

デジタル時代のアーカイブズ

公益財団法人渋沢栄一記念財団 デジタルキュレーター 金甫榮

2023年8月15日発行

<目次>

1. はじめに：アーカイブズ (archives)
2. デジタル記録の特徴
3. どうやって保存するか？
4. アーカイブマティカ (Archivematica)
5. おわりに

【注】

1. はじめに：アーカイブズ (archives)

誰もが「アーカイブズ (archives)」という言葉を一度は耳にしたことがあるだろう。最近ではIT用語として古くなったデータをまとめて保存する際によく使われるようになったが、その語源は古代ギリシャ語の ἀρχεῖον「アルケイオン」とされている¹。その意味は、統治者の執務場所またはそこで作成された記録を保存する場所という意味で使われ、権威と権力の象徴で神聖なものとして考えられていた。アーカイブズ学分野では、長期的に保存する価値のある記録（を保存する場所）という意味として用いられている。「長期」「価値」「記録」には、専門用語として様々な定義や理論が存在するが、本稿では「長期」の意味に注目したい。

博物館や図書館、文書館のような史資料を保存する施設では、特別な収蔵庫があり温室度や害虫防止など徹底した保存対策を講じている。これまでの記録資料のほとんどは紙を用いているため、保存対策も主に紙を対象にしていることが多いだろう。しかし、近年問題となっているのはデジタル記録の保存である。紙のように収蔵庫で保存するというわけにはいかない。これまでの保存対策は、デジタル記録には通用しないのだ。

2. デジタル記録の特徴

では、デジタル記録はどのように保存すればよいだろうか。それを知るためには、デジタルの特徴を理解する必要がある。デジタルの特徴としてよく言われるのが「三つの寿命」である。

一つ目は「媒体の寿命」である。光ディスク（CD や DVD など）は10～30年、ハード

ディスクは約 5 年、フラッシュメモリー（USB メモリーなど）は約 10 年も経てば使えなくなるが、保存環境や扱いによっては寿命を迎える前に突然データが消えてしまうこともある。

二つ目は「ファイルフォーマット（形式）の寿命」である。デジタルは、数字の 0 と 1 の配列で表現されるが、その順序や構造によってフォーマットが異なり、情報を読み取る際には正しいフォーマット情報を取得する必要がある。しかし、ファイルフォーマットはバージョンが古いと読み取れなくなる場合や、なくなってしまうこともある。そのため、ファイルを人間が読めるようにするためには、定期的にフォーマットのリスクを評価しなければならない。

三つ目は「再生機器の寿命」である。再生機器にはハードウェアだけではなく、ソフトウェアも含まれる。例えば、フロッピーディスクに保存された文書を読み取るためには、まずフロッピーディスク専用のドライブでファイルを読み込み、そのファイルを表示するには文書作成で使用したソフトウェアが必要である。

デジタルのもう一つの特徴は、「記録」として必要な要素が分散していることである。記録は独立した単一の文書ではなく、組織の業務に関連づけられた文書群として存在する必要がある。例えば決裁書の作成から保存までを考えてみよう。組織には決裁書の作成者が存在し、権限のあるものが承認を行ない、それが然るべき場所のファイルに綴じられる。この一連の過程は組織で意思決定のプロセスとして認知されており、その過程が証明できるからこそ決裁書が証拠となる記録として有効である。しかし、デジタルの場合はこの一連の過程が、様々なシステムの中でメタデータ（データに関するデータ）としてバラバラに存在する。これらのメタデータを意思決定がなされたコンテキスト（文脈）が分かるように残すことが不可欠だが、分散された情報を集めて保存することは容易ではない²。

さらにデジタルの保存を難しくするのは、いつでも改変できていくらでも複製物を作成できる点である。そのため紙記録で言う原本の概念は崩壊し、デジタルの長期保存には常にいつだれが作成・変更したかを記録しなければならなくなった。デジタル時代のアーカイブズは、データの継続的なモニタリングとその内容を記録することが求められる³。

