

人文情報学から中琉関係史資料へのアプローチ デジタル化資料の長期維持に向けた取り組みの一つとして

富田千夏*

要旨

収集した古文献資料の長期的な維持管理のなかで、資料を電子化することは資料保存の目的や利用者の利便性を高める意義もあり、近年多くの資料保存機関がデジタルアーカイブ事業を展開している。デジタルアーカイブを構築・運営していく上で多くの機関が直面している問題は「如何にして長期的に維持するか」であり、その難しさは現在利用ができない「沖縄の歴史情報研究会」の事例が物語っている。

本稿では、デジタル化された資料の長期維持に向けた取り組みとして、「人文学に情報学の技法や技術を応用する¹」学問である人文情報学 (Digital Humanities) の手法を取り入れ琉球大学附属図書館の「琉球・沖縄関係貴重資料デジタルアーカイブ」をIIIF (International Image Interoperability Framework) へ対応した事例や「沖縄の歴史情報研究会」にて作成された「琉球家譜」のデータを再利用する試みを紹介する。現在利用ができない状況である「沖縄の歴史情報研究会」テキストデータのうち、CD-ROM版に収録されていた「琉球家譜」のデータについては現在の環境に適したデータの形へ変換することで再度利活用が可能である。「琉球家譜」のデータファイルを現在の環境で利用可能なデータに変換した上でデファクトスタンダードであるTEI (Text Encoding Initiative) に適用する試行的な取り組みを通して、デジタル化された資料の長期的な維持について機関側と研究者が出来る事とは何かを検討し、今後のデジタルコンテンツのあり方について課題を共有したい。

キーワード：デジタルアーカイブ、琉球家譜、TEI、人文情報学、Digital Humanities

* 琉球大学附属図書館情報サービス課保存公開係長

¹ 後藤真 「人文情報学と歴史学」 『歴史情報学の教科書-歴史のデータが世界をひらく』 文学通信、2019、p22-22。

1.はじめに：琉球・沖縄関係資料のデジタル化の経緯と電子化の意義

1950年に首里城跡地に琉球大学が開学した際に「琉球大学文化センター」として開館²した附属図書館では、当時から積極的に琉球・沖縄研究に資する資料を収集し、1955年に受け入れた伊波普猷文庫をはじめ、島袋源七文庫（1957年）や仲原善忠文庫（1967年）等のコレクションを収集し、現在に至っている。

文書資料を収集・保存し、公開していく上で資料をデジタル化ししていくことは、資料の保存のためになるべく現物資料へのアクセスを減らす意味³や、利用者の閲覧の利便性が高まる等の意義があり、総合的な学術研究への寄与に期待⁴したものである。琉球大学附属図書館では、1997年に宮良殿内文庫のデジタル化とその公開に始まり、2001年には伊波普猷文庫、2003年に仲原善忠文庫等と20年以上にわたってデジタルアーカイブ事業を実践してきた⁵。沖縄県内に目を向けてみても、近年沖縄県立図書館の「貴重資料デジタル書庫⁶」や南城市が構築運営する「なんじょうデジタルアーカイブ⁷」、さらには沖縄県教育委員会による「琉球王国交流史・近代沖縄史料デジタルアーカイブ⁸」等が公開され、自治体によるデジタル化事業の展開が活発化している傾向にある。⁹

² 琉球大学附属図書館WEBサイト「沿革」を参照（<https://www.lib.u-ryukyu.ac.jp/about/history/>）。

³ 琉球大学附属図書館報『びぶりお』No.116（https://www.lib.u-ryukyu.ac.jp/lib_uploadfile/publication/biblio116.pdf）。

⁴ 琉球大学附属図書館報『びぶりお』No.140（https://www.lib.u-ryukyu.ac.jp/lib_uploadfile/publication/biblio140.pdf）。

⁵ 大谷周平、富田千夏「琉球大学附属図書館のデジタルアーカイブ事業」『沖縄県図書館協会誌』22、2019。

⁶ <https://www.library.pref.okinawa.jp/archive/index.html>。

⁷ <https://nanjo-archive.jp/>。

⁸ <https://ryuoki-archive.jp/>。

⁹ 戦後の沖縄における資料の収集や編纂そしてデジタル化の概略については、山田浩世、小野百合子「戦後沖縄における資料収集・編纂と近年のデジタルアーカイブの取り組み」（『歴史学研究』1024、2022、28-39頁）を参照のこと。

2.資料のデジタル化の課題：データベースの維持と長期保存の問題

デジタルアーカイブ等のコンテンツを構築・運営している機関の関係者にとって大きな関心事であり悩み事の一つに「システムを如何にして維持・管理するか」という点が挙げられる。数年毎に必要に迫られるシステム更新に加えて、更新後も定期的なメンテナンスや保守に関する経費、担当



図1 旧「琉球・沖縄関係貴重資料デジタルアーカイブ」の「ガラスビュー」画面

職員の人件費等のランニングコストを如何にして獲得していくかは大きな課題である。¹⁰

琉球大学附属図書館では、所蔵する貴重資料のデジタル化と公開については1997年から始めており文庫毎に個別のシステムで公開していた¹¹。2012年度に「琉球・沖縄関係貴重資料デジタルアーカイブ」を構築し、各文庫をまとめて一つのシステムとして公開することにした際には、それまでの画像のみの公開から、より使いやすさを重視し、解題（とその英訳）、翻刻、現代語訳（一部は英訳）の付与を行ってきた。また、ビューワーは史料画像と翻刻の情報を重ねて表示する機能（ガラスビュー¹²）を有しており、古文書学習への活用も期待したもので、商用のソフトウェア（Adobe社Flash）を使用していた。しかし、2017年7月に2020年末をもってFlash Playerのサポートを終了することが発表¹³されたことで、必ず直面するシステムの老朽化と共に対応を余儀なくされるこ

