

地域資料のデジタルアーカイブ化による利活用と保存について —青森県史デジタルアーカイブシステム構築と文化財及びテキストデータのあり方—

古川 淳一（青森県環境生活部県民生活文化課）・伊藤由美子（青森県教育庁文化財保護課）

Utilization and Preservation of Regional Historical Records through Digital Archiving
Furukawa Jun'ichi (Aomori Prefecture, Department of Environment and Public Affairs, Lifestyle and Culture Division)
Ito Yumiko (Aomori Prefecture, Board of Education, Cultural Property Protection Division)

- ・自治体史／Local government history ・地域資源／Local resources
- ・デジタルアーカイブ／Digital archives ・青森県／Aomori prefecture

1. 青森県史編さんの経緯とデジタルアーカイブス

「青森県史デジタルアーカイブシステム」は2019年3月20日に青森県環境生活部県民生活文化課が公開を開始したデジタルアーカイブスである。
<https://kenshi-archives.pref.aomori.lg.jp/contents/kenshi-front/>

アーカイブスとしての特徴は、青森県が同課に事務局を置いて刊行した『青森県史』とその編集過程において実施された調査の成果及び収集した史資料をもととしたものであることにある。そこで、まず『青森県史』について簡単に説明する。

(1) 青森県史の編さん

青森県史編さん事業は1996年4月に開始された。それに先立ち、準備段階として編さん方針・目的や組織を定め、委員の人選などが行われている。

青森県史は当初考古、古代、中世、近世、近現代、民俗、文化財、自然の8つの専門部会からなり、全50巻の刊行を予定していた。各専門部会に属する委員とそれを担当する事務局職員・嘱託員とが協同で調査・編集・執筆・校正に当たる態勢である。県史の他、関連刊行物として主に民俗・文化財部会により「青森県史叢書」を刊行し、他に「青森県史研究」「青森県史資料所在目録」「青森県史だより」を刊行していた。ところが、2004年に事業の大幅な見直しを図ることとなり、編さん期間を延ばした上で「県

史」「県史叢書」は統廃合により巻数を見直した。「県史研究」「所在目録」は廃刊、その後「だより」もインターネットによる広報に置き換えていった。刊行を取りやめた県史の中には「年表・索引」がある。事業を見直した当時はデジタルアーカイブスのことを考えていたわけではないが、書籍として年表・索引を編集・校正する作業量、検索語句の選定の難しさを考えると、編集を行う委員や事務局職員にとってデジタルアーカイブスに替えることができたのは幸いであったと言えよう。それにも増して利用者にとって編集者の設定した検索語句に縛られず、自ら自由に検索できる便宜ははかりしれないものと言える。

途中、このような事態を経ながらも、2018年3月に通史編3巻の刊行をもって青森県史は完成した。「青森県史」36巻、「青森県史叢書」14冊、奇しくも当初計画の県史50巻と構成はまったく異なるが同じ巻数となっていた。

表1 青森県史の構成
青森県史

編名	題	副題	刊行年度
資料編	考古1	旧石器・縄文草創期～中期	2016
	考古2	縄文後期・晩期	2012
	考古3	弥生～古代	2004
	考古4	中世・近世	2002
	古代1	文献史料	2000
	古代2	出土文字資料	2007
	中世1	南部氏関係資料	2003
	中世2	安藤氏・津軽氏関係資料	2004

	中世3	北奥関係資料	2011
	中世4	金石文・編さん物・海外資料・補遺	2016
	近世1	近世北奥の成立と北方世界	2000
	近世2	津軽1 前期津軽領	2001
	近世3	津軽2 後期津軽領	2005
	近世4	南部1 盛岡藩領	2002
	近世5	南部2 八戸藩領	2010
	近世6	幕末・維新期の北奥	2014
	近世	学芸関係	2003
	近現代1	近代成立期の青森県	2001
	近現代2	日清・日露戦争期の青森県	2002
	近現代3	「大国」と「東北」の中の青森県	2003
	近現代4	昭和恐慌から北の要塞へ	2004
	近現代5	占領と改革の時代	2008
	近現代6	高度経済成長期の青森県	2013
	近現代7	青森論	2015
	近現代8	日記	2016
民俗編	資料 南部		2000
	資料 下北		2006
	資料 津軽		2013
文化財編	建築		2015
	美術工芸		2010
自然編	地学		2000
	生物		2002
別編	三内丸山遺跡		2001
通史編	1	原始 古代 中世	2017
	2	近世	2017
	3	近現代 民俗	2017