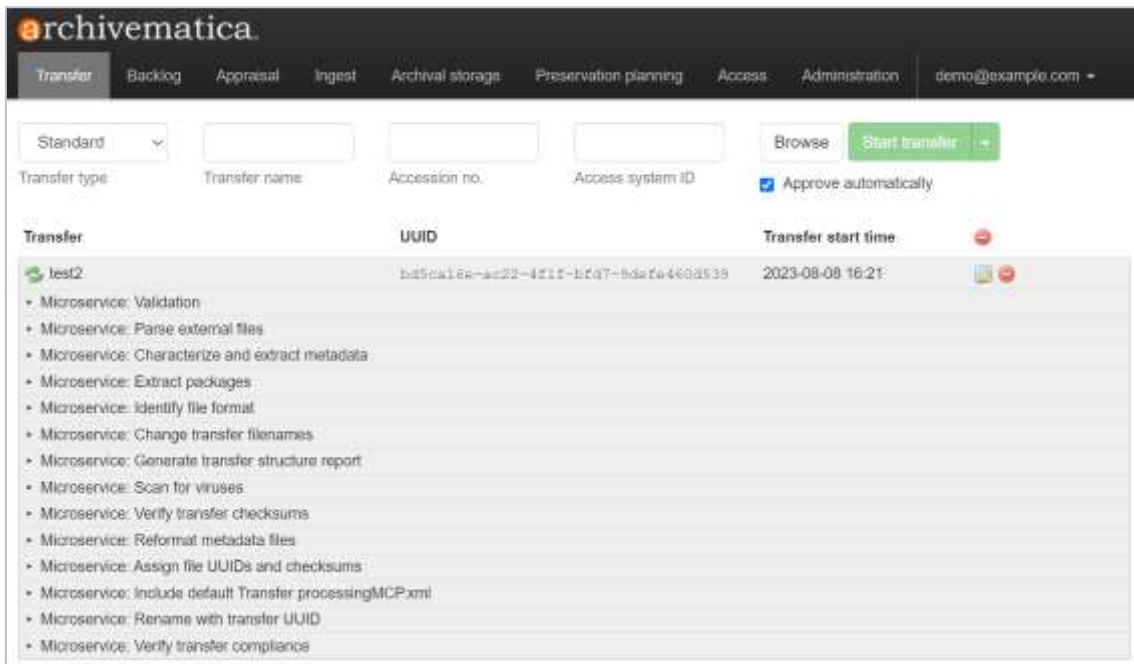
3. どうやって保存するか？

デジタル記録の保存における弱点を克服し、長期的に利用可能にするための国際標準として「OAIS 参照モデル（Reference model for an Open Archival Information System, ISO14721）」⁴がある。この標準は、アメリカ航空宇宙局（NASA）が 1975 年に打ち上げた火星探査機のデジタルデータが、わずか 25 年後に解読不可能となったことをきっかけに誕生した。OAIS 参照モデルは、デジタルを保存する機関またはシステムに必要な機能と情報を概念的なレベルで定めている。しかし、その実装は容易ではなく国内の導入事例は非常に少ない。

4. アーカイブマティカ (Archivematica)