¹⁰ 当館にとっても、デジタルアーカイブに関連する論考で何度も言及している程度には常に付きまとう問題である。大谷周平、富田千夏「琉球大学附属図書館のデジタルアーカイブ事業」『沖縄県図書館協会誌』22、2019。富田千夏「琉球・沖縄関係貴重資料デジタルアーカイブについて」『琉球沖縄歴史』3、2021、122-127頁。

¹¹ 大谷周平、富田千夏「琉球大学附属図書館のデジタルアーカイブ事業」『沖縄県図書館協会誌』22、2019。

¹² バゼル 山本登紀子「越境する沖縄関係資料—阪巻・宝玲文庫資料の電子化共有プロジェクト経験を通して」『阪巻・宝玲文庫の世界 事業報告・研究報告会 報告書：The World of the Sakamaki/Hawley Collection』、2017、9-46頁

¹³ Adobe Blog「Flash とインタラクティブコンテンツの未来（原題：Flash & The Future of Interactive Content）」、2017年7月（<https://blog.adobe.com/jp/publish/2017/07/26/201707adobe-flash-update>）。

とになった。

一方で、「沖縄の歴史情報」は科研費という研究者個人と密接に結びついたプロジェクトによるデータベースの長期的な維持管理に関して、機関運営のデータベースとはまた異なる問題を有している。¹⁴

「沖縄の歴史情報研究会」とは科学研究費補助金（科研費）の特定領域研究（A）「沖縄の歴史情報」（研究番号：06208102, 研究代表者：岩崎宏之）に基づく研究会で、「本研究は、領域研究の成果を取りまとめて研究成果報告書を作成し、領域研究の成果である琉球・沖縄史と環東シナ海地域間交流史に関する各種歴史情報を、学界はもとより広くインターネット等を利用して一般に公開・利用に供することを課題とした。（研究概要¹⁵より）」とあるように、公開当時、Webサイト上では「歴代宝案」や「琉球家譜」「琉球王国評定所文書」といった琉球・沖縄研究における基礎資料のデータベースが公開され、近年まで活用されていた。プロジェクトの規模も大きく、その後の関連研究に大きな影響を与えており、「歴史資料のインターネットを通じた画像公開やフルテキストデータによる資料検索、基礎的なデジタルデータの整備を進めた点で画期的であった¹⁶」とされる。プロジェクトが終了した後も、データベースは維持されてきたが、数年前¹⁷より利用ができない状況が続いている。ただし、全てのデータにアクセスが出来なくなったわけではなく、筑波大学附属図書館では、一部のデータベースが¹⁸利用可能であるほか、研究成果報告書としてCD-ROM版¹⁹（全10巻）が発行されており、どのようなデータが収集・公開されていたのかを確認することは可能である。しかし中には「琉球王国評定所文書」や「歴代宝案」のように、Web上では全文が検索対象となっていたものの、CD-ROM版への収録は一部にとどまっているデータ²⁰もあり、従前の環境を補完するものとは言えない。

¹⁴ 山田浩世、小野百合子「戦後沖縄における資料収集・編纂と近年のデジタルアーカイブの取り組み」『歴史学研究』1024, 2022, 28-39頁。

¹⁵ 特定領域研究(A)「沖縄の歴史情報」については、科学研究費助成事業データベースの該当ページを参照（<https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-06208102/>）

¹⁶ 山田浩世、小野百合子「戦後沖縄における資料収集・編纂と近年のデジタルアーカイブの取り組み」『歴史学研究』1024, 2022, 28-39頁。

¹⁷ 山田浩世、小野百合子「戦後沖縄における資料収集・編纂と近年のデジタルアーカイブの取り組み」『歴史学研究』1024, 2022, 28-39頁によれば、2010年代までは利用可能な状況にあった。

¹⁸ 筑波大学附属図書館「沖縄歴史文献データベース検索」（<https://www.tulips.tsukuba.ac.jp/lib/dbinfo/okinawa-db>）。

¹⁹ 岩崎宏之『文部省科学研究費補助金重点領域研究「沖縄の歴史情報研究」（領域番号：110）（平成6年度-平成9年度）CD-ROM版研究成果報告書』、岩崎宏之、1998。

²⁰ 「琉球王国評定所文書」については目録のみ、「歴代宝案」についてはテキストデータの一部がCD-ROMに収録されている。

前述のとおり、デジタルアーカイブ等のシステムの構築やその維持にはある程度の予算が必要となってくるうえ、永続的なものではなく数年単位でシステム更新の必要に迫られるものである。また、Adobe社Flashの例に代表されるように、特定企業の提供するソフトウェアに準拠した公開システムを採用した場合、企業のサポートが終了した際に必要な機能が失われることも考慮しなければならない。

長期的にデジタルアーカイブを維持するためにはどのようなシステムを導入すべきか、を館内で検討する中でデファクトスタンダードであるIIIFとTEIに着目することになったのである。

3.IIIFとTEI：データの標準化技術

IIIF (International Image Interoperability Framework) ²¹は、WEB上においてデジタルコンテンツを効果的かつ効率的に公開し、共有するための仕組みであり、国内外の多くのデジタルアーカイブシステムで導入され事実上の標準規格（デファクトスタンダード）となっている²²。沖縄県内では前述したとおり、琉球大学附属図書館が2019年に導入したほか、沖縄県立図書館「貴重資料デジタル書庫」でも導入されている。IIIFの特徴の一つとして、IIIFの規格にそって画像の公開をしておけばユーザー側は使い慣れたIIIF対応のビューワーを使用することができ、図2に示すように複数の資料画像を並列して表示することもできる²³。