青森県史叢書

担当部会	タイトル	刊行年度
民俗	奥南新報「村の話」集成 上・下	1997
民俗	馬淵川流域の民俗	1998
民俗	小川原湖周辺と三本木原台地の民俗	2001
民俗	下北半島北通りの民俗	2002
古代	青森県史資料編古代1補遺	2002
民俗	下北半島西通りの民俗	2003
文化財	下北の仏像	2005
文化財	青森県の暮らしと建築の近代化に寄与した人々	2006
民俗	岩木川流域の民俗	2007
文化財	南部の仏像	2008
民俗	西浜と外ヶ浜の民俗	2009
文化財	津軽の仏像	2010
文化財	近現代の美術家	2011

(2) デジタルアーカイブスへの経緯

さて、青森県史デジタルアーカイブスは県史完成の最終年度2017年度からシステムの検討と構築を始め、県史完成の祝賀の席ではデジタルアーカイブ

ス構想についての説明もできるようになっていた。だが、それは後の章に託し、ここでは県史編さん事業のはじめに立てられた編さん方針にみられる特徴、萌芽といえるもの、さらに自治体史において収集した編さん資料の特徴とデジタルアーカイブスとの関係について記すこととする。

県史編さん方針の中には「マルチメディアの積極的活用」が謳われている。編さんを始めた年はWindows95リリースの翌年。まだインターネットは普及しておらず、パソコンを使う者の一部に、電話回線に繋いでパソコン通信を行う者もいたという時代で、さらにパソコンで利用できる文字も現在とは比較にならないほど限られていて、文献史料を入力するにも制約があった。このように県史編さんにコンピュータを活用するといったことがイメージされがたい状況下ではあったが、一方で歴史学へのコンピュータ利用が動き始めた時代でもあった。委員の一部にもそうした思いをもつ者がいて、準備段階で県史編さん方針に盛り込むことが話し合われた結果、当時一般に理解される言葉として「マルチメディアの積極的活用」となったと聞いている。当初にこうした方針を立てたことをどこまで編さん事業に反映させることができたか、デジタルアーカイブスにどのように結実したか、確かなことは言いがたいけれども、常にそうした意識を持って編さんに臨んでいた人々がいたことはデジタルアーカイブスの構想を練る礎となったと言えよう。

(3) 県史編さん資料の特性とデジタルアーカイブス

第一に、県史編さんは多くの時代、分野にわたることである。前述のとおり青森県史はスタート時に8つの専門部会があり、時代としては「古代」「中世」「近世」「近現代」の4部会、分野としては「考古」「民俗」「文化財」「自然」の4部会があった。それぞれ資料の形態、調査・収集方法などに差異が認められるものであり、資料整理のあり方も自ずと異なるところである。こうした調査成果、収集資料の違いは「青森県史デジタルアーカイブス」が5つのデータベースによって構成されていることにつながる。このことについては次章以下で記す。

次に、博物館や美術館、図書館などのアーカイブスと大きく異なる特徴として、県史編さん資料は自ら所蔵するものは限られていて、他の資料所蔵機関や社寺、個人などが所蔵する資料がはるかに多いということが挙げられる。それらを調査して写真撮影や調書作成により県史編さん資料を収集していくのである。ほかに文献史学の分野では先行する書籍等を悉皆調査して翻刻史料を調査カード化する場合もあり、民俗分野などでは資料の中に聞き取り（録音）や動画撮影によるものも含まれる。調査での撮影・調書作成等は資料所蔵先で行うこともあれば、資料量や調査先の環境によっては資料を借用して県史編さん室で行うこともある。いずれにせよ、現代では資料は現地保存主義に立ち、資料は借用した場合も速やかに返却する。そして、編集・執筆は撮影した複製をもとに行うのが一般的である。

