

令和4年度修士論文題目一覧

学位	学生氏名	修士論文題目	論文公開
修士 ライブラリーサイエンス	キ ムヒ 季 夢菲	国立大学アーカイブズにおけるデジタルアーカイブの構築に向けた現状と課題	公開
修士 ライブラリーサイエンス	バ 馬 キンユ	大学図書館における研究データ管理支援人材の特徴 －米国の求人情報に基づく分析－	公開
修士 ライブラリーサイエンス	リ ゲン 李 彦トウ	留学生の授業外学習の実態に基づく図書館サービスの検討	公開
修士 ライブラリーサイエンス	ワタナベ コウヘイ 渡邊 康平	論文の記述を用いたオープンデータの検索支援に関する研究	公開

# 国立大学アーカイブズにおけるデジタルアーカイブの構築に向けた現状と課題

Status and problems of creating a digital archive within national university archives

2FS20201E 季 夢菲 JI Mengfei

近年、デジタル技術の発展に伴ない、デジタルによる情報発信が拡大し、デジタル情報へのアクセスのニーズが上昇した。特に、新型コロナウイルスの影響により、遠隔授業とテレワークの必要が多くなり、デジタルアーカイブ資源の需要が高まっている。このような社会情勢の変化の中、遠隔で様々な活動を可能とする基盤としてのデジタルアーカイブを構築することの重要性が見えて来た。

日本において、国立大学は国立大学法人として、運営、研究、教育等遂行のために収集・管理した情報を公開すべき、高等教育・研究機関として、研究資源、研究成果、教育データなどの知的資源の管理責任も持っている。大学はより多くの情報発信をすること、及びデジタルアーカイブ社会の構築に参画することが期待される。大学アーカイブズは、大学による情報発信の重要な一環であるが、大学アーカイブズにおけるデジタルアーカイブの構築が遅れている。

本論文は、日本の国立大学アーカイブズにおけるデジタルアーカイブの現状について調査を行ったうえ、現状を引き起こす問題や原因を分析し、更に、これらの問題解決に向かい、自分の考案を提出する。

調査を始める前に、まず、アーカイブズとデジタルアーカイブの概念と範疇を明確にすることが必要である。また、同じく大学で情報資源を保存・公開するデジタルアーカイブと機関リポジトリ、この二者の異同をはっきりした。

日本の大学デジタルアーカイブの現状を把握するため、日本の国立大学のデジタルアーカイブに関する情報をなるべく多く集め、調査を行う。情報を集めるツールは国立国会図書館が発表した「文化・学術機関におけるデジタルアーカイブ等の運営に関する調査研究」のような公的機関による統計もあれば、デジタルアーカイブのウェブサイトを一覧にした「デジタルアーカイブリンク」のような企業による統計もある。他にも、国立国会図書館や国立公文書館など、国レベルの文化的情報資源を収集・蓄積・提供する公共機関が各自構築したデジタルアーカイブとの連携状況の確認、国立大学のホームページに掲載されているデジタルアーカイブに関する情報のチェックなどの調査も行い、いろんなツールを使い、できるだけ情報収集をし、最終的に現状の把握に近いことになる。

その中に、特に大学アーカイブズを主体とするデジタルアーカイブに対して、より詳しい情報を手に入れるため、

「国立公文書館等」として指定を受けた館からいくつかの代表例として選択し、聞き取り調査を行う。それぞれの館が目録のデジタル化・公開、コンテンツのデジタル化・公

開、検索システムの構築などの現状とその理由、デジタルアーカイブに対する意欲と努力などについて調査した。

十分な現状調査を行った結果、大学において、デジタルアーカイブの主体が図書館になっているという傾向があることがわかった。大学アーカイブズは、その重要性に対する認識が高くなってきたが、デジタルアーカイブを実際に構築しているところが少ない。その理由として、「予算不足」と「人員不足」の二つがよく語られる問題であるが、「ノウハウ・技術の不足」、「メタデータ問題」、「著作権と個人情報処理」などの問題も挙げられる。他にも、デジタルアーカイブに関わる専門職の設立、プロセスの形成、評価基準の確立など一連の政策とやり方の補充が期待される。

