

電子個人記録の整理方法： デジタル・フォレンジックスによる原秩序の検証

橋本 陽†

はじめに

デジタル・トランスフォーメーション (Digital Transformation、DX) という語を頻繁に目にするようになって久しい。組織における業務をはじめとする人間の活動の中に、発展した情報技術を取り入れ、その効率化を目指すDXが、日本国内全体で進められつつある。DXの障害の一つに挙げられるのが、紙文書である。一般的に、紙文書の電子化も、DX事業の一つに数えられる。ここでいわれる電子化には、すでに存在する紙文書の電子化と電子機器によって作成されるボーン・デジタル (Born-digital) の文書をデジタルの状態のまま保管し利用する状態が含まれる。非効率性が問題視される紙文書であるが、現代社会において作成されるものの大半は、電子機器によって作成されたボーン・デジタルの文書が印刷されたものである。その観点から見れば、文書作成はすでに電子化しているともいえる。

アーカイブズ学が進んだ欧米では、いち早く電磁的に作成された文書を証拠能力のある記録として長期的に保存する方法が議論されてきた。ライフサイクルの視点からは、記録の作成段階、維持管理の段階、保存の段階のそれぞれで機能する電子システムを構築し、各システムを連動させて保管の連鎖を実現することで、信用価値が備わった記録が永続的に保存できると主張されている。つ

まり、現用段階において、文書決裁のワークフローを管理するシステムから、そこで作成された一連の文書を正確性、信頼性および真正性を備えた記録として維持管理するシステムに移送する。最後の非現用段階において、永続的な保管価値のある記録資料群をアーカイブズが管理する保存システムに移管させ、真正性を保証しながら外部からの利用に対応していくのである⁽¹⁾。また、レコーズ・コンティニューム (Records Continuum) の観点からも保管の連鎖は同様のシステムで実装できる。次元1の作成では文書決裁システムを使用し、次元2の捕捉で維持管理システムに文書を記録として取り込み、次元3の組織化も維持管理システム内の機能で実行する。最後に、次元4の多元化の機能は、アーカイブズにおける保存システムに記録資料群を社会全体の記憶として永久的に保管していくことで遂行できる。このような保管の連鎖を理想的に実現するための諸システムの構築は、その整備ができるだけの能力を備えた組織以外には難しい。ましてや、組織の業務から離れた私生活の場面で作成される文書や記録について、個々人が保管の連鎖を完備したシステムをわざわざこしらえるとは考えられない。しかし、現代社会では、個々人の私生活においても電子機器によって文書を作成しない方が例外的である。したがって、歴史的な価値を持つ個人の電子記録を将来に向けて

† 京都大学大学文書館助教

残していくには、組織が実行できるような保管の連鎖とは異なる手段によって、アーキビストが何らかの対策を講じなければならない。

本稿は、そのような性質を持つ個人の電子記録を、どのようにアーキビストが保存を見据えながら整理し公開していくかについて論じる。そもそも電子に限らず個人記録自体が、アーカイブズ学の先進国である欧米においても、組織の記録と比較して研究が遅れてきたと評されている。ところが、1990年代以降、研究が活発化し、電子記録においても議論が及び、保存や整理の実践が行われるようになった⁽²⁾。一方、日本においては、組織における電子記録の保管方法についてすら、学術研究およびそれを踏まえた実務が十分に進められているといえる現状ではない⁽³⁾が、個人の電子記録となれば、さらに成果は限定される。アメリカ議会図書館と国立国会図書館憲政資料室における電子個人記録の収蔵状況を比較した塩崎亮によれば、憲政資料室が受け入れたものは皆無であった⁽⁴⁾。つまり、個人記録を扱う中で日本を代表するような機関であってすら、電子記録を保有していないという事実から、対策と議論が立ち遅れた現状が窺い知れるのである。しかし、今後は、日本のアーカイブズにおいても、個人から寄贈される記録資料群の中に、パソコンなど個人の所有する電子機器に残された記録が含まれるのは避けられない。フロッピーディスクといった旧式の保存媒体にいたっては、すでに受け入れられているものもあるだろう。このことから、現在の時点において、電子個人記録を適切に処理するためのワークフローを整理しておくことは有用であると考えられる。

本稿では、最初に、使用する用語を説明した後で、電子個人記録の整理に関する先行研究を概観する。続いて、記録資料群を整理するにあたっては、出所および原秩序尊重の原則に準拠するのが鉄則であるが、電子個人記録への適用についてはどうあるべきかを検討する。特に原秩序に関しては、文

書形式学 (Diplomatics)⁽⁵⁾ との類似性が見出されるデジタル・フォレンジックス (Digital Forensics, DF) の機能に着目し、その関係性を読み取るための具体的な運用方法を確認する。さらには、これを電子データ長期保存のOAIS参照モデルと結びつけながら、公開していくための方法と留意点について考察する。

1 使用する用語と先行研究の確認

1.1 用語

用語については、まず電子記録保存の国際的プロジェクトであるインターパレス・プロジェクト (InterPARES Project、以下インターパレス) の定義を確認する。文書 (Document) とは、「安定した文体で媒体に付与され (書き込まれ) たメッセージが構成する情報の不可分の単位」であり、「定まった形式と安定した内容」を持つ⁽⁶⁾。記録 (Record または Archival Document) とは、「実際の活動の中で、その活動の手段または副産物として作成あるいは受領され、行動や参照のために取り置かれた文書」⁽⁷⁾ を指す。取り置くという行為は、ある記録を同じ記録の集合体の中に位置づける行為であり、このときにアーカイブ結合性 (Archival Bond) が生じる。アーカイブ結合性とは、同一の記録資料群に属する個々の記録の関係性を指す⁽⁸⁾。したがって、記録とは、文書の中の一つの形態であり、アーカイブ結合性を備えたものを指す。

Encyclopedia of Archival Science は、個人記録を、「個人または家族がその活動の中で作成した文書」であり、「その出所は個人または一つの家族」とであると定めている。具体例をあげれば、書簡、業務や私生活の連絡物、日記、予定表、写真、草稿などである⁽⁹⁾。個人においても、文書ではなく記録とするには、他の記録との関係性の中に取り置き、アーカイブ結合性を得る必要がある。この取り置くという行為を厳格に捉える場合、紙であれ電子であれ、業務分類に応じた正式なファイリ

