

令和 6 年 6 月 10 日現在

機関番号：17102

研究種目：挑戦的研究（萌芽）

研究期間：2018～2023

課題番号：18K18508

研究課題名（和文）デジタルヒューマニティーズを促進するオープンデータ環境およびシステム基盤の構築

研究課題名（英文）Development of an open data environment and system infrastructure to promote digital humanities

研究代表者

石田 栄美（Ishita, Emi）

九州大学・データ駆動イノベーション推進本部・教授

研究者番号：50364815

交付決定額（研究期間全体）：（直接経費） 4,800,000円

研究成果の概要（和文）：オープンデータとして、九州大学附属図書館が所蔵している貴重書をデジタル化し、IIIFに対応したデジタルコレクションに加えた。デジタルコレクションのメタデータを補完することによりアクセシビリティが高まるかを検証した。また、デジタルコレクションのアクセス性を検討するため、大学図書館のデジタルコレクションの有無とオープン性を調査した。その他、オープンデータや研究データ管理・公開に関する動向調査や様々なイベントを開催した。

研究成果の学術的意義や社会的意義

貴重書をデジタル化し、IIIF形式でデジタルコレクションとして公開することにより、オープンデータに貢献した。デジタルコレクションのメタデータを補完することは、アクセス数の増加に貢献する可能性があることを示した。貴重書等のデジタルコレクションの公開の有無とオープン性を調査し、データのオープン化についての在り方を検討する材料を示した。オープンデータや研究データ管理・公開に関するイベントを開催し、それらの普及に貢献した。

研究成果の概要（英文）：As open data, rare books held by Kyushu University Library were digitized and added them with IIIF format to a digital collection. Enhancing metadata of digital collection improved accessibility was examined. Openness of digital collections were provided university libraries are examined. In addition, trend surveys and various events regarding open data and research data management and publications were held.

研究分野：図書館情報学

キーワード：オープンデータ デジタルヒューマニティーズ 研究データ管理 デジタルコレクション

## 1. 研究開始当初の背景

近年、内閣府など政府機関等からも、オープンサイエンスが推進されている。オープンサイエンスとは、研究成果（論文、生成された研究データ等）の容易なアクセス・利用を可能にし、科学技術研究を推進することを意味するが、これにはオープンアクセスとオープンデータが含まれる。オープンアクセスに関しては、大学では機関リポジトリを利用した研究成果のオープン化という形で進められており、国立大学を中心とした大学ではほぼ基盤が整っているといえる。一方、研究データの再利用を目指したオープンデータに関しては、これからの課題となっており、まだその基盤が構築されているとは言えない。生命科学や地球惑星科学などの分野においては学協会等が中心となりデータリポジトリを構築しているところもあるが、人文学分野においてはデータを登録する代表的なデータリポジトリは構築されていない。そのため、人文学分野の研究データを想定したオープンデータの基盤を構築することが必要である。

人文学分野では新しい研究の動きもある。デジタル技術の発達・普及の影響を受け、それらの技術を積極的に取り入れたデジタルヒューマニティーズ (Digital Humanities; DH) である。DH とは、デジタル化やテキスト化等、人文学分野の研究資料をコンピュータで扱いやすい形にした資料を対象にし、質的な分析と量的な分析を組み合わせるなどして行う研究である。デジタル化された資料を用いて研究することは、資料が物理的に存在する場所に行かなくても研究が進められることや資料を複数人で共有することができるため共同研究が進めやすいなどのメリットがある。さらに、今まで手作業で行ってきた分析をコンピュータの技術を用いて量的に分析することなどにより、新たな知見を見出せる可能性があり、近年、人文学研究において注目されている。このようなデジタル人文学への流れは、人文学における研究の方法、体制にまで影響を与えるようになっている。

## 2. 研究の目的

DH における研究を促進するために、オープンデータ環境基盤を構築することを目的とする。大学内でオープンデータを蓄積するためのデータリポジトリを構築し、そこに登録するデータに付与するメタデータを検討し、データリポジトリとしてデジタルコレクションを充実させる。デジタル化された人文学の資料は、国文学研究資料館や人文学オープンデータ共同利用センターなどが積極的にデータを提供しているが、それぞれの大学の図書館も貴重な研究資料を所蔵しており、これらをオープンに研究利用に供する。