それらの問題の多くに対して、それぞれ解決策が考えられるが、特に技術や権利関係問題に対して、専門家に依頼すれば解決に向かうに違いない。しかし、予算と人員の共通問題が客観的に存在している上で、どのようにして、最低限のお金と人手で、デジタルアーカイブの構築と長期的な運営を保証するかが一番重要な課題になる。それに対して、費用と人員を節約し、効率を高める考案をいくつか提出する。

まずは学内の各組織・部署の資源を集約し、業務を行うことである。具体的に、デジタルアーカイブの構築にあたって、必要な作業に対して、一つ一つの部署にそれぞれ機械、専門職などの資源を分けるのではなく、統一的に業務を集約する一か所にまとめて、それぞれの部署にこの特定の作業の協力と支援を行うことである。例えば、コンテンツのデジタル化、システムの開発とプロセスの設計は、集約することにより、大いにコストを削減することが実現できる。また、MLA連携の促進により、資源と経験の共有を進めることができ、協力し統合ポータルを作ることも効率的で役立つ方法として考えられる。最後に、業務の集約や連携の促進を長期的且つ安定的に行うには、体制と仕組みの確保が避けられない基本問題である。デジタルアーカイブの責任主体を確定し、ポストを形成し、それぞれのポストに人手を適当に配置することが基本である。それに、体制を長期的に保証するため、評価体制で正確に評価することも大事である。各機関が積極的に取り組めるよう、制度の見直しの検討も含め、評価基準の確立や意識の向上につながる仕組みが必要である。

これらの考案はあくまでも一部の問題の解決策で、デジタルアーカイブの普及は決して容易に実現できないものの、これから長く経験を積み重ね努力すべきことである。

# 大学図書館における研究データ管理支援人材の特徴—米国の求人情報に基づく分析—

## Characteristics of Research Data Management Support Personnel in Academic Libraries: An Analysis Based on Job Advertisement in the U.S.

2FS21202P 馬 キンユ MA Xinyu

ビッグデータ時代の到来と e-Science の進展によりデータ量が大幅に増加した。学術界においても研究データが急速に増え、これらを十分に記述・共有・保管・再利用するために研究データ管理が重要になっている。しかし、研究者だけでデータを管理することは難しく、国、資金提供機関、大学、図書館等が協力して支援することが必要である。特に、大学図書館は、これまでも学術情報流通において重要な役割を担っており、また、研究活動を支援できる強力な組織としてこれに貢献できる。

現在、日本では、研究データ管理の支援人材の不足という課題に直面している。一方、米国の大学図書館は日本よりも先に研究データ管理サービスを開始し、長年の経験を通して、より完成度の高いサービスモデルを持っている。本研究では、米国の大学でデータ管理支援を行う人材の特徴を把握し、日本の研究データ管理支援人材に必要な知識やスキルを検討することを目的とする。

本研究の分析対象は、米国の大学による研究データ管理支援人材の求人情報である。2021年11月から2022年4月の間に ALA JobLIST、Inside Higher Ed、および IFLA libjobs に掲載された求人情報のうち、「data management」というフレーズを含む求人情報を収集し、重複等を除いた 89 件を分析の対象とした。これらに対し、求人情報に含まれる「部局」、「業務」、「職種名」、「必要な学歴」、「必要な経験」、「必要な能力」と「好まれる能力」を分析した。

分析の結果、求人情報の募集部局は、図書館が 56.2% であり最も多く、次いで研究所が多かった(20.2%)が、その他に学生サービス部門や特定の学部でも募集していた。研究データ管理支援人材の職種名は、48.3% に「data」が含まれており、43.8% に「librarian」が含まれていた。特定の分野が含まれる職種名は 12.3% だった。このことから、研究データ管理支援には、図書館だけでなく、他の部局も関わっていることがわかる。

