePADD は Java 及び JavaScript の両プログラミング言語が用いられている。よって、ユーザーは開発側の意思に左右されることなく恒久的に利用が可能となる。

そして、ePADD は容易に導入できるよう設計されていることは大きなインセンティブとなる。一般的に利用されている Windows・Mac・Ubuntu といったオペレーション・システム（OS）に対応している。さらに汎用 Web ブラウザー（Google Chrome・Mozilla Firefox 等）のユーザー・インターフェース上で操作が可能であることから、PC に苦手な人の技術的障壁を除去してくれるであろう。

ePADD のメリットは導入・操作の容易性だけにとどまらない。email をアーカイブズが受け入れる際に行う作業毎にモジュールが搭載されている。史料、特に紙資料を受け入れる際、評価選別を行い、整理作業（目録搭載）及びユーザーへの提供という手順が通例である。こうした作業を一つのアプリケーションで行えるよう ePADD は設計されている⁽¹³⁾。具体的には、“モジュール”と呼ばれる機能単位で作業できるような仕様となっている。

それぞれのモジュールについては、下記のとおりである。

- 評価選別（Appraisal）モジュール：email の保存に入る前に作成者・資料提供者や受入機関員がデータ内容を確認し、評価選別作業を行うことができる。
- 整理（Process）モジュール：email のデータを整理し、アーカイブズに必要な記述を付すことができる。
- 開示（Discovery）モジュール：アーキビストなどの特定の制限下にあるユーザーがデータの事項群などを検索できるとともに、レポジ

email を歴史資料にするために：OSS・ePADD のススメ

トリにデータを送ることができる。

- 提供 (Delivery) モジュール：アーカイブズとして email の全文にアクセスすることができる。

長期保存については、処理データを“MBOX”形式（電子メールの保存ファイルフォーマット）でエクスポート可能な仕様となっている⁽¹⁴⁾。この機能を介し、OAIS 参照モデルで示される長期保存の要件を満たしながら、大量の電子メールを保存できる。例えば、Artifactual Systems 社が提供する電子データの長期保存用アプリケーション OSS・archivematica⁽¹⁵⁾に MBOX ファイルをアップロードし、email データの長期保存を行っている機関が散見される⁽¹⁶⁾。つまり、こうした機関は archivematica を用いて、ePADD の仕様に含まれない機能の OAIS 参照モデルのプロセスを遂行している。このように、ePADD で評価選別を行った後に電子データの長期保存用アプリケーションを併用することで、email データを確実に保存管理することが可能となる。

また、独自に開発した自然言語処理 (Natural Language Processing: NLP) 機能を搭載し、固有名詞や特定項目表現の抽出・処理が実行できる仕様となっている。

次に、それぞれのモジュールの機能について ePADD の動作の手順を踏まえつつ確認していくことにする。

2.3 モジュールの機能と ePADD の操作

ePADD をインストールする際に難しい作業は特にはない。公開されているマニュアル通りに操作を行えば容易に email の整理・公開環境を構築することが可能となる⁽¹⁷⁾。インストール後、アプリケーションを起動すると、Web ブラウザーが開き、インポートを行う email の設定を“Import”画面にて行う必要がある。その際、Google や Yahoo、Microsoft などが提供している、いわゆる Web メール (Public Email Accounts) とプロバイダーなどが提供している

email アカウント（Private Email IMAP Accounts）が区別されている点に注意する必要がある。また、MBOX 形式のファイルもここでインポートする機能が搭載されている。よって、ePADD 以外のアプリケーションで同型式にアウトプットしたファイルを ePADD に替えて整理作業を行うことができる。

email のインポートが完了すると、4つの行程をそれぞれのモジュールに切り替えて作業を行うこととなる。この行程において評価選別が最も重要であり、大変な作業となる。email を書簡資料に置き換えると判りやすい。アーカイブズに移管される資料は、組織やその内部部署、そして個人など様々である。しかし、それらはその量に関係なく、アイテムレベルまで詳細な目録作成を行うことが日本国内では通例となっている。書簡資料もちろん、アイテムレベルで目録を作成する。したがって、その量によって、作業は膨大な時間とマンパワーを必要とするが、その結果、年賀状や礼状など「記録」であることを否定するものではないにしろ、費やした作業量に比して得られる情報が僅少であるアイテムが多数にのぼることがある。個人文書に含まれる、個人宛領収書や送り状といったものを想像するとわかりやすいであろう。しかし、そうした内容が判明するのは、整理作業を行った結果であることは言うまでもない。結局、悉皆的整理をすることを是とする限り、これは致し方ないであろう。