背景でも述べたように、研究開始当初は、DH を促進するために、オープンデータの環境基盤を構築することを目的としたが、新型コロナウイルス感染症により、研究期間が延長された中で、オープンデータに関する動向にも変化があった。分野を問わず研究データそのものに注目が集まり、オープンデータから、研究データの管理・公開を大学等で支援する体制を整備することが求められるようになった。たとえば、科学技術研究費などの研究助成において研究データ管理計画の作成が求められるようになってきたり、論文や研究データを含む研究成果の即時オープンアクセスなどが求められるようになってきた。また、大学等においては、研究データに関するポリシーを策定することが求められている。適切な研究データ管理の先に、研究データの公開(オープン化)があると考えられるため、本研究課題では、研究データ管理・公開に関する動向調査なども含めることにした。

上記の関心から、様々な観点、アプローチで、本研究課題に取り組んだ。具体的には、以下のことを行うことを目的とした。

### (1) オープンデータや研究データ管理・公開の動向把握と啓もう

オープンデータや研究データ管理・公開の動向を把握する。また、これに関連するイベントを開催することにより、多くの人と動向等を共有することにより、オープンデータや研究データ管理・公開の促進を進める。

### (2) 資料のデジタル化と IIIF への対応によるオープンデータへの貢献

研究資料として価値が高い貴重書を選定し、それらをデジタル化し、IIIF に準拠した形で公開する。

### (3) デジタルコレクションにおけるメタデータ補完の効果

デジタルコレクションにデジタル画像を追加した場合、容易なアクセスができることが、オープンデータのメリットとなる。デジタルコレクションが始まった初期のころにデジタル化され、公開された画像には、メタデータの項目が十分に記入されていないものが存在する。そのため、それらを補完することによる効果をアクセシビリティという観点から評価する。

### (4) デジタルコレクションのオープン性

日本では多くの大学図書館でデジタルコレクションが公開されているが、日本の古い貴重書が対象となることが多いこともあり、国際的視点からみてオープン性があるのかわからない。日本の大学図書館のデジタルコレクションを、国際的視点も含めたオープン性に関する評価項目を用いて、評価を行った。

#### (5) デジタルヒューマニティーズとしての研究の試み

公開されたデータは、有効に使われることが重要である。デジタルヒューマニティーズの研究の一例として、くずし字の自動認識を試みた。

### 3. 研究の方法

#### (1) オープンデータや研究データ管理・公開の動向把握と啓もう

オープンデータや研究データ管理・公開の動向について、海外の大学などを訪問調査することにより、その動向を把握する。また、それらの動向を共有することや海外から専門家を招聘したイベントを開催する。

#### (2) 資料のデジタル化と IIF への対応によるオープンデータへの貢献

九州大学附属図書館が所蔵する貴重書のうち、研究資料として価値が高い資料を選定し、デジタル化し、IIF に準拠した形で公開する。公開先としてデータリポジトリを構築することを検討していたが、九州大学附属図書館のデジタルコレクションは IIF に準拠しているため、データ公開先として用いることにした。

#### (3) デジタルコレクションにおけるメタデータ補完の効果

初期のころにデジタル化され、メタデータの項目が十分に記入されていないものについて補完したため、その補完前と後でアクセス回数がどの程度変化するかを調べることにより、アクセシビリティが向上するかを検討した。

#### (4) デジタルコレクションのオープン性

日本の国立大学図書館のデジタルコレクションを対象に、オープン度合い、ページの組織化の方法などを調査した。さらに、英語のページの有無や日本語ページと英語ページの情報量と違いといった国際的な視点という点も加えて、そのオープン性を評価した。

#### (5) デジタルヒューマニティーズとしての研究の試み

公開されたデータは、有効に使われることが重要である。デジタルヒューマニティーズの研究の一例として、くずし字の自動認識を試みた。

### 4. 研究成果

#### (1) オープンデータや研究データ管理・公開の動向把握と啓もう

オープンデータやデジタルヒューマニティーズの啓もうのため、また、研究データ管理・公開に関して、シンポジウム、ワークショップ等、多くのイベントの企画、開催、講演した。また、講演者として、いくつかのイベントに参加した。

オープンデータ、デジタルヒューマニティーズの啓蒙のため、シンポジウム「オープンデータと大学」(2019年2月7日)を後援した。このシンポジウムにおいては、オープンデータに関して大学がどのように向き合っていくかを議論した。