このように普段の利用には複製を活用する。このことはデジタルアーカイブスを作成して利用者に供することにもつながる。また複製の作成は原資料が被災する事態への備えともなるものである。

だが、デジタルアーカイブスでの活用は、県史編さん終了間近となって考えられたことであり、調査・資料収集を行っていた当時は県史を編さんするため、すなわち書籍に掲載することを前提として資料所蔵者の協力を得るものであった。現在あるいは今後に行われる自治体史編さんであれば調査ないし編集の時点でインターネットでの公開も含めて利用許諾を得ることもあろうが、青森県史編さんの頃はそうした例はほぼ皆無であり、デジタルアーカイブスで公開するにあたってはあらためて所蔵者・機関に対して許諾を得る必要があった。ただし、編さん開始すなわち調査開始からデジタルアーカイブス構築までは20年あまりの年月を経ており、資料によっては所蔵者の代替わりや転居などがあり、許諾を得るのが困難な場合もある。このことは執筆者や掲載写真・図版の著作権者も同様で、連絡が取れない場合もあった。アーカイブス収録の許諾を得ることができたとしても、二次利用については所蔵者により対応が異な

ることになる。このことも次章以下で記そう。

このように県史編さんの成果からデジタルアーカイブスを形成することには制約や困難もあるのだが、一つの県について多岐にわたる時代・分野をまたがって網羅的にかつ自由に言葉を選んで調べることができることは、県史ひいては地域資料の利用価値を著しく増すものである。こうしたアーカイブスの形成について、以下に記していくこととしよう。

2. システムの概要

(1) 事業について

青森県史デジタルアーカイブスは、1－(2)でふれているように青森県史の編さんが終了する間際の2017年度に、終了後に成果の普及・活用を図ることを目的として、講演会の開催、普及用の冊子製作とともに計画されたが、政策経費として青森県史デジタルアーカイブス事業だけが認められた。

事業費は、県立博物館（青森県立郷土館）が製作していたデータベース等の整備費を参考に予算化され、その中には収集した史資料のデジタル化の委託料も含まれた。

デジタルアーカイブスシステム構築の委託の契約方法は、プロポーザル方式を採用し、インフォコム株式会社に委託した。同社のパッケージプランであるInfoLibを基本とし、それに地図検索機能等を追加した。

現在の運用は、経常経費である。

(2) 想定した対象

利用者として想定した対象は、20年以上にわたり県内全域で収集した資料を還元するという意味で、小学生から高校生も含めた一般県民及び研究者である。

(3) 仕様

予算規模と想定する対象より、以下の①～④の仕様とした。

① 5本のデータベース

アーカイブスを構成するデータベースは予算的にも5本が限度であったので、1－(3)で書かれている部会と史資料の内容から次の5つにまとめ、割り振った。

ア 文化財・自然データベース

遺跡、建造物、彫刻等の文化財資料と、標本等自然資料をまとめた。その詳細については、後述の5で述べる。

イ 絵はがき・写真等データベース

写真資料・動画をまとめた。現段階では、近現代の絵はがきと写真、民俗の再現した食事等を掲載しているが、将来的には県史で収集した民俗分野の動画も掲載する予定である。

ウ 古文書・文献史料データベース

古文書・写本・刊本、古地図・絵図類をまとめた。古文書は所蔵種別（個人・機関等）で分けて掲載している。また、近現代のポスター・チラシ・パンフレット、民俗分野で収集した版木、お札等も掲載している。