ングの体系を準備し、その中の該当するファイル内に保管するという作業に該当する⁽¹⁰⁾。しかし、ヴィクトリア・レミューによれば、取り置くというのは、個々人がコンピューター上のフォルダーに文書を一時保存する⁽¹¹⁾行為にまで広げられるという⁽¹²⁾。したがって、個々人が何らかの基準で自らが作成あるいは受領した諸々の文書をまとめていけば、それらはアーカイブ結合性でつながれた個人記録であると判断できる。そういった個人記録が電子生成されたものを本稿では電子個人記録と呼称する。

最後に、整理という言葉については、アメリカ・アーキビスト協会 (Society of American Archivists, SAA) の定義する Processing の意味で使用する。整理とは、「アーカイブズ資料を利用に向けて準備する」ことを指し、その中に編成 (Arrangement) と記述 (Description) が含まれる⁽¹³⁾。

1.2 先行研究

電子個人記録の整理に関する日本国内の研究においては、塩崎による研究が注目される。塩崎は、イギリスのパラダイム (Personal Archives Accessible in Digital Media, Paradigm)、アメリカのエイムス (Born-Digital Collection: An Inter-Institutional Model for Stewardship, AIMS)、全米人文科学基金 (The National Endowment for the Humanities, NEH) によるプロジェクトにおける取り組みを簡潔にまとめ紹介している。NEHの助成プロジェクトには、サルマン・ラシュディーが寄贈したパソコン内の状況をエミュレーターによって再現した有名な整理方法も含まれる⁽¹⁴⁾。また、塩崎は、「個人による情報管理 (Personal Information Management, PIM)」というアーカイブズ学以外の専門領域が、英語圏のアーカイブズ学においてどのように導入されつつあるかについて、その研究成果を紹介している⁽¹⁵⁾。PIMとは、個々人が紙や電子を含む自らの文書、ウェブサイ

トおよび電子メールなどをどのように維持管理しているのかを研究する分野である。認知心理学、ヒューマン・コンピューター・インタラクション (Human Computer Interaction)、データベース管理など多様な研究分野から影響を受けている⁽¹⁶⁾。電子個人記録の整理に関わっては、アーカイブズ受け入れ時に秩序のない事例が数多い個人記録の分析に、PIMの視点が役立つのではないかと主張する研究が見受けられる⁽¹⁷⁾。

塩崎が触れたイギリスのパラダイムは、電子個人記録の整理を試みた初期のプロジェクトである。このプロジェクトは、電子データ全般のための長期保存の枠組みである OAIS 参照モデル (The Reference Model for Open Archival Information System) に基づき、受け入れから公開にいたるまでの筋道を説明している⁽¹⁸⁾。もう一つのエイムスもまた、OAIS 参照モデルを基礎とした整理過程について、多くのアプリケーションや事例を参照しながら、言及する。パラダイムより新しいマニュアルであることを反映し、紹介される技術も新しい。その中には DF も含まれる⁽¹⁹⁾。

電子個人記録ではないが、具体的な整理過程を説明した成果の中で注目されるのは、カナダ建築センター (Canadian Centre for Architecture, CCA) の試みである。CCA は、DF のアプリケーションであるビットキュレーター (BitCurator)⁽²⁰⁾ の機能を使い、受け入れた電子記録の資料群を検証し、そこから保存対象となるものを OAIS 参照モデル準拠のアプリケーションであるアーカイブマティカ (Archivematica)⁽²¹⁾ によって長期保存を実行するというワークフローを実際に運用している。電子メールについても、アーカイブズが適切に処理できるように開発されたイーパッド (ePADD)⁽²²⁾ を使用し、整理の実践に役立っている⁽²³⁾。CCA の試みは建築記録を対象としているが、電子個人記録の整理についても参考となる点が多々認められる。

電子個人記録の整理についての先行研究の中で共通しているのは、OAIS参照モデルに準拠した長期保存の方針を立てるとともに、伝統的な整理の方法論である出所尊重と原秩序尊重の原則を遵守するよう求めている点である。OAIS参照モデルの準拠は、アーカイブズで保存する電子記録の可読性と信用価値を維持するための前提となる。そのためのワークフローは日本でも橋本陽が明らかにしている⁽²⁴⁾。出所尊重と原秩序尊重の原則についても、記録の置かれたコンテキストを伝えることで、信用価値を守る役割を担っており、当然重視されなければならない。しかし、これら整理の原則は、物理的な実体のない電子記録を扱うにあたり、どのように適用を試みるかが国際的な議論となっている。さらに、電子記録をはじめとする現代の問題をアーカイブズ学が取り上げる中で、出所と原秩序という概念自体が変わりつつある。

上記の先行研究の中では、このような変化の中にある概念において、どのようなものを採用し、電子個人記録の整理に活かしていくかが明示されていない。この問題について、ここではアーカイブズ学の原点ともいべき原則を簡潔に見直しながら検討することとする。そこで得られる観点から、整理の過程にDFをどう組み込むのかを考察する。

2 電子個人記録と整理の原則

整理の原則はコンテキストを保持する役割を果たすと書いたが、コンテキストのどの部分に関与するのだろうか。この点について、ジェニファー・ミーハンが要点をまとめている。出所はフォンド外部の構造であり、記録資料群とその作成者や、それを生み出す機能・活動との関連性を表す。出所尊重は、出所となる作成者、その使命、構造、機能を指す出所コンテキスト (Provenancial Context)、作成者が属する法規・行政コンテキスト (Juridical-Administrative Context) に関わる原則である。原秩序は、フォンド内部の構造であり、

内部に含まれる各記録の関連性を表す。原秩序の尊重は、記録が生成される過程の業務手続きである手続きコンテキスト (Procedural Context) とフォンドの構造を示す記録コンテキスト (Documentary Context) に関わる原則となる。ソフトウェアの使用など記録を作り出すのにどのような技術が用いられたかを指す技術コンテキスト (Technological Context) は、出所と原秩序の双方に関与している⁽²⁵⁾。