TEI (Text Encoding Initiative) ²⁴は、デジタルのテキスト資料をよりよく作成し、共有するための仕組みとして1987年にガイドラインが策定された²⁵。欧米における人文学向けの電子化資料の構造化の手法としてデファクトスタンダードとなっているものである²⁶。日本では2016年にTEI協会東アジア/日本語分科会が設置され、国立国語研究所や東京大学史料編纂所等において日本語資料へのTEI準拠の取り組みが進められている²⁷。

²¹ <https://iiif.io>。

²² IIIFについては、永崎研宣「CA1989 - 動向レビュー：IIIFの概要と主要APIバージョン3.0の公開」（カレントアウェアネス：<https://current.ndl.go.jp/ca1989>、2020）、および永崎研宣「デジタルアーカイブシステムにおける標準化技術」（『情報の科学と技術』71(4)、doi:https://doi.org/10.18919/jkg.71.4_165、2021、165-170頁）を参考のこと。

²³ 永崎研宣「デジタル文化資料の国際化に向けて：IIIFとTEI」『情報の科学と技術』2017、61-66頁。

²⁴ <https://tei-c.org/>。

²⁵ 一般財団法人人文情報学研究所[監修]『人文学のためのテキストデータ構築入門』文学通信、2022、36頁。

²⁶ 永崎研宣「デジタル文化資料の国際化に向けて：IIIFとTEI」『情報の科学と技術』2017、61-66頁。

²⁷ 永崎研宣「デジタル文化資料の国際化に向けて：IIIFとTEI」『情報の科学と技術』2017、61-66頁。

2020年代に入り、日本国内においても「渋沢栄一ダイアリー²⁸」「デジタル源氏物語²⁹」等のTEIガイドラインを採用したデジタル化プロジェクトがいくつか進行しており、成果が公表されている³⁰。



図2 沖縄県立図書館蔵『歴代宝案』（写本・影印本）の比較

2019年度に「琉球・沖縄関係貴重資料デジタルアーカイブ」を更新するにあたって、できるだけ安定した継続が望める仕組みを導入することを検討した。結果、デファクトスタンダードであるIIIFは国内外の機関が導入しており他機関との連携も容易であることから、今後の労力を抑制しさらに資料の可視性を高める等の利便性を高めることにつながると考え³¹採用することになった。また、IIIFと共に電子化資料の標準化技術として国内で徐々に広まりつつあるTEIはテキストデータの構造や固有表現をマークアップすることで、専門家にとって使いやすいデータとすることを目的としており³²、詳細な書誌事項を含むメタデータを付与することもできる。将来的に「琉球・沖縄関係貴重資料デジタルアーカイブ」の翻刻データへの適用の可能性も考慮し、職員の有志でTEI協会東アジア/日本語分科会主催の勉強会に参加する中で、TEIの活用への試行として中琉関係史研究において重要な「琉球家譜」のテキストデータのTEI準拠に着手することにした。

²⁸ 渋沢栄一記念財団によるプロジェクト (<https://shibusawa-dlab.github.io/app1/>)。

²⁹ 有志団体による (<https://genji.dl.itc.u-tokyo.ac.jp/>)。

³⁰ 2020年代のTEI関連のプロジェクトは、一般財団法人 人文情報学研究所[監修]『人文学のためのテキストデータ構築入門』（文学通信、2022、43-44頁）を参照のこと。

³¹ 大谷周平、富田千夏「琉球大学附属図書館のデジタルアーカイブ事業」『沖縄県図書館協会誌』22、2019。

³² 永崎研宣「デジタルアーカイブシステムにおける標準化技術」『情報の科学と技術』71(4)、doi: https://doi.org/10.18919/jkg.71.4_165、2021、165-170頁。

権・肖像権) 上の制約があるもの以外については資料提供者側の都合による条件をつけず自由な利用を可能とするようにしている³⁸。

5. 「琉球家譜」のTEI準拠：「沖縄の歴史情報研究会」のデータを再利用する

本稿では、前述CD-ROM版に収録されていた『那覇市史』の家譜資料編のテキストデータを現在の環境において利用可能なデータ形式に変換すると共に、試行として一部のデータをTEIに準拠したマークアップを行い、機械可読可能なXMLファイルとして作成し、研究上の利活用の可能性について提案を行うことを目的としている。

5-1. CD-ROMに格納されたデータを変換し、XMLファイルに組み込む

琉球家譜のデータはCD-ROM版第5巻に収録されており、それぞれ『那覇市史』資料編の久米系家譜(1-6)³⁹、首里系家譜(1-7)⁴⁰、泊・那覇系家譜(1-8)⁴¹として刊行されたものに依拠したテキストデータ⁴²を有するHTMLファイルであった。ページ毎にファイルが分かれていたため、全てのページの文字コードをShift-JISからUTF-8へ変換した後に拡張子をhtmlファイルからtxtに変換し、『那覇市史』の各冊単位で1ファイルとなるように結合処理した。さらにXMLファイルにテキストを取り込み、不要なHTMLタグを除去するなどのクリーニングを行った。その際、HTMLの各ファイルのtitleタグは、『那覇市史』におけるページ番号と一致するため、TEIガイドライン(P5)の<pb> (page beginning) エlement⁴³に変換した。

³⁸ 画像の二次利用についての考え方については、富田千夏「琉球・沖縄関係貴重資料デジタルアーカイブについて」(『琉球沖縄歴史』3、2021、122-127頁)を参照のこと。