それでは、email に対しても同様に作業をおこなうことが可能であろうか。もちろん、時間的制約がなければ可能であろう。しかし、現代人の生活環境において email には全ての情報が送受信されているといっても過言ではない状況にある。例えば、個人の email アカウントにある受信 BOX には、友人や家族からの連絡と同時にインターネット通販の購入・発送通知や、契約しているありとあらゆるところから広告が届く。企業では、休日明けの出勤時には必要な email と職務に関係のない“Junk”や“Noise”を振り分けるだけでも大変な作業となっている。それらがきちんとマネジメントされずに移管されてきた場合、

email を歴史資料にするために：OSS・ePADD のススム

書簡資料と同様の整理を行うことが出来るであろうか。仮に、評価選別の作業において、email の概要を把握し、振り分けることができれば、そうした徒労を避けることができるのではないか。

上記の仮定を実現させたのが、評価選別モジュールである。次の図 1 を見てもらいたい。

同図は、筆者の email アカウントの一つをテスト用（TEST_Collection）にインポートした際の Dashboard である。これを見るとわかるように、email アカウントで作成した分類フォルダは、“Folders” にそのまま反映されるなど、email アカウントの状態がそのまま可視化される。

さらに、“Correspondents（図 2）”を開くと、email をやり取りした相手毎の数量が表示される。email アカウントの持ち主との関係が数量的に表示され、さらにそれを、送受信量とその年代をグラフ化することができる（図 3）。

次に、“Entities”を選択すると、“Place（地名）”や“Person（人）”などの項目を email から抜き出し、その数量が表示される（図 4）。表示された項目

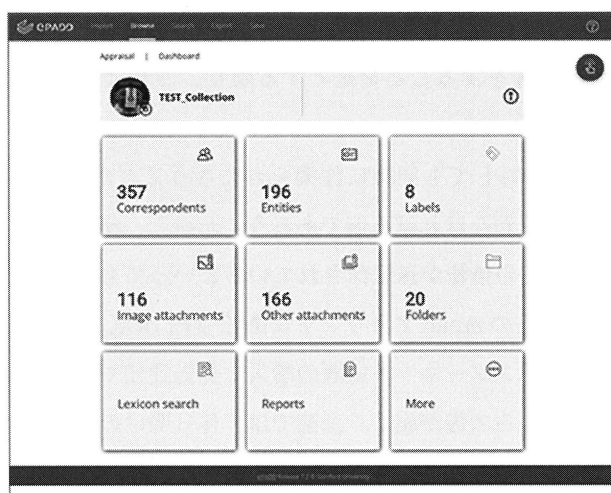
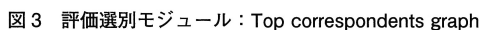
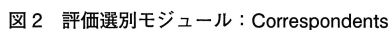
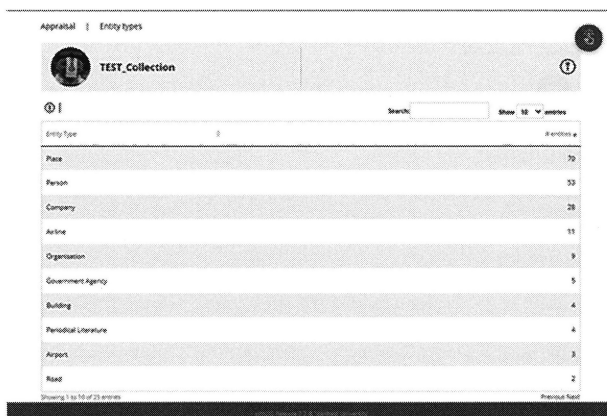