オープンデータの流れの中で重要な要素の一つである研究データサービスに関して、イリノイ大学アーバナ・シャンペーン校から3名の図書館職員を招へいし、1.5日間にわたるシンポジウム・ワークショップ「大学における研究データサービス」(2019年12月5・6日)を開催した。シンポジウムでは、研究データ管理計画の作成や研究データをリポジトリに登録する支援など研究データに関してイリノイ大学図書館で実施されているサービスが紹介された。パネルディスカッションでは、日本の大学でオープンデータに通じる研究データの管理をどのように支援をしていけばよいかを議論した。ワークショップでは、学術雑誌におけるデータポリシーやデータキュレーションチェックリスト等の研究データを公開する際に確認するポイント等について、演習や議論をおこなった。

2020年度は、大学における研究データ管理、および研究データサービスの提供に関する検討を行った。大学において研究データ管理を行うには、システム的な基盤、制度的な基盤、人的基盤を整備することが必要であるが、これらについての検討を行った。その一環として、2020年11月に開催された図書館総合展のフォーラム「誰がやる？研究データ管理サービス」において、「九州大学で実施した研究データサービスに関するイベントの経緯とねらい」と題した講演を行った。研究データサービスに関する専門人材の育成の必要性や大学における研究データサービスに関して日本型モデルの構築の必要性について述べた。また、ディスカッションにおいては大学の他組織との連携が必要であることや研究データ管理支援に必要な知識・スキル等についても議論した。

2021年12月9日には、国際イベント「デジタルトランスフォーメーション時代のデータキュレーションと情報管理」を提案し、その中で「日本の大学において研究データ管理を実現させるための研究」と題した発表を行なった。その後、このテーマについて登壇者と議論した。

さらに、2023年6月19日には、データのオープン化の効果やオープン化を促進するための体

制づくりを議論するための国際シンポジウム「大学における研究データ管理の意義と支援人材育成」を企画、開催した。

その他、研究データ管理に関する支援サービスの実態を調査するため、米国の大学図書館(イリノイ大学アーバナ・シャンペーン校、ミシガン大学)と欧州の大学図書館(エディンバラ大学、グラスゴー大学、ケンブリッジ大学、ストックホルム大学等)の研究データサービス部門にて観察調査とインタビュー調査を実施した。

また、オープンデータの促進を支援する役割もあるデータライブラリアンについての動向も調査した。1995年から2022年にIFLAのメーリングリストに投稿された36の求人情報を分析した。職種名による分析では、昔は特定の分野のデータライブラリアンが求められていたのに対し、近年では、研究データライブラリアンが求められていることがわかった。職務内容の分析からは、2000年から2004年の間は、専門分野のデータに関する知識や統計分析ソフトが使えることなどが求められていたのに対し、2015年から2022年の間では、研究データに関する知識が求められていた。このように、データライブラリアンは、近年、研究データの保存や公開等を行うための研究データ管理としての職務が期待されていることがわかった。

#### (2)資料のデジタル化とIIIFへの対応によるオープンデータへの貢献

オープンデータを増やし、研究に供するため、九州大学附属図書館の貴重資料をデジタル化した。研究期間中、研究資料として価値が高いと判断される雅俗文庫をはじめとし、合計で225点、画像数にして約9700画像を九州大学附属図書館のデジタルコレクションに追加した。このコレクションは、IIIFに対応している。

#### (3)メタデータの補完とその効果

すでにデジタルコレクションで公開されている画像データのメタデータ項目を再確認し、アクセス性を高めるためのメタデータ項目を検討した。検討の結果、試験的に415件に対してメタデータの補完をおこなった。また、デジタル化されてはいるがテキスト化されていない手書きの冊子体目録一冊をテキスト化した。

次に、メタデータの項目に、情報を追加することによる効果を調査した。九州大学図書館デジタルコレクションのメタデータのうち、昔に作成されたメタデータは、基本的な項目しか情報が入力されていないものがある。そのため、メタデータの項目のうち、タイトルのカタカナ表記、タイトルのローマ字表記、代替タイトル、著者のカタカナ表記、著者の識別子(ID)、著者の別名、九州大学の書誌ID、出版年などの項目を対象に該当する情報を入力することでメタデータの補完を行ってきた。これらのメタデータを用いて、調査を行った。2020年と2021年に補完した7,249メタデータを対象に補完前と補完後のアクセス数を調査した。メタデータが更新された月の前後数か月のアクセス数を比較したところ、国内外を含むキャンパス外からのアクセス数が増加していた。このことから、メタデータの補完はデジタルアーカイブのアクセシビリティの向上に貢献する可能性があることがわかった。