³⁹ 那覇市企画部市史編集室編『那覇市史 資料篇 第1巻 6(家譜資料 2 久米系)』那覇市企画部市史編集室、1980年。

⁴⁰ 那覇市企画部市史編集室編『那覇市史 資料篇 第1巻の7(家譜資料 3 首里系)』那覇市企画部市史編集室、1982年。

⁴¹ 那覇市企画部市史編集室編『那覇市史 資料篇 第1巻 8(家譜資料 4 那覇・泊系)』那覇市企画部市史編集室、1983年。

⁴² 「沖縄の歴史情報研究」における琉球家譜のテキストデータ作成に関しては、豊見山和行「琉球家譜」テキストデータベースの作成」(『重点領域研究「沖縄の歴史情報研究」：総括班研究成果報告書』、1998、235-238頁)を参照のこと。

⁴³ <https://www.tei-c.org/release/doc/tei-p5-doc/ja/html/ref-pb.html>。

5-2. 画像処理されていた文字情報をUnicodeに変換する

ある程度、不要なタグを除去した段階で、当時テキストデータとして入力できない文字（外字）の処理を行った。当時テキストデータの作成を担当していた豊見山和行の報告によれば、テキスト上で入力できない文字については、☆印で前後をかこみ、☆艸☆（「葦」）のように表示したとある⁴⁴。この情報をHTMLファイル上では、GIF形式の画像ファイルとして処理しており、ファイル名は「今後の展開も考慮し⁴⁵」Unicode（JIS X 0221）に対応させ、JIS X 0221に存在しない漢字については、画像処理ソフトで作成した画像ファイルとして処置されていた⁴⁶。HTMLファイルには、テキストに組込まれた画像ファイルは、例えば「丰」であればUnicode「4e30」をファイル名としてとimgタグとして処理されていたため、Unicodeへの一括変換が可能と判断した。imgタグとUnicodeへの置換リストをExcelにて作成し、それを基にPythonのプログラムコードを使用して一括変換を行った⁴⁷。当時Unicodeにない文字の画像ファイルについては、今後の作業上で確認しだい都度変換することになっている。

5-3. 「琉球家譜」データのTEIマークアップ

1. 書誌事項等、メタデータのマークアップ

上記までの作業を経て、「琉球家譜」のテキストへのTEIによるマークアップ作業に着手することができる。試行として、『那覇市史 第1巻 6 (家譜資料 2 久米系)』に所収されている「王姓家譜 (小渡家)⁴⁸」のデータを対象とした。

TEIにおいて、ヘッダー (teiHeader) は極めて重要な要素とされ、ガイドラインにおいては必須の項目となっており、元の資料との関係性や作業者の情報等をファイルの中に記述しておくことで、ファイル単体でもそのテキストデータの概要について共有することができる⁴⁹。ヘッダーの要素である<fileDesc> (file description)⁵⁰のなかに、電子ファ

⁴⁴ 豊見山和行「「琉球家譜」テキストデータベースの作成」『重点領域研究「沖縄の歴史情報研究」：総括班研究成果報告書』、1998、235-238頁。

⁴⁵ 桶谷猪久夫「「琉球家譜」の検索システム」『重点領域研究「沖縄の歴史情報研究」：総括班研究成果報告書』1998、239-247頁。

⁴⁶ 桶谷猪久夫「「琉球家譜」の検索システム」『重点領域研究「沖縄の歴史情報研究」：総括班研究成果報告書』1998、239-247頁。

⁴⁷ Pythonのソースコードの作成と変換作業は琉球大学附属図書館の大谷周平氏にご協力いただいた。

⁴⁸ 那覇市企画部市史編集室編『那覇市史 資料篇 第1巻 6 (家譜資料 2 久米系)』那覇市企画部市史編集室、1980、1-17頁。

⁴⁹ 永崎研宣「歴史データのさまざまな応用 -Text Encoding Initiative の現在-」『歴史情報学の教科書 歴史のデータが世界をひらく』文学通信、2019、131-154頁。

⁵⁰ <https://tei-c.org/release/doc/tei-p5-doc/ja/html/ref-fileDesc.html>。

イルに関する書誌事項や、元の書籍に関する書誌事項等を記録することができるので、テキストデータが作成された「沖縄の歴史情報研究」に関する情報や、『那覇市史』の書誌事項、マークアップの作業者の情報について記述することにした。

```

<?xml-model href="../../../tei_all_ja.rnc" type="application/relax-ng-compact-syntax"?>
▼<TEI xmlns="http://www.tei-c.org/ns/1.0">
  ▼<teiHeader>
    ▼<fileDesc>
      ▼<titleStmt>
        <title>久米村家譜</title>
        ▼<respStmt>
          <resp>transcription</resp>
          <orgName>沖縄の歴史情報研究会</orgName>
        </respStmt>
      </titleStmt>
      ▼<editionStmt>
        <edition>TEI Version</edition>
        ▼<respStmt>
          <persName>Chinatsu Tomita</persName>
          <resp>TEI encoding</resp>
        </respStmt>
      </editionStmt>
      ▼<publicationStmt>
        <distributor>沖縄の歴史情報研究会</distributor>
        <idno type="研究課題/領域番号">06208102 </idno>
        <ref target="https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-06208102/"> 沖縄の歴史情報研究</ref>
      </publicationStmt>
      ▼<sourceDesc>
        ▼<biblStruct type="book">
          ▼<monogr>
            <editor>那覇市企画部市史編集室</editor>
            <title>那覇市史 資料編第1巻6 家譜資料2</title>
            <idno type="国立国会図書館書誌ID">000001469074</idno>
            <idno type="DOI">10.11501/9773364</idno>
            ▼<imprint>
              <publisher>那覇市企画部市史編集室</publisher>
              <pubPlace>沖縄県那覇市</pubPlace>
              <date when="1980">1980-3-25</date>
            </imprint>
          </monogr>
        </biblStruct>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  ▼<listPlace>

```

図4 TEIヘッダの冒頭部分

また、file description の中には<sourceDesc>(source description)という電子テキストのもと資料について記述をするエレメントがあるが⁵¹、ここには、テキスト中に登場する個人のリストである<listPerson> (list of persons)⁵²や、場所のリスト<listPlace> (list of places)⁵³を作成する項目がある。場所のリストには位置情報を記述することもできるので、後の機械分析に活用できる情報として入力する(図5および6)。

⁵¹ <https://www.tei-c.org/release/doc/tei-p5-doc/ja/html/ref-sourceDesc.html>.