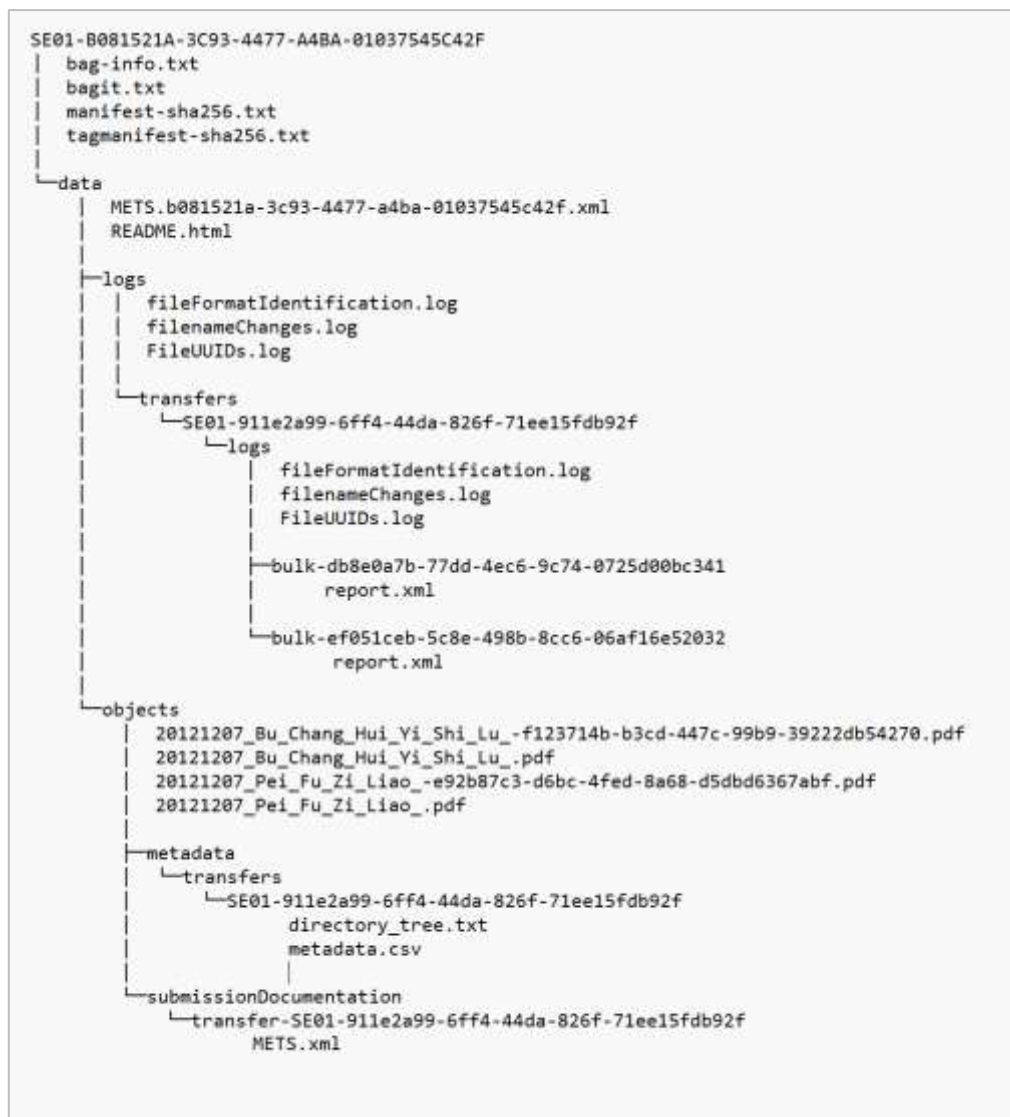
このような時代の変化を受けて、公益財団法人渋沢栄一記念財団（以下「財団」）ではアーカイブマティカ (Archivematica) ⁵ という長期保存を支援するソフトウェアを活用することに挑戦している。アーカイブマティカを活用する利点の一つは、OAIS 参照モデルによる情報の管理ができることである。(図 1) は、アーカイブマティカの操作画面だが、保存対象となるデータを取り込めばサイト上で処理の進捗状況が確認でき、最終的に長期保存に適切な情報パッケージが作成され、指定のストレージに保存できる仕組みとなっている。