出所は、伝統的には、具体的な記録の集まりであるフォンドを作成する一つの組織体、集団、一人の個人であり、フォンドと1対1の関係にあると捉えられてきた。ところが、現代の組織においては内部機構の変遷が目まぐるしく、シリーズ単位で顕著に見られるように、記録資料群の生成に複数の部署が関わるようになった。また、部署が遂行する諸活動の副産物や手段が記録であることから、諸活動が構成する機能も出所の一部であるという見方もアーカイブズ学の基礎的な知識となっている。さらに、その次の段階として、記録で描かれる主体についても、それがなければ記録は生まれなかったという発想によって、並行出所 (Parallel Provenance) や社会的出所 (Societal Provenance) という考えも主張されるようになった。複雑化する現代社会に呼応して、出所の概念もまた複雑になっているのである⁽²⁶⁾。

複雑化する出所の尊重は、アーカイブ記述によって実行できる。シリーズ・システムに代表されるように、出所とそれに関わる記録資料群の記述を別々に作成し、それぞれの記述を紐付けることで複数の出所と記録資料群の複雑な関係性を表現できる。この紐付けの中に、機能や活動の記述を加える試みもすで実践されている⁽²⁷⁾。さらに、場所および記録の扱う主題や主体など、そしてそれらと記録の間にある多様な関係性を追加して表現するために、新しい国際記述標準である Records in Contexts (RiC) が作成され、改訂と各地にお

ける実践が進められている⁽²⁸⁾。

電子個人記録の出所の尊重についても、シリーズ・システム型のアーカイブ記述が有効である。完全に一人の個人が私有していたパソコンのような事例であれば、資料群作成者の記述一つと、パソコン内にある資料群それぞれの記述を紐付ければよい。ほかには、家族などのグループが共有して使用した機器が寄贈された場合、その構成員が複数関わったシリーズやファイルも含まれるだろう。その場合は、構成員それぞれの記述と資料の記述を紐付けることで、資料と出所の関係性を表現できる。また、個人の場合、仕事と私生活の境目が曖昧な活動から生じる記録資料群もある。こちらについても、資料群の記述と複数の活動の記述の紐付けで対応可能である⁽²⁹⁾。

出所と同様、原秩序についても概念は複雑化している。原秩序は、記録資料群の作成者が資料群内部に構築した秩序であると一般的に認識されてきた。しかし、現代になって、電子記録が構成するフォンドについては、その構造の支柱をなすシリーズやシリーズに含まれるファイルといった物理的な近接性は意味をなさなくなったことも一因となり、原秩序もまた再考されるようになった。アーカイブズ移管時に見出されるような記録間にある単一の関係性ではなく、現用期における変更の跡も含む多層に折り重なった論理的な関係性を認識、想像または再生することが原秩序の尊重であると主張する専門家も見られるようになった⁽³⁰⁾。

そもそも、紙で作成されたものであっても、個人記録の資料群は混沌としていることが多く、原秩序尊重の原則の準拠が難しい。目の前にある資料群に秩序を見つけられないことから、原秩序を抽象的に把握しようとする流れが生まれる。例えば、ミーハン⁽³¹⁾は、原秩序尊重を、コンテキストを分析するための抽象的な概念の枠組みだと認識し、記録資料群とその作成者との関係性などを根拠にして、アーキビストが記録の配列を決定するべきだ

と主張している⁽³¹⁾。

電子個人記録についても同様の議論が起こっている。ジンファン・ニウは、電子記録全体において原秩序がどうあるべきかを扱う中で、抽象的概念秩序、アプリケーション・ソフトウェアのユーザー・インターフェース上の秩序、ストレージ上の秩序の3つに区分している。抽象的概念秩序は、パソコンや外部ストレージなどいくつかの媒体に分かれていたとしても、記録資料群作成者の組織構成や活動などのコンテキストを反映して、分散した保存媒体を貫いて存在する抽象的な原秩序を指す。特に、組織のような正式の管理を受けない個人の記録に関しては、抽象的概念秩序はその個人の頭の中にしかならない可能性があり、明確に示されていないことがあるとの見解をニウは示している⁽³²⁾。同じく、電子記録の原秩序を扱ったジェーン・ジャンは、事例研究によって、整理の現場でどうそれが認識されているのかを明らかにした。電子個人記録についてはウェルカム図書館(Wellcome Library)の事例を取り上げている。そこでは、個人の残したフォルダーの構造を残しながらも、最終的には、寄贈資料群に含まれる紙媒体と電子媒体の両方を貫いた論理的な構造に編成することが方針とされている。また、ジャンは、自分が取り上げた事例からいくつかの結論を導き出した。その一つは、各記録には他の記録との無数の関係性があるとみなされながらも、現状では紙の時代と同じような一つの階層構造に落とし込む編成が行われているというものである。さらには、その階層構造が原秩序とされ、それがもともとの作成と使用の証拠として記録を閲覧する手段ともなっていると説明している⁽³³⁾。ジャンの論考からは、アーキビストの実践においても、ニウのいう抽象的概念秩序を唯一の原秩序として捉え、その復元を目指す傾向があるのがわかる。この抽象的概念秩序は、現実に認識できるユーザー・インターフェース上の秩序や物理的なストレージ上の秩序

を超えたところに想定される配列であり、ミーハンの主張する抽象的概念枠組みとしての原秩序と同様の概念であると判断できる。

日本国内においては、安藤正人の主張にも類似した点が見出される。安藤は、記録資料群作成者が決定した記録の位置関係であり、かつ明示されているものを原秩序と理解した。そして、原秩序尊重の目的は、原秩序の中に潜む体系的構造を明らかにすることだと述べている⁽³⁴⁾。この体系的構造とは、「記録史料群が内包している、記録発生源母体の組織と機能を反映した」⁽³⁵⁾配列を指す。さらに、これは、記録資料群作成者が現用段階において行った資料群の配列の変更までも含む概念であり、その経過を示す根拠となる資料があれば、「より立体的に豊かに解明される」と説明している⁽³⁶⁾。つまり、体系的構造とは、目にできる原秩序とは異なる別の秩序であり、物理的な配列状況と同一視できるような一意の構造ではない。この点で、具体的な実体を離れた抽象的な構造を想定したニウやミーハンの概念的秩序と近似したものとみなされる。また、秩序の変遷過程を含めているところも、英語圏における原秩序をめぐる議論と共通する要素として認められる。

ここで疑問に挙がるのが、次の2点である。

1. 抽象的概念秩序や体系的構造の存在を確定する方法
2. 秩序変遷の過程を確定する方法

これらの疑問を、電子個人記録において解明できるのかを問う必要がある。疑問1で問われるのは、電子記録の資料群に残された配列とは異なる秩序や構造を見出す方法となる。疑問2は、アプリケーション・ソフトウェアのユーザー・インターフェース上において電子ファイルを移動させた経過など、個人が行う電子ファイルの操作の歴史を逐一追うことができるのかという問いとなる。