⁵² <https://www.tei-c.org/release/doc/tei-p5-doc/ja/html/ref-listPerson.html>.

⁵³ <https://www.tei-c.org/release/doc/tei-p5-doc/ja/html/ref-listPlace.html> なお、<listPerson>については、ヘッダのほかに<text>の中の<back>に記載することも可能である(一般財団法人人文情報学研究所[監修]『人文学のためのテキストデータ構築入門』文学通信、2022、175頁)。

```

</listPlace>
</listPerson>
  <person xml:id="王可法_17-2020">
    <persName>王可法</persName>
  </person>
  <person xml:id="王裕之_17-2120_05">
    <persName>王裕之</persName>
  </person>
  <person xml:id="王三錫_17-2120_06">
    <persName>王三錫</persName>
  </person>
  <person xml:id="王三德_17-2120_08">
    <persName>王三德</persName>
  </person>

```

図5 <listPerson>人名のリスト表示

```

</listPlace>
  <place xml:id="福州琉球館">
    <placeName>福州琉球館</placeName>
    <location>
      <geo resp="tomita" source="https://geo.gl/maps/yY7RFieNpAyMfMfB6">
        26.067548187175556, 119.31713986572119 </geo>
      </location>
    </place>
  <place xml:id="八重山">
    <placeName>石垣島</placeName>
    <location>
      <geo resp="tomita" source="https://geo.gl/maps/w4690bb6wBxeuDuPA">
        24.339327107585798, 124.1489963484789 </geo>
      </location>
    </place>
  <place xml:id="宮古島">
    <placeName>宮古島</placeName>
    <location>
      <geo resp="tomita" source="https://geo.gl/maps/mVfuEdDQPhaPgwLW8">
        24.8068259977504, 125.27501021798428 </geo>
      </location>
    </place>

```

図6 <listPlace>地名のリスト表示

TEIヘッダの最後には<encodingDesc> (encoding description)エレメント⁵⁴を付与し、<projectDesc>(project description)⁵⁵として、この作業の経緯を明記しておくとともに入手元のCD-ROM版への書誌事項へのリンクを参照情報とすることにした⁵⁶。また、各家譜について、那覇市が発行している『氏集』の番号を参照情報とするために同書の書誌事項についても明記した(図7)。

```

</listPerson>
</sourceDesc>
</fileDesc>
<encodingDesc>
  <projectDesc>
    <p>このファイルは、文部省科学研究費補助金重点領域研究「沖縄の歴史情報研究」(領域番号:110)(平成6年度-平成9年度)CD-ROM版研究成果報告書第5巻所収のデータをTEIに準拠させたものです。</p>
    <ref target="https://ci.nii.ac.jp/ncid/BA4021040X">沖縄の歴史情報(CD-ROM)</ref>
  </projectDesc>
  <refsDecl>
    <p>『氏集番号』は那覇市発行『氏集』に掲載されている家譜目録に付与された固有IDをさす</p>
  </refsDecl>
  <ab>
    <bibl>
      <title>氏集：首里・那覇 第4版</title>
      <editor>那覇市市民文化部歴史資料室</editor>
      <publisher>那覇市市民文化部歴史資料室</publisher>
      <date when="2004">2004.3</date>
      <idno type="国立国会図書館書誌ID">000007374366</idno>
    </bibl>
  </ab>
</encodingDesc>
<profileDesc>
  <langUsage>
    <language ident="zh-JP">中国語</language>

```

図7 <encodingDesc>の情報

54 <https://www.tei-c.org/Vault/P5/4.3.0/doc/tei-p5-doc/ja/html/ref-encodingDesc.html>.

55 <https://www.tei-c.org/Vault/P5/3.4.0/doc/tei-p5-doc/ja/html/ref-projectDesc.html>.

56 『人文学のためのテキストデータ構築入門』170頁に、「このエレメントは、この文書を他の人がコンピュータで処理・分析する際の拠り所となるものですので、可能な限り詳細に記述しておくことが重要です(一般財団法人 人文情報学研究所[監修]『人文学のためのテキストデータ構築入門』文学通信、2022)」とあり、今後情報を追加していく予定である。

2. 基本的な久米村家譜の文書構造をマークアップする

TEIヘッダの情報の次は「琉球家譜」の本文テキストへのマークアップに進む。久米村家譜の基本的な文書構造を『王姓家譜（小渡家）』を例として図式化すると図8のようになる。

まず、世系図があり、各個人の履歴となる「紀録」になるが、紀録部分は、生没年や父母や妻子等の情報部分の他、官爵の情報、職歴に該当する「助庸」、知行地に関する情報である「采地」、子女の婚姻に関する情報である「婚嫁」に分かれている。

この各個人の「紀録」と内部の情報の構造を以下のようにマークアップする。

- 各個人の「紀録」毎に<div> (text division)エレメント⁵⁷を使用して分割する。
- 「紀録」の内部構造は<ab>(anonymous block)エレメント⁵⁸を使用し、「官爵」「助庸」等の項目毎にアトリビュート（属性：@type）を付与することにした。