エ 県史テキストデータベース

青森県史全36巻のうち、通史編3巻と資料編「古代2出土文字資料」「近現代7日記」を除く31巻と、県史叢書14巻のテキストを公開している。「近現代7」を除いているのは個人のプライバシーにかかる記述を多く含むことによる。

テキストデータの詳細については、6で述べる。

オ 県史関係図書・論文データベース

県史編さんの過程で集めた図書資料や参考図書、各所から寄贈された図書など類の書誌情報を掲載している。

② デジタル展示室

2017年度当時は、まだデジタルアーカイブスというものが一般に知られていなかったことから、一般利用者がデジタルアーカイブスに親しんでもらうきっかけとなることを目的として作成した。

ここでは、「青森県の山・川・湖」など設定したテーマに沿って、各データベースからデータを選び、その画像と解説文を掲載し、「詳細を見る」から、元のデータを見ることができるようにしている。

③ 検索機能

ア 各データベース内の検索機能

各データベースの検索機能としては、検索した

い語句を入力する文字検索の他、県史の書名など対象とする項目を指定して検索語句を入力する詳細検索、及び検索結果にかかる位置情報を地図上に表示する地図検索を設けた。

また、資料の分野や種別等のカテゴリーによって表示される画面と、青森県の特徴である「津軽」・「南部」・「下北」の各地域別にデータが表示される画面を作成した。

イ トップ画面の検索機能

文字や単語等のキーワードを入力し、5本のデータベースを横断して検索する機能と、文字や単語等のキーワードを入力し、地図上の位置情報およびデータを見ることができる地図検索機能を作成した。

④ データベース間の相互リンク

5本のデータベースは、収集した史資料、県史を編さんする際に参考とした図書類、県史テキストの3つに分類することができ、情報が重複しているものがあることから、相互リンクの機能を付した。

例えば、史資料を検索したら、関連する県史テキストをみることができ、逆に県史テキストから基になる史資料の情報をみることができ、さらに参考文献の書誌情報を見ることができるようになっていく。

(4) データの保管

サーバーは庁内には置かず、2TBの容量のクラウドにデータやコンテンツを収めている。容量は各分野で掲載を予定する画像・動画の量から算定した。

3. オープンデータ化

青森県史デジタルアーカイブスを公開するにあたっては、画像及びテキストデータについて、基本的にオープンデータとすることとした。ここでは、オープンデータ化する際の観点とオープンデータ化するための手続きについて述べる。

(1) 「地域資源」という観点

県史編さん事業の終了が近づくなか、青森県史を編さんする過程で収集した県域にわたる史資料の利

県史編さん資料を「地域資源」としてデジタルアーカイブで活用する前提として、いつでも、だれもが簡単に利用できることが必要であると考え、メタデータを含むテキストと画像を基本的にオープンデータとして提供することを検討した。



「青森県史編さん資料等の利用及び公開等に関する要綱」を定め、その中でデジタルデータは、原則としてクリエイティブ・コモンズ・ライセンスにおける表示4.0国際(CC BY4.0)の下で提供することとした。また、利用条件がある個々のデータについては利用方法をデータに注記している。

公開後もしばらくは連携に至らなかったが、2020年8月のジャパンサーチ正式公開からちょうど1年後の2021年8月に連携を開始した。ジャパンサーチを通じた検索、データの閲覧利用の拡大により、青森県史デジタルアーカイブズの利用数は急増している。

(1) 項目設定

各資料の持つ属性を最大公約数的に項目化した。資料名＝遺跡名・文化財名・和名(標本)、産出地及び出土地＝化石の産出地、遺物等の出土地、所在地＝建造物の所在地、標本の所蔵機関場所とかなりまとめて設定した。資料の年代等についても、時代・年代・和暦に項目を分け、もっとも適した項目に入力するようにした。

その結果、項目数が他の4つのデータベースに比べて多くなった。その際、項目としてまとめるまでに至らないが、資料情報として無視することはできないもの、あるいは検索する際に必要な語句等があった。そこで、「キーワード」「備考」の項目を設け、補足すべき内容も含めて収めるようにした。