図 1 評価選別モジュール：Dashboard

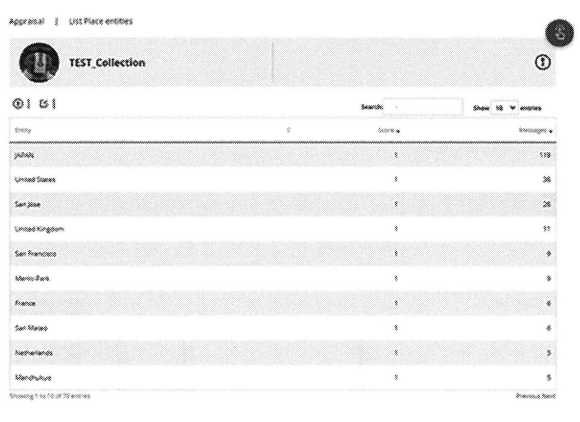


email を歴史資料にするために：OSS・ePADD のススメ



Entity Type	# entities
Place	70
Person	53
Company	28
Airline	11
Organisation	9
Government Agency	5
Building	4
Periodical Literature	4
Airport	3
Road	2

図 4 評価選別モジュール：Entity types



Entity	Score	Relevance
Japan	1	119
United States	1	36
San Jose	1	26
United Kingdom	1	11
San Francisco	1	9
Marin Park	1	9
France	1	4
San Mateo	1	4
Netherlands	1	3
Marshall Islands	1	3

図 5 評価選別モジュール：Entity types> List Place entities

を選択すると、その詳細が表示される（図 5）。

図 3～5 の Entity types にて可視化された項目は、“Named Entities Recognizer (NER)” と呼ばれるエンティティ認識機能の恩恵である。NER は email のアドレス・ヘッダーやテキスト本文から人物・組織・場所など、様々なワードを抽出する機能となっている。この機能により、膨大な email の

評価選別を行う際に、可視化された情報がその手助けをしてくれる。

以上のように、Appraisal モジュールでは、email のやり取りや、その内容の詳細までもが可視化され、メール本文を確認できるような仕様となっている。また、それらが先述したような評価選別における有用な機能であることがわかる。ここでいう有用性は、email の移管を受けたアーキビストだけに留まらない。資料の寄贈者にとっても重要であり、email アカウントとドナーが異なる場合は、大きなアドバンテージとなる。

例えば、故人の親族が寄贈者となり文書を受け入れる場合、開示・非開示はもちろん、その前段階において、どのような資料（email）があるのかを確認し、個人的な email——例えば、資産情報などを評価選別の段階で把握し、移管文書から除外または非開示にすることが可能となる。つまり、紙ベース資料の整理段階において、把握することが困難であった個人的情報を容易に選別することができることから、アーキビスト・ドナー（関係者）両者が共同で確認しあいながら作業を行うことができる。これは、電子記録ならではの利便性であり、有用な機能となっている。

また、評価選別の後の整理モジュールでは email のアーカイブズに必須の記述を付すことができる（図 6）。こうした記述を評価選別からの連続的動作で行うことができるのもまた効率的な面から考えると有用であると言える。

一方で、ここで注意が必要なことは、ePADD 上での作業は、その対象はあくまで email に限定されていることにある。通例、email のみが移管されることはなく、フォンドの 1 シリーズとして書簡があるのと同様に、email もまた、電子記録群のなかの 1 シリーズである。よって、記述する際に、フォンドに合致したものにする必要がある。