#### (4)デジタルコレクションのオープン性

オープンデータ環境が現時点でどの程度整備されているかを把握するため、大学図書館が保持している貴重書等のデジタルコレクションを対象に、どの程度、公開されているか、また、それらは国際的な視点からみてアクセス性が高いといえるのかを検証した。

国際比較としてまずインドネシアの大学図書館におけるデジタルコレクションの公開の程度を把握した。その結果、インドネシアにおいては、デジタルコレクションを公開している大学図書館は少なく、またその画像に関する使用条件や権利などが明記されていないことが明らかになった。

次に、日本の国立大学図書館86館を対象に貴重書等のデジタルコレクションの公開の有無とオープン性を調査した。国際的なデータの流通という観点として、英語のページがどの程度提供されているかも調査した。調査の結果、57%の図書館でデジタルコレクションが提供されていた。デジタルコレクションがある49館のうち、制限なく利用ができるのが全体の8.2%であった。英語での情報提供は、ホームページやメタデータに関しては27%から35%程度で入手可能または一部入手可能であったが、ポリシーでは15%程度にとどまった。デジタルコレクションを介したデータのオープン化の国際化に関しては課題があることがわかった。

また、オープンデータの現状把握として、研究に使用されているデータへのアクセス性も調査した。既存論文のレファレンスに記載されたURLを分析し、データ共有の経年変化やアクセス性がどの程度保持されているかを分析した。Scopusから、2012年から2021年に発表された人文科学系の論文を対象に、引用数の多い方から年ごとに各2,000件の論文の書誌情報を収集し、Referenceセクション情報として収録されているテキストからURL及びDOI表記を抽出し集計した。DOIの明記が進み情報へのアクセス性は高まっているが、URLはその半数が無効になっていることが明らかになった。

#### (5) デジタルヒューマニティーズとしての研究の試み

デジタルヒューマニティーズ研究の一環として、くずし字の自動認識手法を2つ提案した。一つ目は、CNN (Convolutional Neural Networks) の画像特徴を抽出する特性を利用し、文字枠の選定を行うと同時に、当該文字の枠も学習させる方法である。二つ目は、区切り枠のみの予測であり、従来の手書き文字認識によく使われる多文字画像を単一文字ずつに分割してから認識するという手順をくずし字の認識に適用するという手法である。これらの提案手法を、誤認識率と枠の一致率という評価手法を用いて評価した。7万個の3文字データを用いて訓練し、7000個のテスト用データを用いて評価したところ、二つ目の提案手法の誤認識率は、すでに開催されたくずし字チャレンジ最優秀賞チームの誤認識率に比べ、よい結果を示した。また、枠の評価については、一つ目の提案手法の枠予測結果は、二つ目の提案手法よりよい結果を示した。

また、物体認識に基づいて、くずし字のセグメンテーションと認識を同時に行うくずし字認識法を開発した。この研究成果は文字認識に関するトップ国際会議の併設ワークショップの1つである HIP2019 に受理された。また、提案手法を応用した手法が PRMU 研究会のくずし字チャレンジ 2019 において第4位に入賞した。