(2) 他機関との連携

文化財・自然データベースは、県外の機関が所蔵する青森県に由来する資料が多く含まれていた。これは、青森県内に所在する亀ヶ岡遺跡等の古くから著名な遺跡について、県外の大学や博物館により調査が行われたことに起因する。

また、自然分野でも、ブラキストンが収集した青森県が産出地の標本を所蔵する北海道大学やフォーリー神父が収集した植物標本を所蔵する京都大学がある。

これらの資料の一部は、所蔵する機関のデータベースで公開されているものがあり、資料を所蔵する機関と連携し、相互にリンクできるようにした。これにより、青森県史デジタルアーカイブスに掲載できなかった項目を、リンク先である所蔵機関のデータベースを見ることにより補完することができることとなった。

(3) 位置情報の公開

青森県史デジタルアーカイブスの特徴の一つとして、スマートフォンなどのモバイル機器によって青森県史のデータを手軽にどこでも利用できることがある。

青森県史の本の重さは1kg以上あり、1冊でも持ち歩くのは難しい。

モバイル機器に対応することで県内外の人々が、

興味がある文化財を現地で探索し、内容等を調べることができるように、遺跡や許諾を得た社寺等の建造物などの位置情報を公開している。

6. テキストデータのあり方

県史という書籍と編さん資料をもとにしたデジタルアーカイブスは、博物館などのそれとは異なり、モノとしての資料のみではなく、翻刻・活字化された内容をデジタルテキストとしてアーカイブスに収録し検索・閲覧利用に供することはその利便性を大きく増すものであるとともに、テキストデータベースはアーカイブス全体の中核とも位置づけられるものである。

県史各巻が印刷業者から納品される時には一部の巻を除きPDFファイルも合わせて納品するようにしていたが、納品されたPDFファイルは後述する理由もあって全文を検索して結果を一括表示するなどの用途に適するものではなく、著作権にも関わる画像や図版を含みそのままアーカイブスで表示させるのも不適切である場合もあった。検索及び検索結果の利用を容易にする観点からもアーカイブスに含めるのはPDFではなくデジタルテキストであることとし、PDFに埋め込まれたテキストデータをもとにして、時間と労力は要するが、デジタルテキストの作成を進め、それを公開することとした。

ただ、県史で扱われるテキストには、執筆された解説文の類もあれば、史料の翻刻もある。翻刻史料の中にも古文書・写本や手紙・日記といった手書きの類もあれば、既に活字化して出版されたそれらをもとにする場合もあり、新聞や雑誌のように原史料自体が活字である場合もある。県史が多くの時代・分野にわたるものであることから、掲載されたテキストも時代・分野により様相を異にするのは当然であり、ましてそれを検索・閲覧可能なデジタルテキスト化して一つのテキストデータベースに収めることにはいくつかクリアしなければならない問題がある。

(1) 文字コード

文献史料の翻刻はもちろんのこと、解説的な文章

に含まれる専門的な学術用語にも、難字や異体字など文字の問題が伴う。県史編さん当初から編集・執筆にはパソコンのワープロソフトを使用するのが主流になっていたが、まだ使用できる文字はJIS第一・第二水準を基本としてそれ以外の文字は外字を作成するのに限られていた。印刷料金の積算に外字作成数も含まれるような時代であった。専門部会によっても積極的に印刷業者に外字を作成させてできる限り原史料に忠実に活字化して掲載する場合もあれば、常用漢字への置き換えを原則とする専門部会もあり、このことは各専門部会それぞれが立てた方針によるものである。県史全体として調整を図ることはしなかった。20年を超える編集の終末期に至るまでの間にパソコンでの編集・執筆に使用できる文字も飛躍的に拡大することになったが、50冊の「県史」「県史叢書」各巻ごとに文字のあり方がまちまちとも言えるものをテキストデータベース化するのには困難が伴う。