疑問1に関する問題として、抽象的概念秩序や体系的構造について記録資料群作成者による配列と認識できる根拠がないという点が挙げられる。例えば、ニウは電子個人記録の抽象的秩序はその作成者の頭の中にしかないと説明していたが、当事者でないアーキビストがそれを知る手立ては示されない。ジャンの事例研究においても、抽象的概念秩序はむしろ整理担当のアーキビストが、記録資料群作成者、その機能および記録そのものの関係を調べ、想定していたものである。ミーハンのいう抽象的な概念の枠組みとしての原秩序もまた、アーキビストが決定する配列であった。また、体系的構造については、その要件の一つに機能が挙げられていた一方で、ヨハネス・パブリッツがまとめたヨーロッパの記録管理史⁽³⁷⁾から明らかのように、作成者の作るファイリングの体系が必ずしも機能を基礎としてきたわけではない。つまり、記録資料群が機能を反映した秩序を必ず内包しているとはいえないのである。さらには、組織と違い一貫した管理体系が与えられない個人記録においては、原秩序すら残っていない事例が多く、原秩序すら認識できない資料群から、その内部にある体系的構造まで見出すのはほとんど不可能に近い。もちろん、そのような記録資料群について、機能を考慮しながら編成を行うというのは当然の措置である。しかし、それはもはやアーキビストの仕事である。したがって、電子個人記録において、原秩序と別の配列は、もはや当の記録資料群作成者が与えたものではなく、アーキビストが構想した構造であり、人為的秩序と呼ぶのが適切だろう⁽³⁸⁾。また、仮にPIMによって、個人の情報の管理方法にいくつかのパターンがあることがわかったとしても、そのパターン自体が抽象化されたものであるのに加え、どのパターンを整理対象の記録資料群にあてはめるかは、アーキビストの判断に任せられる。そのため、PIMの知見を活かして電子個人記録を構造化するにしても、その構造を原秩序と

するのではなく、アーキビストが設定した人為的秩序とみなすのが穏当である。

なお、原秩序のない記録資料群に人為的な秩序を作るのは、入門書でも紹介されるアーキビストの通常業務である。オーストラリアのアーキビストの手引きである *Keeping Archives* の第3版も、「人為的シリーズ」の必要性について言及している。そこでは、人為的シリーズは、同一の機能、類似した内容あるいは特定の形態から構成できると述べられている⁽³⁹⁾。人為的シリーズの設定には、機能、内容および形態を調べるという行為が当然必要となるが、これは抽象的概念秩序や体系的構造を推論するとき求められる手続きと同じである。このように整理の実務の面においても、人為的秩序は、抽象的概念秩序や体系的構造と違いがない。

疑問2については、まず原秩序とはどの段階の秩序を指すのかを検討したい。インターパレスでは、アーカイブズが受け取る前の現用期の最後の秩序であると定義している⁽⁴⁰⁾。これは、ヨーロッパのいくつかの地域においても採用される定義である。イタリアについては、エリオ・ロドリニが、記録資料群が受けた管理体系の変遷も維持する必要があるという考えから、原秩序を現用期の最後の秩序とすると主張している⁽⁴¹⁾。ドイツのアーカイブズ学の泰斗であったパブリッツもまた、最初の秩序と二次的な秩序を比較し、最初の方の構造を侵害していなければ、二次的なものを受け入れることが経済的にも合理的であると説明している⁽⁴²⁾。オランダにおいても、もともとは原秩序を最終的な配列と捉えていたようである⁽⁴³⁾。電子記録の原秩序においては、ヘザー・マクニールはコンピューター内のフォルダーの構造にも該当させられると述べた⁽⁴⁴⁾。この考えに則れば、個人から受け取った時点での電子機器内にある秩序を原秩序と捉えることができる。また、先述したように、個々人がコンピューター上のフォルダーに文書を一時保存することも、記録を成立させるための置き

に該当し、そのフォルダー内にあるそれぞれの記録にアーカイブ結合性が認められる。置き置き以降の記録管理の様態によっては、フォルダー構造は変遷していくだろうが、最終秩序としての原秩序内にアーカイブ結合性の認められるフォルダーも残ることを考えれば、古典的な原秩序の尊重の意義は電子個人記録においても決して失われていない。なお、デイヴィッド・ベアマンは、記録間の関係性を記述したメタデータの必要性を主張した⁽⁴⁵⁾が、それはアーカイブズが受け入れた時点での構造をメタデータとして保存しておくことで対応できる。

インターパレスでも原秩序を現用期の最後の秩序であると定義したが、その秩序にいたる変遷過程の重要性は認識されている⁽⁴⁶⁾。ただし、この変遷過程をたどるのは、近代のレジストリー・システムに代表されるような成熟した記録管理制度下で保管されるもの以外の記録資料群においては難しい。変遷過程を知るのに求められるのは、ライフサイクルの作成・維持管理・永久保存の各段階のうち、現用期の記録管理を担当する維持管理の段階の機能である。この維持管理の段階で、組織変遷を受けてシリーズを使用する部署が変更される、シリーズが別のシリーズにつけかえられるなどの変化が記録されていなければならないのである。電子記録の維持管理システムでは、これらの変化をメタデータとして残しておくことが真正性保持の条件ともされている⁽⁴⁷⁾。このような管理を受けない電子個人記録において、それが形成する構造の変化の過程を追えるかどうかは、後段、言及する。

これまで述べてきた電子個人記録に対する出所および原秩序尊重の原則の適用方法についてまとめれば、出所と記録の関係性については、シリーズ・システム型の記述で表現できる。一方、原秩序は、アーカイブズが受け入れる直前に当たる現用期における最終時点の構造とし、それ以外の抽

象的な構造概念は、人為的秩序とみなし区別することができる。ここで考えなければならないのは、電子個人記録が形成する資料群の原秩序をどのように認定し記録していくのかという問題である。これについては英語圏においてDFを用いたアプローチが考案されている。