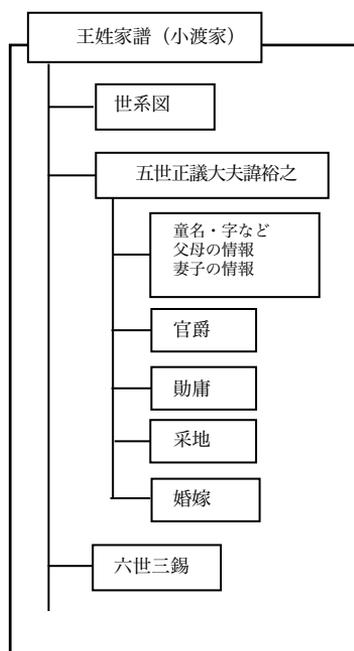


図8 久米村家譜の基本的な構造図

3. 世系図（家系図）をマークアップする

「沖縄の歴史情報研究」では、「琉球家譜」のテキストデータ作成に際し、文字情報のデータ化を基本としたために、『那覇市史』の各家譜の冒頭に掲載されている世系図（家系図）については省略し、WEB上で画像の掲載として提供していた⁵⁹。

TEIでのマークアップにあたり、『那覇市史』に掲載されている情報はなるべく入れたいため、これらの世系図の情報もマークアップしたテキストとして入力することにした。

TEIガイドラインに掲載されている「19 Graphs, Networks, and Trees⁶⁰」は、家系図や、写本系統図のようなツリー構造のデータをマークアップする際に使用できる記述方法

⁵⁷ <https://www.tei-c.org/Vault/P5/2.9.1/doc/tei-p5-doc/ja/html/ref-div.html>。

⁵⁸ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ab.html>。

⁵⁹ 豊見山和行「『琉球家譜』テキストデータベースの作成」『重点領域研究「沖縄の歴史情報研究」：総括班研究成果報告書』、1998、235-238頁。

⁶⁰ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/GD.html>。

である。具体的に「19.1.2 Family Trees⁶¹」として家系図をマークアップする方法もあるが、これは父母および子とその婚姻等の関係性を詳述することが想定されているため、父と子の関係しか記述されていない琉球家譜の世系図⁶²に適用するには沿わないと考えた。今回は「19.2 Trees⁶³」を参照して、世系図に記載された通りの情報をマークアップすることにして、図9の世系図の情報を図10のようなデータとして作成した。

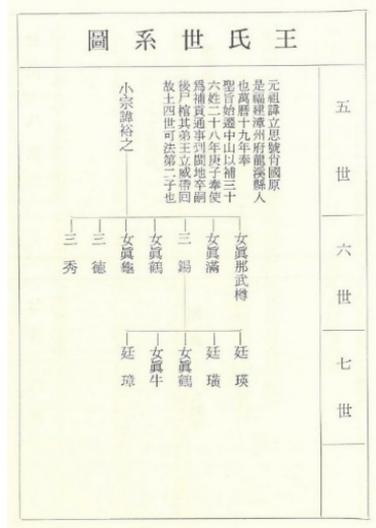


図9 王世家譜 (小渡家) の世系図 (部分) 『那覇市史 資料篇 第1巻 6 (家譜資料 2 久米系)』 (1980) より

```

</teiHeader>
▼<text type="genealogy">
  ▼<body>
    <pb n="1"/>
    ▼<div n="17-2120" type="氏集番号">
      ▼<tree n="17-2120" ord="partial" arity="7" order="43">
        ▼<root value="王裕之_17-2120_05" children="#真鶴_17-2120_06 #王三錫_17-2120_06 #真龜_17-2120_08 #真滿_17-2120_06 #王三德_17-2120_06 真那武樽 #王三秀_17-2120_06">
          ▼<label>
            元祖諱立思號肖國原是福建漳州府龍溪縣人也萬曆十九年奉17-2120_06 聖旨始遷中山以補三十六姓二十八年庚子奉使為補貢通事到閩地卒嗣後尸棺其弟王立威帶同故土四世
            <persName sameAs="王可法_17-2120">可法</persName>
            第二子也 小宗諱裕之
          </label>
          </root>
          ▼<iNode value="真鶴_17-2120_06" parent="#王裕之_17-2120_05" children="#unknown">
            <label>女真鶴</label>
          </iNode>
          ▼<iNode value="王三錫_17-2120_06" parent="#王裕之_17-2120_05" children="#真鶴a_17-2120_07 #王廷瑛_17-2120_07 #真牛a_17-2120_07 #王廷瑛_17-2120_07">
            <label>三錫</label>
          </iNode>
        </tree>
      </div>
    </body>
  </text>

```

図10 世系図のマークアップ

61 <https://tei-c.org/release/doc/tei-p5-doc/en/html/GD.html#GDFT>.

62 妻や子の婚姻関係については個人の「記録」部分にそれぞれ詳述されている。

63 <https://tei-c.org/release/doc/tei-p5-doc/en/html/GD.html#GDTR>.