原則としては書籍の用字をそのままにデジタル化するが、書籍編集時につくられた外字は、現在では文字コードが当てられるようになった字形があれば、いまだ含まれないものもあり、後者をどのようにデジタルテキストで表現するかは、結局アーカイブス作成時に各巻を担当したデジタルテキスト入力者の工夫・判断にゆだねられるところとなってしまう。字形を示すのに画像を添付する場合や、文字を解字（分解）して部首名を列挙して表す場合、文字コードが当てられている字形の近い異体字に置き換える場合など、文字・字形それぞれ判断せざるを得ないのが実情であり、対応を統一することは難しいところである。

また、デジタルテキスト入力作成時だけでなく、検索利用時における検索語句入力や結果表示に関わる問題点もある。先述のとおり、現在では県史編さん時に比べパソコンで利用できる文字が飛躍的に拡大していて、これは unicode の普及によるところが大きい。現在ではインターネットで使用される文字コードは unicode (UTF-8) が標準となっていると

いって良からう。ただ、青森県史デジタルアーカイブスでも一時、unicode の追加漢字面（コードが5桁に及ぶ領域）に当てられている文字が文字化けしたり、消えたりする現象が発生したことがあった。入力やデータのアップロードに用いるソフト・ファイル形式に起因する障害であって、現在は解消しており、検索結果の表示に問題はない。

しかし、県史の書籍に載せられていた難字・異体字をデジタルテキストの中に表示するに当たって、unicode におけるこのような領域の文字を用いて表現できるのであればそれを使用するのはあるべき姿勢ではあるのだが、そもそも利用者がそのような文字を入力して検索しようとする可能性は著しく低く、ごく一般的な文字で検索するのが普通であろう。もし史料による用字・字形の相違を研究しようとする者であれば、それは原史料あるいはその画像に遡って調べるべきことであって、デジタルアーカイブスとくにテキストデータベースは参考程度にしかならず今のところそうした用途には適さない。また、パソコンで使用しているフォントの中には unicode (UTF-8) に対応しているフォントであっても先に述べた追加漢字面など使用頻度の低い領域には文字を当てていないフォントもある。利用者にとっては検索語句入力時点で壁に阻まれることとなるが、アーカイブス運営の側では対応しがたい。

こうした状況への有効な対応策としては、シソーラスの作成が第一に挙げられよう。既存のデジタルアーカイブスの中にはシソーラスを含めたものも多くある。青森県史デジタルアーカイブスもシステムに辞書登録機能を有しており、語句 A を検索すれば語句 B も検索結果に含まれるように登録することは可能であるのだが、デジタルテキスト入力時の異体字や同義語に対する知見や作業量の制約、現在公開しているデジタルテキストを通して確認することも難しく、いまのところシソーラスの作成には至っていない。

(2) 割り付け

県史に掲載したテキストの内、解説原稿や近代以

降の史料にはほとんど見られないが、近世以前の史料には割り付けの問題がある。これはデジタルテキストの表示に当たって書籍の割り付けをできる限り再現したいということを、検索利用という目的のもとでどの程度かなえるかという問題につながる。

まず漢文体史料に付された返り点・訓点。これは原史料にはなく、県史に先行する翻刻や、県史編集の際に付された場合、典拠とした古文書や写本にもともと付されていた場合とに分けられるだろう。結論を記すと原則としてこれはどの場合であれ削除して白文の形に整理した。文中にまじると検索の妨げになるとともに、検索結果として表示されたデジタルテキストが訳のわからない読みづらいものになるからである。

訓点の中でも送り仮名は少々問題がある。漢文体史料において小書きしている送り仮名は、削除すれば事足りるが、漢字仮名交じり文の史料の場合には簡単にはいかない。もちろん通常の漢字仮名交じり文では「テニヲハ」や送り仮名を残さなければならぬことは言うまでもない。しかし史料によっては文の一部だけ漢文体をまじえていて返り点が付されていたり、おまけに送り仮名まであったり、また「也」「矣」「於」といった助字を記しているながら、その読みとなる送り仮名を二重に記しているものがあつたり、こうした返り点や送り仮名は原則として削除するが、文として形の整わないものとなってしまうこともあり、結局はそれぞれの文に応じた判断をせざるを得ない。