整理モジュールで必要な記述を付した上で、開示モジュールに進める。このモジュールは、Web サーバー上で動作するように設計されているため、ユーザーガイドに従って、Web サーバーにマウントする必要がある⁽¹⁸⁾。作業を終

email を歴史資料にするために：OSS・ePADD のススメ

The screenshot displays the 'test-collection' interface in the ePADD system. At the top, there are links for 'Processing' and 'About this collection'. Below this is a large, empty rectangular area with a copyright symbol in the top right corner. The main interface is divided into two columns. The left column, titled 'Summary', contains a list of metadata fields: Institution (myhome), Repository (myhome), Collection ID (test-001), Collection Title (test-collection), Date Range (2007-02-04 to 2020-12-29, 328 messages updated), Messages (2,313, Sent by owner: 668), Attachments (282, Images: 156, Documents: 956, Others: 60), and Contact email (hori@92mail.com). The right column, titled 'test-collection', features a search bar and several tabs: 'test case', 'Scope and Content', 'Rights and Conditions', and 'Notes'. The 'test case' tab is currently selected, showing a list of test cases. Below the tabs is a section for 'Add accession' with an 'ENTER' button.

図6 整理モジュール

えると、整理を終えた記録にリモートユーザーとしてアクセスすることができ、エンティティやemailのテキストメッセージを検索することが可能となる。このモジュールを使用するのは、emailを一般ユーザーに提供する前段階であり、除外措置等の確認を行う際に有効である。

開示モジュールでの作業を終えると、ようやく提供モジュールに進めることができる。これまでのモジュールでの作業で問題がなかったemailに、制限のないフルテキストアクセスをユーザーに提供することができる。もちろん、そのなかでユーザーが希望するemailがヒットするよう、検索することが可能となっている。

3. 運用における諸問題

以上、見てきたように、ePADDのモジュールはアーカイブズ業務において

有用な機能を有している。一方で、運用にあたり色々な問題があることも分かっている。これは、特に日本語で使用する際に大きな障壁となる。

第一に、全ての表記が英語であることは、これまでに提示した図から分かるであろう。この点については、日本語に表記を改変することで解決できるため、大きな問題とはいえない⁽¹⁹⁾。ここで問題視しなければならないのは、ePADD の根幹的機能といえる、email 本文の選別機能が日本語対応していない点である。具体的には評価選別モジュールにおける“Lexicons”と NLP がその対象である。

3.1 Lexicons の日本語対応

前章で“Lexicon Search”を Dashboard の図上で確認したが、本機能を利用することで、評価選別作業はさらに効率的に進めることが可能となる。この機能は、個人情報の項目を設定し、それらを含む内容が email にあるか否かを確認できるようになっている。例えば犯罪情報などの用語や個人の財産情報が入っているかどうかといったアーカイブズで提供する際に制限する必要があるものについて、把握できるよう関連用語を設定する仕様となっている。この項目に、日本語で設定したものはない。よって、こちらが設定ファイルを作成し、ePADD に読み込ませる必要がある。

そこで、ここでは“test.japanese.lex.txt”というテキストファイルを作成し、そもそも日本語に対応するのかを試みた（図 7）。あくまで試論ということで、読み込ませるファイルには簡易的に「家族」・「職場」・「仕事」・「プライベート

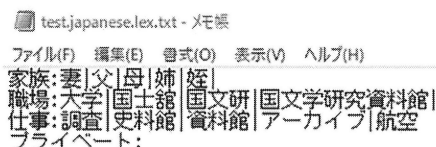


図 7 “test.japanese.lex.txt”

email を歴史資料にするために：OSS・ePADD のススメ

ト」という項目を設定し、それぞれに関連ワードを追記した。本テキストファイルを、“epadd-appraisal> user> data> lexicons”のディレクトリに置き（図8）、改めて“TEST_Collection”を読み込ませた。その結果、Appraisal モジュールの“Lexicons”に反映され、図9のような結果を得た。同図からわかるように、“test.japanese.lex.txt”で設定した4項目とそれぞれの関連ワード反映されている。

ここでは、あくまで“Lexicons”が日本語に対応するのかという目的で設定ファイルを作成したが、この機能を用いることで、評価選別の作業を飛躍的に向上できると考える。例えば、email アカウントの持ち主の経歴や親族の氏名など事細かに設定することで、email 全体の内容を分類・把握することが可能となる。繊細な問題についても同様である。とはいえ、一般的な地名や人物名等についても作成するとなると、大変な労力となる。そこで、有効な打開策となりうるのがNLPの機能である。