4. 渡唐関係記事のマークアップ

琉球家譜における個人の記録のうち業績履歴に関する「助庸」には、進貢や接貢等の渡唐に関する具体的な内容が記録されている部分となっている。

本データのもととなった科研「沖縄の歴史情報」では、テキストデータの作成時に、より検索性を高めるために底本である『那覇市史』から若干の変更が行われており、年号が「同年」とあれば前の行の年号の繰り返しとして入力されている他、刊行時の書式どおりではなく文章の区切り位置で改行するなどの処理がされている⁶⁴。年号については、「十一」「拾一」といった表記の揺れは多少あるにせよ、ある程度機械的に処理できることがわかり今後の作業の効率化につながる⁶⁵。改行位置についても修正はせずにそのまま活用することにした。各行ごとに<s>(s-unit)⁶⁶エレメントで区切り、属性(@type)として、渡唐の目的（進貢・接貢・勤学等）を付与することにした。つづいて、文章中の地名や人名といった固有表現を中心にタグを付けていくが、地名や役職名等において同じものを指しているながら違う表記がなさ

```

</s>
▼<s type="進貢">
  <date calendar="#cal_QL" when="1726-03-05">雍正四年丙午二月初二日</date>
  奉命為
  <name type="job">進貢二號船通事</name>
  <date calendar="#cal_QL" when="1727-01-02">十二月十一日</date>
  隨
  ▼<persName>
    <name type="job">耳目官</name>
    <name type="唐名">毛汝龍</name>
    <roleName>田里親雲上</roleName>
    <forename type="名乗">盛武</forename>
  </persName>
  ▼<persName>
    <name type="job">正議大夫</name>
    <name type="唐名">鄭廷極</name>
    <roleName>宇地原親雲上</roleName>
  </persName>
  <placeName sameAs="#那覇">那覇</placeName>
  一齊開洋
  <date calendar="#cal_QL" when="1727-01-26">翌年丁未正月初五日</date>
  在
  <placeName sameAs="#久米島">古米山</placeName>
  開洋初十日到
  <placeName>蘭</placeName>
  公事已竣
  <date calendar="#cal_QL" when="1727-08-02">六月十五日</date>
  <placeName sameAs="#五虎門">五虎</placeName>
  開洋
  <date calendar="#cal_QL" when="1727-08-07">二十日</date>
  歸國
</s>
<pb n="3"/>

```

図11 渡唐関係記事のマークアップ

⁶⁴ 豊見山和行「「琉球家譜」情報化の問題点」『重点領域研究「沖縄の歴史情報研究」：総括班・沖縄3研究班合同研究会報告集』1996、16-17頁を参照。なお、変更の内容については、<teiHeader>の<encodingDesc> (encoding description)に記述する予定。

⁶⁵ 今回は渡唐の記事については試行として旧暦の月日までを西暦に変換して入力した。入力の際には中央研究院数位文化中心より提供されているデータベース「兩千年中西曆轉換」 (<https://sinocal.sinica.edu.tw/>) を活用している。

⁶⁶ <https://www.tei-c.org/release/doc/tei-p5-doc/ja/html/ref-s.html>。

れている場合もあることから、それらを検索や分析の際に同一のもととして判定できるよう、属性 (@sameAs) に情報を付与している⁶⁷。

5-4. データの活用：人文情報学の中琉交流史研究への可能性

TEIに準拠したXMLファイルとして作成した「琉球家譜」のデータは、XML向けに提供されるツール等で活用が可能である。

最もシンプルな活用方法としては検索用のテキストデータとして使用することであるが、正規表現やXPathによる検索を使用すればより高度な活用が可能となる。

正規表現 (Regular Expression) は、「任意の文字列の集合を、記号を用いて一つの文字 (列) で表現する方法⁶⁸」で、置換や検索の際に便利な記述方法である。例えば、「琉球家譜」のテキスト上では「大通事」と「都通事」が混在していることがあるが、その場合はテキストエディタの検索ツールにおいて正規表現を有効にした上で「(大|都)通事」と検索ボックスに指定することで、「大通事」も「都通事」を検索結果として表示される。正規表現については、詳述されているWEBサイトも多くあるので参照されたい。

XPathについては、XMLファイル上にタグ付けされた情報を用いてデータを絞り込むことが可能なルールであり⁶⁹、本稿のテキストでは例えば<placeName>や<persName>でタグ付けされた地名や人名の抽出や検索、年号や位置情報に基づいたデータの抽出などが可能である。<s>(s-unit)エレメントでタグがつけられた行のうち、進貢や接貢等の属性で絞り込むこともできる。

PythonやJavaScriptなどのプログラミング言語を使用して機械処理すればさらに活用の幅が広がる。マークアップの方法次第では、テキストに付与した位置情報を利用して地図データ上で可視化ができるなど、より中琉交流史研究へ資するデータを作成できると考えている。

67 一般財団法人人文情報学研究所[監修]『人文学のためのテキストデータ構築入門』文学通信、2022、175頁。

68 一般財団法人人文情報学研究所[監修]『人文学のためのテキストデータ構築入門』文学通信、2022、185頁。

69 一般財団法人人文情報学研究所[監修]『人文学のためのテキストデータ構築入門』文学通信、2022、185頁。

6. 電子テキスト資料を長期的に維持するには：課題と展望

テキストデータや画像データとして電子化した文書資料を長期的に維持する為に重要な事として、管理側にとって継続しやすい仕組みや長期保存しやすいデータのあり方を模索していくと同時に、「如何に多くの利用者に利用されていることを示せるか」という点も挙げられる。デジタルアーカイブなどのコンテンツ経費を継続して確保していくためには、システムの有効性をステークホルダーに対してどのように示せるかも重要であり、そのためにはデジタルアーカイブがどの程度教育・研究に寄与し、社会貢献につながったのかが「可視化」できることが求められている。