このほか、漢文体であれ、漢字仮名交じり文であれ、返り点、送り仮名以外の割り付けの問題としては、ルビ、傍書、小書き、割書などが挙げられる。青森県史デジタルアーカイブスでデジタルテキスト化するに当たってはこれらは、「<rubi ……><bousho ……><kogaki ……><wari ……>」といった書き方で該当箇所を括って文中に挟むようにした。htmlのタグのような書き方であるが、そのようなたいしたものではなく、筆者(古川)にもそうした知識は不足している。なんとか史料のものと割り付けをプレー

ンなデジタルテキストに表現できないか、誰が入力者でも簡単な決まり事で処理できないか、さまざまなケースに応用できる汎用性のある処理方法とならないか、という苦し紛れのやり方である。実際にマークアップ言語によってはなにかしら処理できることなのかもしれないが、何人かが分担して「県史」「県史叢書」のテキスト入力に当たっていたこと、デジタルテキストを登録するデータベース項目の設定、検索操作と検索結果の表示が安定して行えるかななどを考慮すると、タグまがいのものをまじえたテキストとするのが作業を進める上で無難であった。

また、検索利用に供するデジタルテキストであることを第一に考えると、単語、熟語といった単位が細切れになってしまうことは絶対に避けなければならない。このことは実際には原史料の割り付けを尊重することよりもはるかに重要なことである。原史料における割書の改行などは必ずしも反映させる必要はない。意味がある改行ならばスペースでも挟むのが適当である。傍書やルビを挟む位置も単語の文字列の真ん中などに置いてはならない。デジタルテキストの入力・作成に当たっては、印刷業者が納品してきたPDFファイルに埋め込まれたテキストを切り出して利用することがほとんどであったが、その場合、往々にしてルビや傍書の位置には印刷業者の用いたDTPソフトでの入力・校正作業の痕跡が残ってしまっており、印刷のためにきれいに整形されている仕上がりとはかけ離れたものである場合も起こりうる。そうした場合に検索利用に供することを前提とした校正が行われなければならないことは言を俟たない。

(3) 人名等の表記・注記

これは上記2点とは異なり、デジタルテキストの形態の問題ではない。テキストを検索に供する上で利用者への配慮の問題である。

古代史の分野でいえば、日本史の教科書であれば人名に「姓(かばね)」を付すことはない。「坂上宿禰田村麻呂」「文室真人綿麻呂」「源朝臣義経」ではなく「坂上田村麻呂」「文室綿麻呂」「源義経」と記

される。しかし、史料では姓を付して記される場合が多いし、文脈によっては「氏（ウヂ）姓」が略される場合、名も記さず官職名や通称で記される場合もある。これはどの時代でも同様であろう。しかし、検索利用者はいちいち姓を付して検索しようとはしないだろうし、官職名で特定の個人を検索しようとする者も少ないだろう。しかし、それでは調べたい人物についてデータベースに収録している史料を検索結果として網羅することはできない。史料のテキストのみが検索対象となっていればやむを得ないところである。これはもはやデジタルテキストの入力・作成者、つまりデータの作成者が、どれだけ自らの知識や県史の書籍本体に載っている注記をもとに検索利用者の便宜を図るか、ひいてはデータベースの利用価値を高めることができるかという問題である。県史テキストデータベースにはテキスト本文とは別に「キーワード」項目がある。筆者はこれに人名の別表記、姓をはずしたもの、あるいはその逆をいれるなどしたが、なかなか県史各巻の他分野のテキスト全体について徹底できるものでもなかった。「キーワード」項目は通常本文中の重要語句を掲げる項目であろうが、筆者の場合は本文中に記されている語句はほとんど挙げずに本文にない別表記ばかり検索対象とするために並べたデータも多々ある。県史のテキストデータベースとしては不統一なところでもあるのだが、時代によっては官職名などについて安易に人物比定できない場合もある。こうした処理において古代史の分野は比較的取り組みやすかったからできたとは言えよう。