3 デジタル・フォレンジックス (DF) による原秩序の認定

DFとは、古賀崇によれば、「データ修復などの手段を通じ、デジタル媒体（ハードディスク、サーバなど）上の証拠保全を行うことを意味する」⁽⁴⁸⁾。主に司法の領域で使用され、電子データや文書の証拠能力を評価した上でその保全が図られる。古賀が指摘したように、記録の正確性、信頼性、真正性から構成される信用価値を評価する文書形式学との接合がインターパレスの研究者たちによって試行されるとともに、北米においては電子個人記録の整理に使用されている⁽⁴⁹⁾。

インターパレスにおいて、DFと文書形式学の接合が意図された理由を、次のようにロジャースがまとめている。

電子文書形式学とデジタル・フォレンジックスは理論と方法論の面で多くの類似性があり、重要な課題を共有している。そのような課題の一つは、真正性を評価することと、電子記録（文書形式学）または電子証拠物（デジタル・フォレンジックス）のコンテキスト、出所、関係性および意味を確立することである。文書形式学の原則とデジタル・フォレンジックスのツールを組み合わせることで、電子環境における記録を確認し分析するという問題に対処できる⁽⁵⁰⁾。

ここでいうDFのツールとは、アプリケーションやソフトウェアを指す。ロジャースは、つまり、

文書形式学の原則をDFのアプリケーションで実行できると主張しているのである。それにより、それぞれの記録の間にある関係性も読み取れるというところに注目すれば、アーカイブ結合性や原秩序の解析に対しても有効であることが期待される。

現在、アーカイブズで使用されるアプリケーションの中で、DFの技術が採用されたものの代表格が、ビットキュレーターとアーカイブマティカである。ビットキュレーターは電子個人記録の分析のために開発された。アーカイブマティカは、電子記録全般をアーカイブズで長期保存するためにOAIS参照モデルに準拠して開発されたアプリケーションであり、一部DFのツールが用いられている。ここからは、電子個人記録を対象にDFが具体的にどのように働くのかを見るため、ビットキュレーターの機能を簡単に紹介する。なお、検討したバージョンは2.0.12である。

ビットキュレーターが進める作業の流れは、ディスク・イメージの作成、復元を含むファイルとファイル・システムの分析、メタデータの抽出、機微情報の発見、複製情報の発見と除去となる。

ディスク・イメージとは、フロッピーディスク、DVD、USBメモリあるいはパソコンのハードディスクなどのストレージ媒体にあるデータを、ファイル・システムの構造や制御情報などとあわせてすべて複製し、1つのファイルにしたものである。ビットキュレーターによる処理の最初の手順として、整理対象となる電子記録の入った機器自体は操作できない設定にした上で、ウイルスチェックを行い、機器内のディスク・イメージをビットキュレーターの入ったコンピューター上に作成する。その後は、電子機器との接続を解除し、ディスク・イメージを対象として処理が進められる。

続いて、ファイルとファイル・システムの分析が行われる。ここでいうファイルとは、ストレージで保管されるデータを指す。アーカイブズの階

層構造の名称であるファイルと区別するため、電子ファイルと表記する。ファイル・システムとは、OSの中で、電子ファイルの操作や管理を行う機能である。それにより、ユーザー・インターフェース上で、フォルダーといったディレクトリを作り階層構造を設定できる。ビットキュレーターでは、この電子ファイルとそれが集積して作り上げる構造の解析などを行える。電子ファイルの単位では、そのフォーマットの認定、一意のIDの付与、消去された電子ファイルの復元などが実行される。ただし、この復元に関しては、電子ファイルまでは実現できず、かつて存在したことを示すメタデータのみに限られる場合もある。また、ファイル・システムの分析によって、ストレージ内のディレクトリ構造が読み取られる。分析の結果は、グラフやXML形式のメタデータとして出力できる

分析作業の中で、機微情報となる可能性のある情報も抽出される。メールアドレス、アメリカの社会保障番号などである。また、各電子ファイルのハッシュ値の計算によって、複製された電子ファイルを発見し除去することも可能である。

さらに、読み取ったディレクトリと電子ファイルの中で、別のシステムに移送するものを選択できる。CCAでは、長期保存する電子記録をアーカイブマティカに移送するためのアプリケーションを用意し、この機能を利用している⁽⁵¹⁾。

ビットキュレーターの持つ機能において原秩序と関わるのは、ファイル・システムの分析である。これにより、受け入れた機器内で電子ファイルがどのような構造の中に位置付けられていたかがわかる。ここで考慮されなければならないのは、すべての電子ファイルが、電子記録ではないという点である。電子記録であるには、少なくとも、作成者自身が構築したファイル内に取り置かれている必要性があり、それ以外は電子文書として認識するのが正確だろう。このほか、個人の所有する電子書籍や雑誌なども機器内に含まれていること

が十分想定される。電子ファイルの中で、アーキビストが扱う対象となるのは、書籍や雑誌ではなく、記録あるいは記録として認められる可能性のある文書であり、こちらの構造に注目する必要がある。この構造の中にも、記録資料群作成者が管理してきたと認められる原秩序と、文書が雑然とした原無秩序に区別される。各電子記録の間にあるアーカイブ結合性は、原秩序のあるグループの方で見出すことができる。さらに、ビットキュレーターのメタデータ作成機能により、コンピューター上に存在する原秩序と原無秩序の状況をXMLファイルに書き出し、保存していくことが可能である。

なお、ビットキュレーターの持つ複製された電子ファイルの消去機能は、アーカイブ結合性を破壊する可能性がある。完全に同じ文書であっても、異なるファイルに取り置かれた場合、そこには異なるアーカイブ結合性が存在するため、それぞれの文書は異なる記録であると認識される⁽⁵²⁾。この点を無視して、複製された電子ファイルを消去すれば、該当する記録だけでなく、それとアーカイブ結合性でつながれた他の記録の特性も消してしまうこととなる。

また、ビットキュレーターのバージョン2.0.12を検討した限りでは、原秩序にいたるまでの記録資料群に与えられた構造の変遷過程は追いきれないようである。ただし、消去した電子ファイルは、かつてあったディレクトリの中で少なくともメタデータだけでも復元できる。そのため、記録資料群作成者が、何を残そうとし、何を消去したかというところにかぎっては、資料群の変遷過程をたどることが可能である。

このように、文書と記録のフォーマットや、原秩序をはじめとする関係性を読み取ることができると、DFは文書形式学を電子個人記録に適用するための手段となりうる事が確認された。その後の整理過程については、電子記録の真正性を維持したまま長期保存を行うため、OAIS参照モ