また、コンテンツを管理している機関だけの問題にせず、研究者や利用者をも巻き込んでデータの長期維持のあり方を検討し、活用していくことも可能であろう。その意味では「みんなで翻刻」や「デジタル源氏物語」の取り組みは機関側が提供する共有可能なコンテンツを有志の研究者によるプロジェクトによって、研究への活用や利便性の高いデータの作成に繋がる先例だと考えられる。本稿で取り扱った「王姓家譜」のエンコーディング作業は試行であり、本格的に「琉球家譜」のTEI適用を進める場合は、データの利活用や分析の方向性を考え、どのような構造にするべきかを検討したうえで、「琉球家譜」やTEIへの関心がある研究者との連携によって取り組む事が、よりデータの長期的な維持や活用が望めるのではないかと、という点を挙げて本稿のまとめとしたい。

謝辞

本発表は国立大学図書館協会海外派遣事業の助成を受けたものである。

本研究においては一般財団法人人文情報学研究所主席研究員永崎研宣氏および琉球大学附属図書館情報管理課雑誌情報係長大谷周平氏に多大なる協力やご教授を賜った。また、TEI-C東アジア/日本語分科会の勉強会への参加によって、TEIはじめ人文情報学に関する知見を広げることができた。ここに記して感謝の意を示す。

【参考文献】

1. 一般財団法人 人文情報学研究所[監修]『人文学のためのテキストデータ構築入門』文学通信、2022。
2. 岩崎宏之『文部省科学研究費補助金重点領域研究「沖縄の歴史情報研究」(領域番号:110)(平成6年度-平成9年度)CD-ROM版研究成果報告書』、岩崎宏之、1998。
3. 大谷周平, 富田千夏「琉球大学附属図書館のデジタルアーカイブ事業」『沖縄県図書館協会誌』22、2019。

4. 桶谷猪久夫 「「琉球家譜」の検索システム」 『重点領域研究「沖縄の歴史情報研究」：総括班研究成果報告書』1998、239-247頁。
5. 後藤真 「人文情報学と歴史学」 『歴史情報学の教科書-歴史のデータが世界をひらく』文学通信、2019、p22-22。
6. 高良倉吉 「総合解説」 『平成23年度琉球大学附属図書館貴重書展 文献資料にみる八重山・琉球』琉球大学附属図書館、2011、1頁。
7. 富田千夏 「琉球・沖縄関係貴重資料デジタルアーカイブについて」 『琉球沖縄歴史』3、2021、122-127頁。
8. 豊見山和行 「「琉球家譜」情報化の問題点」 『重点領域研究「沖縄の歴史情報研究」：総括班・沖縄3研究班合同研究会報告集』1996、16-17頁。
9. 豊見山和行 「「琉球家譜」テキストデータベースの作成」 『重点領域研究「沖縄の歴史情報研究」：総括班研究成果報告書』、1998、235-238頁。
10. 永崎研宣 「デジタル文化資料の国際化に向けて：IIIFとTEI」 『情報の科学と技術』2017、61-66頁。
11. 永崎研宣 「歴史データのさまざまな応用 -Text Encoding Initiative の現在-」 『歴史情報学の教科書 歴史のデータが世界をひらく』文学通信、2019、131-154頁。
12. 永崎研宣 「CA1989 - 動向レビュー：IIIFの概要と主要APIバージョン3.0の公開」、カレントアウェアネス: <https://current.ndl.go.jp/ca1989>、2020。
13. 永崎研宣 「デジタルアーカイブシステムにおける標準化技術」 『情報の科学と技術』71(4)、doi:https://doi.org/10.18919/jkg.71.4_165、2021、165-170頁。
14. 那覇市企画部市史編集室 『那覇市史 資料篇 第1巻 6 (家譜資料 2 久米系)』那覇市企画部市史編集室、1980。
15. バゼル 山本登紀子 「越境する沖縄関係資料-阪巻・宝玲文庫資料の電子化共有プロジェクト経験を通して」 『阪巻・宝玲文庫の世界 事業報告・研究報告会 報告書：The World of the Sakamaki/Hawley Collection』、2017、9-46頁
16. 山田浩世, 小野百合子 「戦後沖縄における資料収集・編纂と近年のデジタルアーカイブの取り組み」 『歴史学研究』1024、2022、28-39頁
17. Adobe Blog 「Flash とインタラクティブコンテンツの未来(原題:Flash & The Future of Interactive Content)」、2017年7月25日、<https://blog.adobe.com/jp/publish/2017/07/26/201707adobe-flash-update>

人文情報學之於中琉關係史資料的展望 —數位化資料的長期維持之一環—

富田千夏*

摘要

在對收集而來的古文獻資料之長期維持管理的課題上，資料的數位化處理無論在資料保存的目的上或是利用者的便利性而言都有更高一層的意義，於是近年來多數的資料保存機關開始推動數位化資料事業。然而很多機關在數位化資料的架構和運作上所面臨的問題是「如何長期維持」，這個難處就從目前無法利用的「沖繩歷史情報研究會」的實例開始說起。本稿針對數位化資料的長期維持課題，就「人文學上情報學技法和技術的應用」之人文情報學的手法採用，試舉琉球大學附屬圖書館的「琉球沖繩關係貴重資料數位化資料」之IIIF (International Image Interoperability Framework) 化處理事例以及重新啟動「沖繩歷史情報研究會」所作「琉球家譜」資料的嘗試。現在無法使用之「沖繩歷史情報研究會」的原檔中，以CD-ROM版收錄的「琉球家譜」資料，可以轉換成適用於現在系統的形式，再度受使用。藉著「琉球家譜」的資料文件的現行系統，採用實際標準的TEI (Text Encoding Initiative) 試行轉換，希望能就有關數位化資料的長期維持，一起檢討所屬機關及研究學者所能做的努力，期待今後能將數位化資料內容有關的課題共同分享。

關鍵字：デジタルアーカイブ、琉球家譜、TEI、人文情報學、Digital Humanities

*琉球大學圖書館情報服務部保存與宣傳科科長