

歴史資料保全のためのデジタルデータ

文●後藤 真

はじめに

本章では、情報技術を応用した保全の状況について触れる。歴史資料に関わる部分のみならず、今やありとあらゆる部分において情報技術が活用される状況になっているのは、いうまでもない。世界中にインターネットの網が張り巡らされ、どのような場であっても、Webを介して情報を得ることができるようになりつつある。とはいえ、情報技術もWebも、道具であるという本質は何らの変わりがない。本章では、資料ネットに関わる情報技術について触れ、資料を災害から守り、未来に継承するための方法について述べる。

1. デジタル・Web・公開と共有

■デジタル化の本質と特徴

デジタル化の本質は「複製と再加工が容易である」ということである。複製が容易であるということは、データを必要に応じて多くの人と共有でき、万一の際の消失に対応しやすいということである。また、再加工可能ということは、たとえば歴史資料の場合は、その資料群の目録から印刷媒体へと変換することも容易であり、緯度経度などを付与することによって地図上に表示させる、検索が簡単にできるようなデータベースを作り上げる、目録と画像とテキストなどを効果的につなげ、情報を豊かにするなどの効果があるということである。

と、同時にケアをし続ける必要がある記録方式である点も特徴である。保持する媒体は、機械的に壊れる可能性が低い光ディスク等であっても、ディスクに傷が入るなど、何らかの理由があれば、データはまとめて破壊される。また、1・0のデータが完全な形で残されていても、その媒体を読み込む機械が失われると読み込むことができない。光ディスクであれば、それを読み込む機械が、あるいはコンピュータと機械をつなぐケーブルの規格（たとえばUSBの一部規格は見かけることが少なくなっている）が失われるとアクセスできなくなる。上記が可能であっても、データのフォーマットが変わるとアクセスは困難になり、ソフトウェアの形式が変わってもアクセスは困難になってしまう。

一方で、データの複製と再加工は容易であることから、フォーマットの変化に常に対応する形で、データをマイグレーション（データを利用可能な形に変換しつつ移行）して持ち続けられれば、完全な形でデータは保たれる。その点からも「ケアをし続ける」

ことが重要な記録形式なのである。

■Web に情報をアップするということ

そして、上記のデジタル化の本質は「Web の本質」と異なることも指摘しておかなければならない。デジタル化することは、即 Web に情報が上がるようなイメージを持つことも多くなってきたが、これは必ずしも同一ではない。デジタル化したものを Web に掲載せずに保管することも重要な選択肢であることを忘れてはならない。Web は上記のデジタル化の本質をより強化するツールとして理解することが望ましい。空間を超え、複数の人や場所での複製保管を可能にする。それは、万一の災害の際に、遠隔地で相互バックアップを実施するための大きな助けとなる。また、再加工し GIS (Geographic Information System = 地理情報システム) 等で資料の所在情報などを地図に表示させ、可視化するなどの成果を閲覧する際にも Web を介してより多くの人に閲覧してもらうことでこそ、その機能を最大限発揮させることになるであろう。

さらに、Web に情報をアップするというのも、即全面公開を意味するものではない、ということも、近年の動向としては指摘しておいた方がよいであろう。近年はさまざまなクラウドサービスが充実することにより、Web 上で、関係者のみ公開を行う方法が増えてきた。Web にアップする以上、セキュリティへの対応は欠かすことができないが、一方で、単に自前で準備するサーバだから安全であり、クラウドサーバだから危険というような状況もすでに過去のものとなっている。むしろ、セキュアなクラウドを選択することが、Web 上での情報共有ではむしろ効果的であるともいえるであろう (ただし、この場合も Web サービスそのものが同じ形式で永続する可能性は低いため、クラウド間でのマイグレーションなどの手当ては常に必要であると考えておくべきであろう)。そのため、デジタルデータの管理という観点からは、この Web クラウドでの共有 (全面公開・限定公開) という部分が、今後さらに大きな位置を占めるようになってくるといえるであろう。

Web における全面公開という機能を考えるならば、いまだに「フロー」の側面が大きいのも事実である。資料の情報を可視化し、所在を伝えることにより災害時の保全に役立つ、資料への理解を深め、文化への理解から資料保全の活動への理解へとつなげるなどの機能である。目録の公開状況 (著作権上オープンなライセンスを宣言するなど) によっては、複製を第三者に持ってもらうことも期待する。ただし、これは「複製が別の場所にある可能性を高める」ことになって、全面解決の方法ではないことは注意しておきたい。