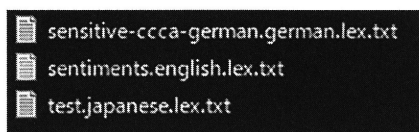
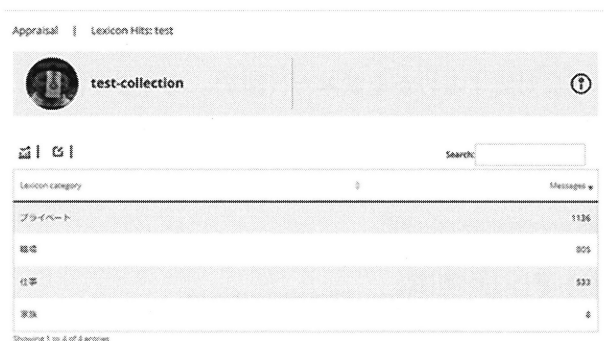


図8 “test.japanese.lex.txt”のファイル設置

A screenshot of the 'Appraisal' module interface. The top bar shows 'Appraisal | Lexicon Hits: test'. Below this is a header for 'test-collection' with a user icon and a help icon. The main area contains a table with 'Lexicon category' and 'Messages' columns. The table lists four categories: 'プライベート' (1136), '結婚' (805), '仕事' (523), and '家族' (4). A search bar is located at the top right of the table area. At the bottom, it says 'Showing 1 to 4 of 4 entries'.

Lexicon category	Messaget
プライベート	1136
結婚	805
仕事	523
家族	4

図9 Appraisal モジュール：Lexicons> Lexicon Hits: test

3.2 NLPの日本語対応

NLPとはそもそもどのような機能なのだろうか。NLPは自然言語処理(Natural Language Processing)の頭文字であることは先に触れた。自然言語とは、我々が日常用いる自然発生的に生まれた言語であり、それに対して人工的に作られたプログラミング言語などを人工言語と呼ばれている。そして自然言語のコンピューター処理に関する学問・研究分野をNLPと定義されている⁽²⁰⁾。

NLPを実装することで、ePADDでは何ができるのかということになるが、“User Guide”において詳細は触れられておらず、「カスタムの自然言語処理ツールキットを用いて、エンティティを識別、抽出、管理する」といった説明にとどまっている⁽²¹⁾。とはいえ、NLPはより日常的に使用される広範囲な用語等を読み取り、そのなかから必要な用語を抽出する機能であることは間違いがなく、“Lexicons”と併用することで、さらにemail群の内容を詳細に把握することが可能となり、重要な機能といえる。

そこで、“Lexicons”と同様に、NLPの日本語化を試みた。調べてみると、“src> java> edu> stanford> muse> ner> dictionary> EnglishDictionary.java”にファイルが実装されていた。そこには、英語以外の言語においても同様に設定すれば対応する、といったことがプログラム上に記載されていた⁽²²⁾。

しかし、ファイルを追加したものの日本語化を進めることができなかった。これは、日本語だけに限らず、英語以外、どの言語も同様だと思われる⁽²³⁾。よって、早急にこの問題を解決する必要があるが、当該ディレクトリにファイルを追加して処理させるだけで対応できるのか否かも不明である。そのためには、関連するプログラムコードを確認しつつ、該当箇所以外にプログラムの改変等が必要な場合も予想されることから、開発者であるスタンフォード大学等と連絡する必要がある。

おわりに

以上、email を整理保存する上で OSS アプリケーション ePADD の有用性を確認してきた。日本国内では、email のアーカイブズの保存について議論が進まない一方で、email は個人・団体・政府と社会全体で利用されている重要なツールであることから、少しでも問題の解決に寄与したいという考えから本論を執筆した。