デルに準拠した編成と記述を進めることとなる。

4 原秩序認定後の整理過程

OAIS 参照モデルに準拠した整理方法については、ここで言及したビットキュレーターにアーカイブマティカを組み合わせたCCAの事例が参考になる。CCAでは、ディスク・イメージを保管し、受け入れ時の状況を保存し、それとは別に編成を行い、シリーズを設定している⁽⁵³⁾。それにより、受け入れ時の原秩序・原無秩序から人為的秩序へと変化した過程を認知できるようになっている。

このような構造の変更をわかるようにしておくことは、電子データの長期保存には不可欠の要件となっている。OAIS参照モデルでは、長期保存のためのメタデータである保存記述情報(Preservation Description Information、PDI)を構成する要素の一つに、出所情報があり、そこに電子ファイルに起こした変化について書き記すように定められている⁽⁵⁴⁾。文書形式学およびアーカイブズ学においても、このようなメタデータは、真正性の一要素であり、不当な変更が行われていないことを示す完全性の証明に必要なものとして提示されている⁽⁵⁵⁾。なお、アーカイブマティカには、受け入れ時の配列を別のものに変更する編成機能があり、変更過程の情報もメタデータに残すことができる⁽⁵⁶⁾。

受け入れ時の原秩序・原無秩序を残したまま、公開するという考えもあるかもしれないが、個人記録全般において、混沌としている場合が多く、保存機器も多岐にわたることも想定すれば、一貫した人為的秩序をアーキビストが提示した方が、利便性が高まる。さらに、編成時には、記録資料群を作成した個人の情報や機能を調査するため、それらと関連づけた構造を作った方がアーキビストの読み取ったコンテキストも伝えやすい。一方で、最近は自然言語処理の技術を用いたトピックモデルによって、自動的にテーマなどを解析できるよ

うになっており、アーカイブズでも採用されるようになってきている⁽⁵⁷⁾。しかし、現状では、アーカイブズの記述標準を利用して実現できるような横断検索にまで対応していないため、やはり、フォンド-シリーズ-ファイルといった伝統的な階層構造にしたがって編成し、それに応じた記述を作成することの利点は大きい。

階層構築の方法については、加藤聖文が提示した「可変的なシリーズ設定」⁽⁵⁸⁾の考えに従い、アーキビストが、かけられる労力と時間を合理的に判断しながら、人為的秩序を設定すれば十分だろう。受け入れ時の原秩序はディスク・イメージに残されているのに加え、メタデータにも記載できるのだから、整理する記録資料群の性質に応じて自由に編成を行うことができる。

アーカイブマティカを使うならば、このような人為的秩序の設定を行いながら、電子個人記録を長期保存用のアーカイブ情報パッケージ(Archival Information Package、AIP)にまとめられる。一般利用には、配布情報パッケージ(Dissemination Information Package、DIP)を用いる。DIPにおいては、メールアドレス、社会保障番号に代表されるIDなどの機微情報、寄贈者の依頼による秘匿事項などを隠した状態にしておく必要がある。

公開に際しては、利用者が調べる検索システム内のアーカイブ記述において、資料受け入れから、AIPとDIP生成にいたるまでのデータ操作の過程を詳細に記載しておくことも求められる。そのような記述によって、利用者に、不当な改変のない真正な記録資料群を提供していることが伝えられる。

以上の整理を実行するにあたっては、生存する個人から電子記録の寄贈を受けることが極めて重要となる。DFによって、個人が削除した記録も復元されるし、機微情報もまとめて抽出される。それらの取り扱い、作成者となる個人と入念に相談する必要がある。さらに、電子記録をウェブ

上で公開すれば、複製、加工、流通が紙のものよりも、はるかに容易であるため、著作権も十分に考慮されなければならない。特に問題となるのは、著作者人格権であり、例えば一般公開されなかった記録や文書の著作者に認められる公表権が主張された場合、長期にわたって公開が不可能となる⁽⁵⁹⁾。著作者人格権が行使されないための条件についても、寄贈者とアーカイブズの間で十分に協議しておくに越したことはない。

おわりに

伝統的な整理の方法論である出所尊重の原則と原秩序尊重の原則には、記録資料群のコンテキストを保護し伝える機能がある。一方で、出所と原秩序の概念は、時代が下るにつれ、複雑化していく。電子記録の時代には、それに合わせた原則の使い方が求められるし、個人記録については、その特徴を踏まえての対応が必要となる。したがって、電子個人記録には、電子と個人両方の特質について考慮した上で、原則を遵守するための整理方法を用意しなければならない。

出所尊重の原則については、アーカイブ記述で対処できることが確認された。特に、複数の主体や機能が関与している場合は、シリーズ・システム型の記述が有効である。一方で、原秩序尊重の原則については、組織のような正式な維持管理の規則による統制を受けていない個人記録に対して、単純な適用は難しい。ここでは、原秩序を現用段階の最終秩序とする伝統的な考えを採用し、電子個人記録の原秩序は個人が残した最終的な配列であるとみなした。しかし、個人記録としての性質から、原秩序以外の混沌とした状態にある部分があることは避けられず、それに対してはアーキビストが行う編成と記述による介入が求められる。この介入によって構築される秩序を人為的秩序とし、それにより、利用者にアーキビストが理解したコンテキストを伝えるとともに資料の利便性を

高められる。

電子個人記録が構成する原秩序の調査には、DFが有効である。DFは記録の性質や記録同士の間にある関係性を調査する文書形式学を電子記録において実行する手段となる。その有効性は、英語圏ですでに運用されているビットキュレーターによって、確認された。一方で、原秩序にいたる記録資料群の持つ構造の変遷過程は、ビットキュレーターでは追いきれなかった。ビットキュレーターで状況を確認して以降は、OAIS参照モデルに準拠し、電子個人記録が持つ電子データとしての側面に配慮しながらの整理方法が求められる。これについては、アーカイブマティカのようなアプリケーションを使えば、人為的秩序への変遷過程をメタデータとして残すことも含め、対応可能である。