## 5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 7件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Widiatmoko Adi Putranto, Regina Dwi Shalsa Mayzana, and Emi Ishita	4. 巻 14458
2. 論文標題 The Openness of Digital Archives in Japanese Universities and Its Opportunities	5. 発行年 2023年
3. 雑誌名 In: Goh, D.H., Chen, S.J., Tuarob, S. (eds) Leveraging Generative Intelligence in Digital Libraries: Towards Human-Machine Collaboration.	6. 最初と最後の頁 123 ~ 138
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-981-99-8088-8_11	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Widiatmoko Adi Putranto, Regina Dwi Shalsa Mayzana, Emi Ishita	4. 巻 13636
2. 論文標題 Opening Access to Digital Collections: The State of Cultural Materials in Indonesian Higher Education Institutions	5. 発行年 2022年
3. 雑誌名 In Proceedings of International Conference on Asia-Pacific Digital Libraries (ICADL2022)(Lecture Notes in Computer Science)	6. 最初と最後の頁 231 ~ 240
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-031-21756-2_19	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tokinori Suzuki, Emi Ishita, Xinyu Ma, Widiatmoko Adi Putranto, Yukiko Watanabe	4. 巻 13636
2. 論文標題 Data Librarians: Changes in Role and Job Duties Over 25 Years	5. 発行年 2022年
3. 雑誌名 In Proceedings of International Conference on Asia-Pacific Digital Libraries (ICADL2022)(Lecture Notes in Computer Science)	6. 最初と最後の頁 485 ~ 491
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-031-21756-2_40	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tetsuya Nakatoh, Hironori Kodama, Yuko Hori, Emi Ishita	4. 巻 13133
2. 論文標題 Enriching the Metadata of a Digital Collection to Enhance Accessibility: A Case Study at Practice in Kyushu University Library, Japan	5. 発行年 2021年
3. 雑誌名 In Proceedings of International Conference on Asia-Pacific Digital Libraries (ICADL2021)(Lecture Notes in Computer Science)	6. 最初と最後の頁 411 ~ 418
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-91669-5_32	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tang Yiping, Hatano Kohei, Takimoto Eiji	4. 巻 -
2. 論文標題 Recognition of Japanese Historical Hand-Written Characters Based on Object Detection Methods	5. 発行年 2019年
3. 雑誌名 Proceedings of the 2019 Workshop on Historical Document Imaging and Processing(HIP 2019)	6. 最初と最後の頁 72-77
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3352631.3352642	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tetsuya Nakatoh, Sachio Hirokawa	4. 巻 11569
2. 論文標題 Evaluation Index to Find Relevant Papers: Improvement of Focused Citation Count	5. 発行年 2019年
3. 雑誌名 Human Interface and the Management of Information. Visual Information and Knowledge Management - Thematic Area, HIMI 2019, Proceedings (Lecture Notes in Computer Science)	6. 最初と最後の頁 555-566
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-22660-2_41	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 宮崎 智, 桂 尚輝, 長井 歩, 青池 亨, 唐 一平, 鈴木 拓矢, 玉木 徹, 内田 祐介, 西山 正志, 緒方 貴紀, 白井 啓一郎, 中村 和晃, 北本 朝展, カラーヌワット タリン	4. 巻 IEICE-119, No. IEICE-PRMU-347
2. 論文標題 第23回PRMU研究会アルゴリズムコンテスト実施報告	5. 発行年 2019年
3. 雑誌名 電子情報通信学会技術研究報告	6. 最初と最後の頁 41-47
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tang Yiping, Kohei Hatano, Emi Ishita, Tetsuya Nakatoh, Toshifumi Kawahira	4. 巻 -
2. 論文標題 Construction of Japanese Historical Hand-Written Characters Segmentation Data from the CODH Data Sets	5. 発行年 2018年
3. 雑誌名 Proceedings of the Eighth Conference of Japanese Association for Digital Humanities (JADH2018)	6. 最初と最後の頁 183-185
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 中藤 哲也, 石田 栄美
2. 発表標題 オープンデータとオープンアクセスの現状分析について
3. 学会等名 情報処理学会 第86回全国大会
4. 発表年 2024年

1. 発表者名 Emi Ishita, Kohei Hatano, Tetsuya Nakatoh
2. 発表標題 The Development of an Open Data Infrastructure to Promote Digital Humanities
3. 学会等名 10th Asia Library and Information Research Group Workshop (ALIRG2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Yiping Tang, Kohei Hatano, Emi Ishita, Tetsuya Nakatoh, Toshifumi Kawahira
2. 発表標題 Construction of Japanese Kuzushi Characters Segmentation Data for Recognition Project
3. 学会等名 10th Asia Library and Information Research Group Workshop (ALIRG2018) (国際学会)
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	中藤 哲也  (Nakatoh Tetsuya)  (20253502)	中村学園大学・栄養科学部・准教授    (37109)	



## 6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	畑 堃 晃平  (Hatano Kohei)  (60404026)	九州大学・基幹教育院・准教授    (17102)	

## 7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計1件

国際研究集会 国際シンポジウム「大学における研究データ管理の意義と支援人材育成」	開催年 2023年～2023年
---	--------------------

## 8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------