とはいえ、詳しく確認していくなかで、ePADD を実際に運用するにあたって色々な問題があることもわかった。例えば、NLP の日本語化や “Lexicons” のパッケージをどのように作成するのか指標が無いなかで、各機関が ePADD をどのように運用していくのかという問題がある。そもそもの問題として、email はおろか、電子記録の保存管理に対応している機関が僅少である国内状況において、email に焦点を絞った問題提起をすることじたいが時期尚早といえるかもしれない。

しかしながら、そうした問題の解決を待つ間に、email は保存されずに消えていくのではないか。日本語化対応については、日本のアーカイブズ学研究において電子記録を専門に研究するシステム・エンジニアが一人でもいれば恐らく解決できるだろうが、無いものを待ち、嘆くよりも、まずは一歩を踏み出すことを重視したい。

ところで、本稿を執筆している際に、筆者が受け持っているアーカイブズ学を受講する学生の多くが、日常では email ではなくソーシャルネットワークサービス（SNS）の各種チャットアプリを利用していることを耳にした。人々を結ぶツールは、予想以上のスピードで転換している。それに抗うかのようにみえるが、それでも「記録」を残すことを諦めずに取り組んでいきたいし、少しでも解決にむかうような提案をし続けていきたいと考える。

注

*本稿で提示した URL の最終確認日は 2022 年 1 月 10 日である。

- (1) アドビシステムズ(株)「2019 Consumer Email Survey」(2019年10月10日、同社 Web サイト：https://www.adobe.com/jp/news-room/news/201910/20191010_2019-consumer-email-survey.html)。
- (2) 一般社団法人ビジネスメール協会「ビジネスメール実態調査」(2021年6月1日、同協会 Web サイト：<https://businessmail.or.jp/research/2021-result/>)。
- (3) ISO 15489-1については、対象が第1版ではあるものの小谷允志「記録管理の国際標準 ISO15489」(『情報管理』第48巻第5号、2005年)を参照のこと。
- (4) ここでは、ISO15489-1に変更を加えることなく作成された日本産業規格 X0902-1:2019「情報及びドキュメンテーション—記録管理—第1部：概念及び原則」(<http://kikakurui.com/x0/X0902-1-2019-01.html>)を使用した。同、5頁。
- (5) 同上、2頁。
- (6) 同上、6-7頁。
- (7) Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System: Recommended Practice CCSDS 650.0-M-2*, Magenta Books, June 2012. (<https://public.ccsds.org/pubs/650x0m2.pdf>).
- (8) 同上、p.1-1.

OAIS 参照モデルについては、以下を参照のこと。栗山正光「OAIS 参照モデルと保存メタデータ」(『情報の科学と技術』第 54 巻 9 号、2004)。／BDC ブログ「情報の長期保存に有力な OAIS 参照モデル第 1 回」(国立映画アーカイブ、2016 年 12 月 22 日：https://www.nfaj.go.jp/nfc_bdc_blog/2016/12/22/%E6%83%85%E5%A0%B1%E3%81%AE%E9%95%B7%E6%9C%9F%E4%BF%9D%E5%AD%98%E3%81%AB%E6%9C%89%E5%8A%9B%E3%81%A0ois%E5%8F%82%E7%85%A7%E3%83%A2%E3%83%87%E3%83%AB-%E7%AC%AC%E5%8F%BC%91%E5%9B%9E%E5%8F%BC%88%E5%85%A8/#log_OAIS)。

- (9) Lee, C. A., *Open Archival Information System (OAIS) Reference Model*, Marcua J. Bates, Mary Niles Maack, Ed.; *Encyclopedia of Library and Information Sciences*, Third Edition; Boca Raton, 2015; pp. 4020-4021.

なお、CCSDSが開発の中心的役割を担ったのには、NASAをはじめとする宇宙事業において重要なデジタル記録がフォーマットの進化等により電子情報の読み出しが不可能になる問題への対応でもあった。この点については、次の記事が詳しい。Tampa Bay Times, *"Digital dark age" looms as file formats evolve, disappear*, September 1,

email を歴史資料にするために：OSS・ePADD のススメ

2005. (<https://www.tampabay.com/archive/2003/01/27/digital-dark-age-looms-as-file-formats-evolve-disappear/>).