今後の課題としては、ビットキュレーターを日本語の資料に対応できるように改変することが挙げられる。日本語の形態素解析のほか、アメリカの社会保障番号とは異なるマイナンバーのパターン認証の機能などを付与しなければ、日本の環境で使用できるものにならない。これ以外にも、今回扱わなかった電子メールも、現代社会においては一個人が膨大な量を抱え込んでおり、その整理はアーキビストにとって大きな問題となる。日本でもイーパッドを利用した対策が論じられている⁽⁶⁰⁾が、実用にはいたっていない。さらに電子メールの著作者は、アーカイブズに寄贈される電子個人記録資料群の作成者とは異なるため、著作権の処理がさらに難しくなる。この処理を的確に行えなければ、プライバシー権など機微情報を含まないメールであったとして、長期にわたって非公開とせざるを得ないかもしれない。

電子個人記録を扱う上での諸々の問題について、特に収集アーカイブズとしての側面を持つ各機関において、遅かれ早かれ解決を迫られることになる。実際に資料を受け入れるまでの間に、適切な対策

を練っておくことが、未来に現在を伝えるアーカイブズの役割を果たす上で重要となることは間違いない。

[註]

以下の注で示されるURLの参照日はすべて、2023年1月11日である。

- (1) 橋本陽「電子記録研究の現在：2004年以降の回顧と展望」、『アーカイブズ学研究』第36号、2022年、26-43頁。
- (2) Catherine Hobbs, 'Personal Records', Luciana Duranti and Patricia C Franks (eds.), *Encyclopedia of Archival Science*, Lanham: Rowman & Littlefield, 2015, pp. 266-270.
- (3) 橋本「電子記録研究の現在」。
- (4) 塩崎亮「ポーンデジタル個人文書の所蔵・受入状況：国立国会図書館・米国議会図書館目録の内容分析」、『図書館界』第72巻第4号、2020年、172-182頁。
- (5) 岡崎敦「文書形式学」、高山博・池上俊一編『西洋中世学入門』、東京大学出版会、2005年、59-72頁。Diplomaticsは古文書学とも訳されるが、佐藤雄基によって、日本の古文書学は、明治期にヨーロッパから取り入れられた際に、文書整理と名称付与のために発展したと推定されている（「序章 本書の課題と視点」、『日本中世初期の文書と訴訟』、山川出版社、2012年、7-8頁；「明治期の史料採訪と古文書学の成立」、松沢裕作編『近代日本のヒストリオグラフィー』、山川出版社、2015年、32頁）。ここでは、文書および記録のもつ内外の形式から真正性を評価するというヨーロッパ本来の意味で使用するため、古文書学ではなく、文書形式学と表記する。
- (6) InterPARES Trust, "Document", Terminology Database, <http://interparestrust.org/terminology/term/document/en>.
- (7) InterPARES Trust, "Record", Terminology Database, <http://interparestrust.org/terminology/term/record/en>.
- (8) InterPARES Trust, "Archival Bond", Terminology Database, <http://interparestrust.org/terminology/term/archival%20bond/en>.
- (9) Hobbs, 'Personal Records', pp. 266-267.
- (10) 橋本陽「アーカイブズ学の視点から捉える電子記録管理システムの要点：DoD 5015.2とアルフレスコによる検証」、『京都大学大学文書館研究紀要』第19号、2021年、29-43頁。
- (11) セーブ (Save) の意味。アーカイブズにおける永年保存と区別するため、一時保存という言葉を使用する。
- (12) Victoria L. Lemieux, *Searching for Trust: Blockchain Technology in an Age of Disinformation*, Cambridge: Cambridge University Press, 2022, p. 151, note 9.
- (13) Society of American Archivists, "Processing", The Dictionary of Archives Terminology, <https://dictionary.archivists.org/entry/processing.html>.
- (14) 塩崎亮「特殊コレクションとしてのポーンデジタル個人文書—英米の事例」、『現代の図書館』第58巻第2号（通巻234号）、2020年、100-107頁。
- (15) 塩崎亮「個人による情報管理とパーソナルデジタルアーカイビング：アーカイブズおよび記録管理領域に対する示唆」、『聖学院大学論叢』第32巻第2号、2020年、1-16頁。
- (16) William Jones and Jaime Teevan, 'Introduction', William Jones and Jaime Teevan (eds.), *Personal Information Management*, Seattle: University of Washington Press, 2007, pp. 3-4.
- (17) Jordan Bass, 'A PIM Perspective: Learning Personal Information Management Research in the Archiving of Personal Digital Records', *Archivaria*, No. 75, 2013, pp. 49-76; Anna Chen, 'Disorder: Vocabularies of Hoarding in Personal Digital Archiving Practices', *Archivaria*, No. 78, 2014, pp. 115-134.
- (18) PARADIGM, *Workbook on Digital Private Paper*, 2005-2007, <https://ora.ox.ac.uk/objects/uuid:116a4658-deff-4b06-81c5-c9c2071bc6d0>.
- (19) AIMS Work Group, *AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship*, 2012, <http://www2.lib.virginia.edu/aims/whitepaper/>