(図 1)アーカイブマティカの操作画面



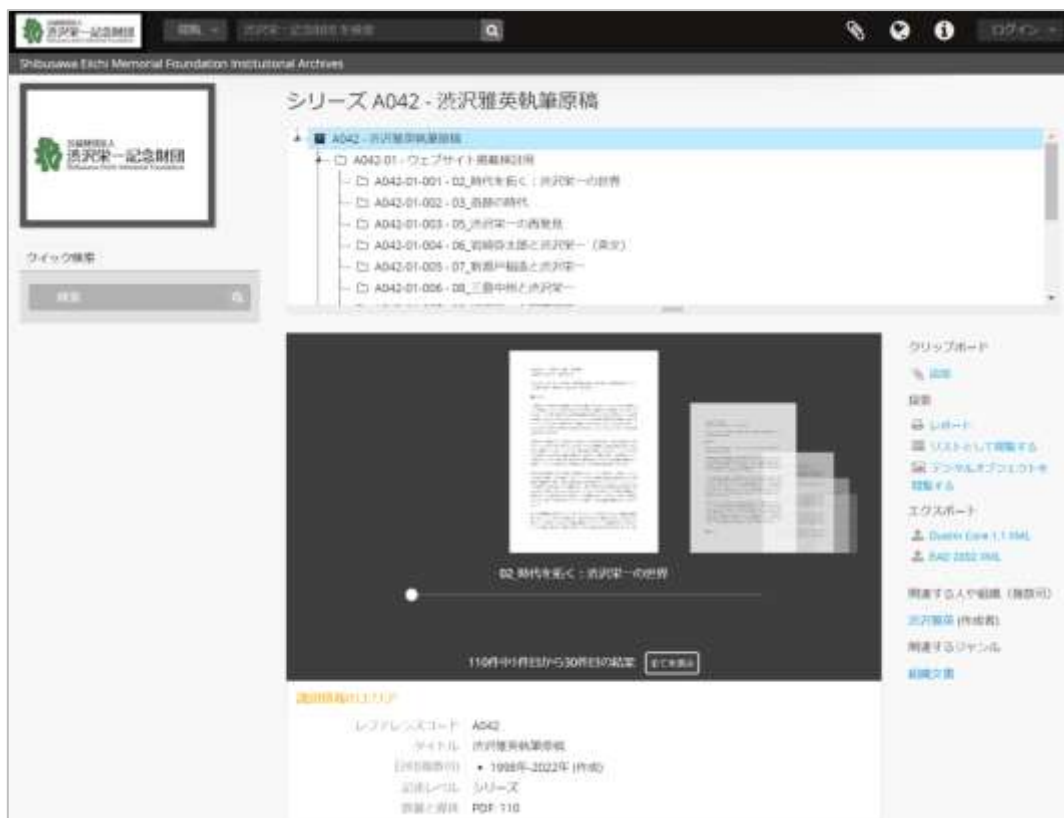
(図 2) は、2 つの PDF 形式の文書ファイルをアーカイブマティカに取り込んで作成した情報パッケージの例である。この図だけでは、意味することが読み取りにくいかもしれないが、データを長期的に利用可能にするためには、沢山の情報が作成され、関連付けられ、機械が分かるように記録されなければならないということが、このツリー構造から伝わってくる ⁶。

図 2—アーカイブマティカで作成した情報パッケージ



財団にとってアーカイブマティカを活用するもう一つのメリットは、すでに導入しているアトム（Access to Memory: AtoM）⁷ との連携が可能ということである。アトムは、アーカイブズ資料の情報を管理・利用するためのオープンソースソフトウェアである。財団では、2019年に「組織アーカイブズ閲覧システム」の基盤システムとしてアトムを導入し、国内で数少ない活用事例となった。（図3）は、最近財団内部で公開した渋沢雅英相談役の講演関連の原稿である。渋沢相談役は渋沢栄一の曾孫であり、23年間財団の理事長を務めた。在任期間中（1997-2020）には数多くの講演を行ない、その記録は組織アーカイブズとして貴重な資料となるため、アーカイブズとして内部で利用できるようにしている。もし、アーカイブマティカとアトムの連携によるデジタルの長期保存と利用のシステムが完成すれば、国内で初めての導入事例となるだろう。しかし、導入の必要性や実現可能性については綿密な検討が必要であろう。