一方で、Web に公開しデータを長期的に保全しようとする考え方として DOI (Digital Object Identifier) というものが存在する。これは、Web 上にあるデジタル資源 (アナログな資料を指し示すデジタル目録も含まれる) について、永続的に一意な (つまり、ほかに重複するものがない) 情報の URL を指し示す技法である (この技法について詳しく示す紙面はないため、『DOI ハンドブック <

handbook/translations/japanese/hb.html』などを参照いただきたい)。これまでは Web 上にある論文等に用いられてきたが、近年はそれ以外の資料に付与することも行われるようになってきた。この DOI を公開されたデジタル資源に付与することで、永続的なアクセスを担保するのの一つの方法である。ただし、DOI はあくまでも、デジタル資源へのアクセス経路を確保するだけであり、最終的なデジタル資源そのものは、デジタル資源の公開者が維持し続ける必要がある。この点からも DOI もデジタル化の本質を抜け出すものではなく、あくまでも補助ツールであることがわかる。

2. 資料保全のために構築するデジタルデータとは

■ 資料所在情報のデータ化と複製

それでは、このような観点から見た場合、資料保全のためのデジタルデータとはどのようなものかを考えるべきであろうか。Excel や資料撮影といったノウハウとは異なる視点から、述べておくことにしたい。

まず、やらなければならないのは、資料所在情報のデータ化と複製である。歴史資料保全の本質が現物の保全である限り、まずはそれらの状況を把握することが求められる。しかし、詳細な目録をとるほどの余力がない場合は多いであろう。そのような状況の中でも必要なのは、最低限の資料の所在（資料群がどこにあるのか。どのような資料なのか）のみでも記したデータを作成しておくことである。そして、そのデータは必ずしも公開できるものではないため、そのデータについては遠隔の信頼できる機関との連携のもとで共有しておくことが重要であろう。無論、その共有先は一機関ではなく複数であってもよい。

また、共有の方法として、セキュアなクラウドを選択することも、可能性としてはありうるであろう。それにより、データを入れた物理媒体が万一被災しても、データは守られる。また、被災地の人が資料保全活動を行うことが困難な場合にも、そのクラウドに別の地域の人がアクセスできれば、そこから保全活動を行うこともできる。そのような体制を作り上げたのちに、それらの資料群の中で公開可能なものがあるのであれば、それを抽出し、公開用の画像撮影や詳細な目録を作成するという流れが、より現実的なものであると考える。このような二段階のデータ作成を行うことで、資料所在情報のデータをより容易に蓄積することができると考えられる。

■ 記録を目録とともに残す

また、その際に重要なのは、「いつ・だれがそのデータを作ったのか」の記録を目録と合わせて残すことである。筆者はこれまでに過去の資料目録の電子化を進めてきているが、その中には「いつの段階の資料の情報で、どのような意図でこのように作られたり修正されたりしたのかが不明になっている」目録が、多く見られる。このことにより、再度目録と資料そのものを突合せさせるような状況が生じており、いざというときに活用が困難になる可能性がある。実際には、データ作成者の名前があっても、

のちの段階で作成者にその目録記述の意図を確認することは困難ではあるだろうが、状況を推定する際の重要なヒントになりうる。無論、データ作成の意図や作業経過を記録できるのであれば、より望ましい。

このように考えるならば、資料保全のためのデジタルデータとしては、資料そのものの情報もさることながら、「いつ、だれが作成したどこにある資料の情報なのか」といった、一段階メタな情報の記録が求められていることがわかる。