- (10) 例えば Presevrca 社 (Web サイト：<https://preservica.com/about>) は電子記録の保存サービスを提供しているが、無料版の設定もあるものの、同一アカウントで利用できるのは1名までであり、ストレージの容量も 5GB と個人利用においても厳しい条件となっている。有料版では月額 199 米ドルで 3 人・250GB とアップグレードされる (<https://starter.preservica.com/pricing>) もの、やはり機関等での恒久的利用は難しく、それ以上の容量プランとなると、個別相談対応となっている (<https://preservica.com/digital-archive-software/products-editions>)。機関での利用を想定する場合、それぞれ利用条件は異なるであろうものの、バックアップ等を考慮すると相当な予算が必要となるであろうことは想像に難くない。
- (11) MySQL については、Web サイト (<https://www.mysql.com/jp/>) を参照のこと。
- (12) スタンフォード大学図書館 Web サイト“ePADD”-“About” (<https://library.stanford.edu/projects/epadd/about>)。なお、本稿執筆中にハーバード大学・マンチェスター大学・スタンフォード大学は共同でプロジェクト：Integrating Preservation Functionality into ePADD (ePADD+) を開始し ePADD に email の長期保存機能を搭載するなど、アプリケーションの機能を進展させている (スタンフォード大学図書館 Web サイト：“ePADD” (<https://library.stanford.edu/projects/epadd>) / 同 “EA:BCC (Phase4)” <https://library.stanford.edu/projects/epadd/about/eabcc-phase-4>)。この新機能については、今後検証した上で別稿にてその結果を報告することを予定している。
- (13) 上に示したように、最新版には長期保存の機能が追加されているが、筆者は未検証のため、この点は別稿にゆずる。
- (14) MBOX 形式については、アメリカ議会図書館 Web サイト：“Digital Preservation at the Library of Congress” のうち “MBOX Email Format” (<https://www.loc.gov/preservation/digital/formats/fdd/fdd000383.shtml>) を参照のこと。
- (15) archivematica については Artifactual 社 Web サイト “Wiki-Main Page” (https://wiki.archivematica.org/Main_Page) を参照のこと。
- (16) 例えば Canadian Center for Architecture. CCA Digital Archives Processing Manual (<https://github.com/CCA-Public/digital-archives-manual?fbclid=IwAR0dVNIAodAd9kKmV06VfBkPFO2SPdiMaCMwNShY869Z9Ix8BKkfysM9XK4>)。
- (17) ePADD の最新版 Ver. 8 の導入については、“Getting Started”, Stanford University’s Department of Special Collections & University Archives, *ePADD Installation and*

User Guide, April 2021. pp. 7-17. (https://docs.google.com/document/d/1CVIpWK5FNs5KWVHgvtWTa7u0tZjUrFrBHq6_6ZJVfEA/edit).

- (18) 同前、72-73 頁。
- (19) 日本語化することは可能であるが、厳密には色々な問題がある。第一にプログラムに詳しい人物が、表記を全て日本語に書き換える必要がある。書き換えた後、ePADD がバージョンアップするたびに、その作業は続くこととなる。よって、ePADD を提供しているスタンフォード大学図書館が、多言語化に着手しなければその面倒は永遠に続くこととなる。
- (20) 黒橋禎夫『改訂版 自然言語処理』（一般財団法人放送大学教育振興会、2019 年）9 頁。
- (21) Stanford University's Department of Special Collections & University Archives, op. cit. p. 26.
- (22) 同ファイル最上段には、“A language related util class, created so as to be able to easily expand to other languages by just creating and using a similar dictionary for the language” とある。
- (23) ePADD をインターネットで検索してみると、英語圏以外の web サイトを見つけることが出来なかった。現段階では推察の域を出ないが、本問題の影響が大きいと思われる。

※本稿で扱った ePADD については、入澤寿美学習院大学名誉教授と有志（筆者も参加）が研究成果（入澤道場 Web サイト：<http://irisawadojo.blogspot.com/search/label/ePADD>）を公表している。参照されたい。