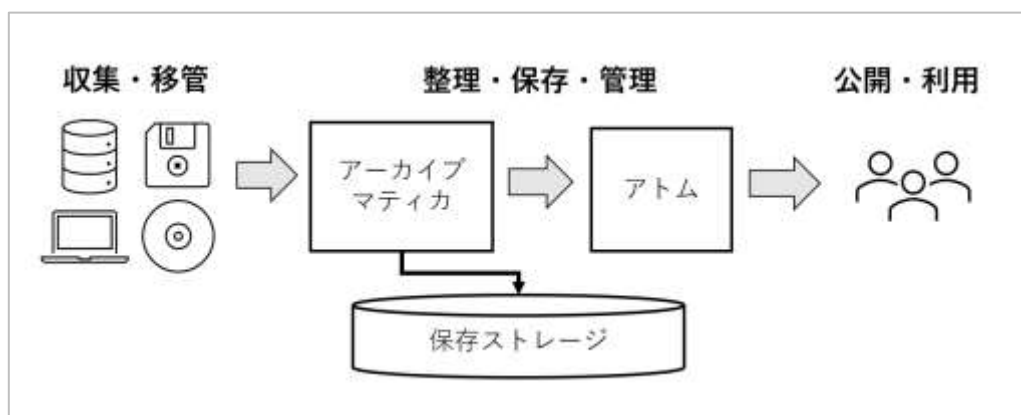
図 3—アトム(AtoM)を用いた「組織アーカイブズ閲覧システム」の画面



5. おわりに

本稿では、デジタルの長期保存における課題について述べ、長期保存を支援するツールとしてアーカイブマティカを、整理されたアーカイブズ資料を管理・公開するツールとしてアトムを紹介した。(図 4) はこれらをアーカイブズ資料の受入から利用するまでのワークフローの中に位置づけた一例である⁸。

図 4—アーカイブマティカとアトムを活用したワークフローの例



しかし、冒頭で述べたようにデジタルをアーカイブするためには「価値」と「記録」という側面も検討しなければならない。企業のような組織においてはとりわけ重要な側面であるが、ごく簡単に言えば、「活動の証拠となるもの」を保存することであり、(図4)の「収集・移管」の段階(さらにその前の段階)に深く関わる。これについては、次にいくつかの参考文献を紹介することに止めたい。

記録の作成からアーカイブズの管理までを理解する上で参考になるもの

- (1) エリザベス・シェパード、ジェフリー・ヨー著、森本祥子ほか訳『レコード・マネジメント・ハンドブック-記録管理・アーカイブズ管理のための』、日外アソシエーツ、2016、393頁。
- (2) 金甫榮「業務分析に基づく民間組織の記録とアーカイブズの管理に関する試論」、『アーカイブズ学研究』29巻、2018、4-29頁。

https://doi.org/10.32239/archivalscience.29.0_4

記録の作成からアーカイブズの管理までを理解する上で参考になるもの

- (3) 橋本陽「守るべき記録の特性と進めるべき業務の電子化」、『アーカイブズ学研究』37巻、2022、22-32頁。
- (4) 金甫榮「電子記録の保存に必要なことー基礎的な理論に基づく実践を目指してー」、『記録と史料』33号、全国歴史資料保存利用機関連絡協議会、2023、2-9頁。

[注](参照日は全て2023-8-1)

- 1) ブリュノ・ガラン著、大沼太兵衛訳『アーカイヴズ 記録の保存・管理の歴史と実践』、白水社、2021、13-14頁。
- 2) より詳しい説明は次が参考になる。公益社団法人日本文書情報マネジメント協会(JIIMA)「電子文書信頼性ガイドライン(第一版)」、2019。
https://www.jiima.or.jp/wp-content/uploads/policy/190328_denshibunsho_sinraiseikoujou_guideline_V1_1.pdf
- 3) 詳細は次を参照のこと。金甫榮「組織アーカイブズにおける真正なデジタル記録の長期保存の要件: Archivematicaを用いた検討」、『アーカイブズ学』38号、2023、4-35頁。
- 4) ISO 14721: 2012 (Space data and information transfer systems - Open archival information system (OAIS) - Reference model)
- 5) アーカイブマティカ <https://www.archivematica.org/en/>
- 6) 情報パッケージに関する詳細は、前掲3を参照のこと。
- 7) アトム <https://www.accesstomemory.org/en/>
- 8) アーカイブマティカを活用したワークフローを検討した事例は次を参照のこと。金甫

https://www.shibusawa.or.jp/center/ba/bunken/doc021_kim01.html

榮「真正なデジタル化資料の長期保存と公開：Archivematica と Omeka S を用いた事例」、『デジタルアーカイブ学会誌』6 巻 s3 号、s147-s150 頁。https://doi.org/10.24506/jsda.6.s3_s147