- AIMS_final_A4.pdf.
- (20) BitCurator, <https://bitcurator.net>.
- (21) Artefactual Systems Inc., Archivematica, <https://www.archivematica.org/en/>.
- (22) Stanford Libraries, ePADD, <https://library.stanford.edu/projects/epadd>.
- (23) Canadian Centre for Architecture, *CCA Digital Archives Processing Manual*, <https://github.com/CCA-Public/digital-archives-manual>. CCA の整理過程については、日本語での紹介がある。平野泉「『デジタルをおそれずに』—電子記録の保存に関する2つの取り組み—」、『アーカイブズ学研究』第33号、2020年、78-84頁を参照。
- (24) 橋本陽「電子記録をどう整理するか—インターパレスとイタリア・アーカイブズ学における知見に依拠して—」、『レコード・マネジメント』No. 71、2016年、24-38頁。
- (25) Jennifer Meehan, 'Arrangement and Description: Between Theory and Practice', Caroline Brown (ed.), *Archives and Recordkeeping: Theory into Practice*, London: Facet Publishing, 2014, pp. 66-67. コンテキストの種類については、Weimei Pan, 'Context', *Encyclopedia of Archival Science*, p. 152を参照。
- (26) 出所概念の変遷については、Jennifer Douglas, 'Origins and Beyond: The Ongoing Evolution of Archival Ideas about Provenance', Heather MacNeil and Terry Eastwood (eds.), *Currents of Archival Thinking*, Second Edition, Santa Barbara: Libraries Unlimited, 2017, pp. 25-52を参照。
- (27) 日本における試みに関しては、森本祥子「アーカイブズ編成・記述の原則再考—シリーズ・システムの理解から—」、国文学研究資料館編『アーカイブズの構造認識と編成記述』、思文閣出版、2014年、71-96頁；千代田裕子「新たな資料編成・記述方法の導入を目指して 第2部：シリーズ・システムにおける「function」概念の整理」、『東京大学文書館紀要』第39号、2021年、10-19頁；橋本陽「変化する組織が生み出す記録資料群の編成と記述：京大・シリーズ・システムの試行」、『京都大学文書館研究紀要』第20号、2022年、15-29頁を参照。
- (28) 国際的な実践の取り組みについては、元ナミ・橋本陽「アーカイブズ記述の新標準：Records in Contexts (RiC) の実装方法」、『デジタルアーカイブ学会誌』Vol. 6、2022年、S230-S233にまとめられている。
- (29) Adrian Cunningham, 'Collecting Archives and the Australian Series System', Australian Society of Archivists, Committee on Descriptive Standards, *Describing Archives in Context: A Guide to Australian Practice*, Canberra: Australian Society of Archivists, 2004, p. 72; 森本祥子「アーカイブズ編成・記述の原則再考」89-91頁；加藤聖文「近現代個人文書の特性と編成記述—可変的なシリーズ設定のあり方—」、『アーカイブズの構造認識と編成記述』、195-196頁。
- (30) Geoffrey Yeo, 'Continuing Debates about Description', *Currents of Archival Thinking*, Second Edition p. 166. 原秩序概念の変化の歴史については、Douglas, 'Origins and Beyond' を参照。
- (31) Jennifer Meehan, 'Rethinking Original Order and Personal Records', *Archivaria*, No. 70, 2010, pp. 27-44.
- (32) Jinfang Niu, 'Original Order in the Digital World', *Archives and Manuscripts*, Vol. 43, No. 1, 2015, pp. 61-72.
- (33) Jane Zhang, 'Original Order in Digital Archives', *Archivaria*, No. 74, 2012, pp. 27-44.
- (34) 安藤正人『記録史料学と現代—アーカイブズの科学をめざして—』、吉川弘文館、1998年、119頁。
- (35) 同上、111頁。
- (36) 同上、120頁。
- (37) Johannes Papritz, *Archivwissenschaft*, Band 1 und 2, Marburg: Archivschule, 1983.
- (38) 橋本陽「個人文書の編成—環境アーカイブズ所蔵サリドマイド関連資料の編成事例—」、『レコード・マネジメント』No. 66、2014年、50頁。
- (39) Lyn Milton, 'Arrangement & Description', Jackie Bettington et al. (eds.), *Keeping Archives*,

- Third Edition, Canberra: Australian Society of Archivists, 2008, p. 262.
- (40) Giovanni Michetti et al., 'Intellectual Control', Luciana Duranti and Corinne Rogers (eds.), *Trusting Records in the Cloud*, London: Facet Publishing, 2019, p. 156.
- (41) Elio Lodolini, *Archivistica: Principi e Problemi*, 14ª Edizione, Milano: Franco Angeli, 2014, p. 216.
- (42) Papritz, *Archivwissenschaft*, Band 3, S. 71.
- (43) Peter Horsman, "Het zat in de lucht." De concepten achter de Handleiding', *Tekst en context van de Handleiding voor het ordenen en beschrijven van archiven van 1898*, Hilversum: Uitgeverij Verloren, 1998, pp. LXXXIX-XC; 'Taming the Elephant: An Orthodox Approach to the Principle of Provenance', Kerstin Abukhanfusa and Jan Sydbeck (eds.), *Principle of Provenance: Report from the First Stockholm Conference on the Archival Principle of Provenance, 2-3 September 1993*, Stockholm: Swedish National Archives, 1994, p. 58.
- (44) Heather MacNeil, 'Theory and Practice: Between Two Paradigms', *Archivaria*, 1994, No. 37, p. 10.
- (45) David Bearman, 'Item Level Control and Electronic Recordkeeping', *Archives & Museum Informatics*, Vol. 10, No. 3, 1996, pp. 197-198.
- (46) Michetti et al., 'Intellectual Control', pp. 158-178.
- (47) 橋本「アーカイブズ学の視点から捉える電子記録管理システムの要点」、33-34頁。
- (48) 古賀崇「記録管理・アーカイブズにおける「デジタル・フォレンジック」に関する一考察—国際比較に基づき—」、『レコード・マネジメント』No. 73、2017年、72頁。
- (49) 同上、77-80頁。
- (50) Corinne Rogers, 'Diplomatics of Born Digital Documents — Considering Documentary Form in a Digital Environment', *Records Management Journal*, Vol. 25, Iss. 1, 2015, p. 9.
- (51) Canadian Centre for Architecture, 'Arrangement of Born-Digital Archives', <https://github.com/CCA-Public/digital-archives-manual/blob/master/guides/arrangement.md>.
- (52) Luciana Duranti, 'Archival Bond', *Encyclopedia of Archival Science*, p. 28.
- (53) Canadian Centre for Architecture, 'Arrangement of Born-Digital Archives'.
- (54) Braian Lavoie, *The Open Archival Information System (OAIS) Reference Model: Introductory Guide*, Second Edition, 2014, p. 18, <http://dx.doi.org/10.7207/twr14-02>.
- (55) Luciana Duranti, 'From Digital Diplomatics to Digital Records Forensics', *Archivaria*, No. 68, 2009, p. 53.
- (56) Artefactual Systems Inc., "Arrangement", *Archivematica*, <https://www.archivematica.org/en/docs/archivematica-1.13/user-manual/appraisal/appraisal/#arrangement>.
- (57) BitCurator NLP, <https://bitcurator.net/bitcurator-nlp/>.
- (58) 加藤「近現代個人文書の特性と編成記述」、181-199頁。
- (59) 早川和宏「民間（収集）アーカイブズの保存活用を巡る法的課題—その利用を中心に—」、『国文学研究資料館紀要 アーカイブズ研究篇』第13号（通巻第48号）、2017年、61-83頁；阪本尚文・堤裕一郎・横島善子「未公開エゴ・ドキュメント研究と著作権法—高橋幸八郎関係書簡を事例として—」、『商学論集』第89巻第3号、2020年、13-24頁。
- (60) 堀内暢行「emailを歴史資料にするために：OSS・ePADDのススメ」、『国士館史学』第26号、2022年、116-136頁。