人名表記の他にも地名などで用字が史料により異なる場合、やはりキーワード項目に他の史料での別表記を掲げておくことや、「陸奥出羽両国」といった本文の場合「陸奥国」「出羽国」とキーワードとしていずれでも検索対象となるようにするなどの配慮はした。人名や地名に限らず、検索利用され検索結果の価値を高めるため別の表記を何らかの形で掲げる工夫は必要であるが、それはデジタルテキスト作成者個々の認識と判断によるもので、あまり一律に目

標として設定できるものでもないし、行き過ぎて詳細になりすぎるのも作業が滞ることになるだろう。

本章全体を通じて記したことは、文字データであれ、割り付けの表現であれ、人名等の表記であれ、さまざまな問題はあっても、まず大切なのはデジタルテキストは検索利用に供するものであること、そのために検索者の便宜を図ること、そして検索結果は活用しやすいものであること、それらのためにどのように考慮、工夫をして問題をクリアするかということである。ここに記した問題に対してもいろいろなアプローチの仕方はあるだろうし、この他にもいろいろな問題はあるだろうし、技術的に解決できることもあるだろう。これはこうしたテキストデータベースを今後作成する方々に期待するところである。

7. 今後の課題及び方向性

最後に今後の課題等について記しておく。

まず、本アーカイブスは県史および編さん資料に基づくものであるから、数的・量的に有限のものではあるのだが、現時点でも未整理・未収録のデータは残っており、データ拡充の余地が残されている。特に県史に掲載した画像・図版でテキストに添えられていないものがある。前述のように民俗分野の調査成果である動画や音声資料も未収録のものが多い。

また、近年は多くの資料所蔵機関でアーカイブス、データベースが構築されてきているが、そのなかには県史の典拠としたデータが収録されているところも見られる。こうした他機関のアーカイブスについて二次利用条件によっては転載やリンクを張るなどすることで利用者の便宜を図ることも重要である。

さらに、現在、編さん資料の整理も進めているところであるが、アーカイブスのデータが編さん資料の配架情報と完全には結びついていない。これを一元化することにより将来、資料の利活用が円滑に行えるようになると考えている。

さて、今後の方向性として次のことを想定してい

る。県内には、青森県立図書館をはじめとした文化財に関するデジタルアーカイブスがすでに構築されている。現状、利用者はそれぞれのデータベースにアクセスし閲覧している。

一方で、2022年4月の博物館法の改正により、今後県内の博物館等でデジタルアーカイブスが構築され、増加することが予想される。

文化財を「地域資源」と捉えるならば、文化財に関するデジタルアーカイブスは、利用者にとって使いやすく、より多様な利用を生み出すものでなければならない。県内の文化財に関するデジタルアーカイブスをまとめたプラットフォームを構築し、ワンストップで利用することも一つの方法であると考えている。すでに長野県がこのようなサイトを構築しているが、今後の課題として検討する必要がある。

なお、この8月に青森県史デジタルアーカイブスは内閣府の知的財産戦略推進本部に置かれたデジタルアーカイブジャパン推進委員会実務者検討委員会から第1回のデジタルアーカイブジャパン・アワードを受賞した。永年にわたる県史編さんに尽力された方々、アーカイブス収録データの入力に携わった職員等の皆さん、デジタルアーカイブスについて御指導いただいた国立歴史民俗博物館の後藤真氏をはじめとする多くの方々に感謝するものである。

文責：古川＝1, 6, 7前半。伊藤＝2～5, 7後半

【参考文献】

- 1) 石川智志・渡辺一生『エリアケイパビリティ-地域資源活用のすすめ-』2015 総合地球環境学研究所