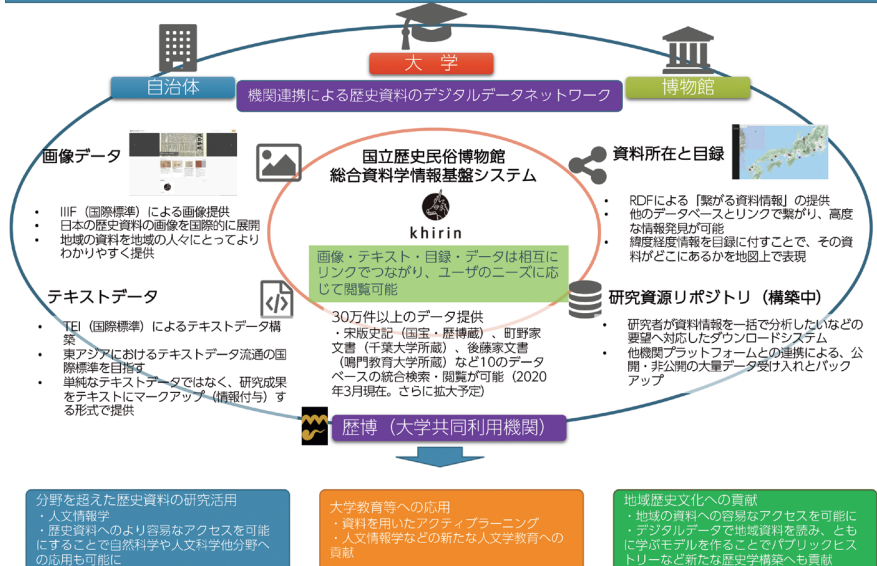
3. 歴史文化資料保全のための目録とデータ保全のための 歴博の取り組み

■情報基盤システム「khirin」

現在、国立歴史民俗博物館では khirin という、地域における歴史資料のデジタルネットワークの構築を進めている^①。khirinは、特定のシステムを指し示すものではなく、全国の歴史資料のデジタルインフラとしての複数の機能を持った総体を指す。

歴史文化資料保全の文脈においては、全面公開と部分公開が可能であり、複数の資料目録を統合的に検索可能で、かつ関連する資料をリンクでつなぎながら閲覧することができる khirin ld と、khirin c という名称のシステムが位置づけられている。これはリンクが可能であるというのみならず、歴史資料の目録であれば、かなり柔軟にデータを入れることができ（1セル1データの Excel であれば、どのような形式のも

大学と自治体・博物館の歴史資料データインフラである総合資料学情報基盤システムkhirin



①情報基盤システム「khirin」

のでも投入可能である)、必要に応じて緯度経度を付して、その資料の場所を地図上で示すことも可能になっている。

この目録情報に、画像等のデータを結び付けて公開する場合には、khirin a という画像公開システムにデータを入れることで、対応できるようになっている。また、非公開の情報については、歴博内のサーバやディスクでデータを預かるなどの対応も実施しており、歴史資料保全のためのデータの蓄積と公開状況に応じた形で、対応可能なものとした。外部機関とも連携しつつクラウドによるデータ蓄積システム (khirin r) の構築も進めており、そこには保存用に加えて、機械的な解析もできるような形でデータを置くことも想定している。また、デジタルデータに関わるデータ作成の情報(いつ・だれが・どのように)といったものも、このクラウドの中に入れることで、よりデータの保全を行いやすい形とすることができるであろう。

歴史資料の実物を保全するためのデータについては、無論これ以外にも、現物が被災した際の複製としてのデジタルデータ、あるいは大量にあり、そもそも日常から保持が難しいものの複製デジタルデータなどや、さらに高度な活用も想定されるであろう。しかし、まずは所在目録と、データの来歴を作り、それを的確な形で共有する(共有先は、無論歴博以外の多様な選択肢があってよい)ことから始める必要がある。デジタルデータの特性を把握しつつ、変わりゆく技術的状况の中で、絶対を守るべきものと、柔軟に対応するものの両面を見ていくことが必要であろう。

さらに深く知りたいときは

- ① **Digital Preservation Handbook** : <https://www.dpconline.org/handbook>
Executive Guide on Digital Preservation : <https://www.dpconline.org/digipres/implement-digipres/dpeg-home>

デジタル保存に関しては、日本語の文献は多くない。英語では上記を参考にできる。

- ② 『歴史情報学の教科書』文学通信、2019年

歴史資料のデジタル化に関する基本的な知識や情報は、本書にも網羅的に描かれている。目録・画像・テキストやクラウドソーシングなどの基本的な考え方について、詳しく把握するためには本書を参考にできる。なお、本書は無料でのアーカイブ公開も存在する。

<https://bungaku-report.com/blog/cat43/>

- ③ 岡田一祐 『ネット文化資源の読み方・作り方 図書館・自治体・研究者必携ガイド』文学通信、2019年

本書発行段階までのいわゆるデジタルアーカイブの解説と評論が行われている。本書に書かれているコメントや評価は、実際にデータ構築を行う際の重要なヒントとなるとともに、2010年代後半のデジタルアーカイブ全般を見通せる。