求人情報に記述されている業務内容を集計したところ、「研究データ管理」という包括的な用語が使われているものが 17.3% だった。研究データのライフサイクルにおける活動を具体的に記述したものは、「データ分析」「データ収集」「学術情報流通」が最も多く挙がっており、先行研究と比較して、データ分析への要求が強くなり、学術情報流通が新たに加わっていることがわかった。

必要な学歴については、大学図書館 50 件の求人情報のうち、35 件は図書館情報学の修士号、14 件は図書館情報

学または関連分野の修士号を求めている。関連分野には、統計、計算、情報技術、コンピュータなどが含まれる。求められる経験では、図書館でのレファレンスサービスの経験の要求が最も多く、次に図書館を通じた研究サービスや支援の提供、指導であった。

求められる能力(必要な能力・好ましい能力)で、最も多く言及されていたのは、「優れた文書および口頭でのコミュニケーション能力」「対人関係能力、チーム環境において他者と効果的に協働する能力」「研究データ管理に関するリテラシー教育を行う経験」「データ管理の知識と実践」であった。また、DEI(多様性・公平性・包摂性)に対する理解、サービス意欲も求められる能力に含まれていた。

上記の結果から、米国の大学図書館における研究データ管理支援人材の特徴を把握した。募集部局から見ると、図書館だけでなく、研究所、学生健康部、事務室、学部、情報技術部、学生課などの部局も研究データの管理に携わっている。職種名に「data」が多く含まれている職員だけでなく、多くのライブラリアンが研究データ管理支援を行っている。ほとんどの場合、図書館情報学または関連分野の修士号が求められ、特に米国図書館協会が認定する学位の要求が多い。それ以外に求められる修士号の専攻分野は、情報処理に関するものが多い。求められる業務として、研究データに関する取り組み以外は、学術情報流通がますます重要になっている。また、図書館でサービスを提供した経験が希望されている。最も求められる能力は、チームワーク力やコミュニケーション力といった汎用的な対人能力であり、データ管理に関する専門的な知識や実践、指導経験も求められている。

日本においても研究データ管理支援人材に必要な知識やスキルが提案されているが、それらと比較してみると、米国の大学の研究データ管理支援人材には DEI 意識と卒業後の業務経験が求められていることが明らかになった。DEI の普及が日本では始まって間もないこと、米国と日本では採用の仕方が異なり、卒業後の業務経験が日本では求められないことが、その理由と考えられる。今後は、日本の大学図書館の研究データ管理支援人材も DEI を意識しながら業務を遂行する必要がある。また、大学在学中などにインターンシップを実施したり、入職後に研修を行うことが考えられる。例えば、研究データ管理支援において要求が高かった一般的なレファレンスサービスや研究支援・指導などを組み込むことが有効であろう。

# 留学生の授業外学習の実態に基づく図書館サービスの検討

Discussion of library services based on the actual status of out-of-class learning of international students

2FS21201W 李彦トウ LI YANTONG

近年、社会、経済と労働市場のグローバル化により、教育における国際化はますます重要になっている。各国は留学生人材に注目し、留学生の人数が増え、属性も多様化している。日本は2008年に「留学生30万人計画」を発表し、日本社会のグローバル化を目指した。その後、留学生の人数が顕著に増加し、2019年に留学生30万人を達成した。しかし、2019年から新型コロナウイルスが世界中に急速に広がり、留学生の数も減少した。新型コロナウイルス感染防止のため、各国で遠隔授業が始まった。遠隔授業に伴い、学生は授業の理解度が低い、学習負担が多い等の困難に直面した。留学生も遠隔授業を受けており、一部の留学生は日本に入国できない状況になった。コロナ禍で授業形態が転換されると、授業課題が増え、留学生の授業外学習時間も増加し、自分の希望する学習ができるかや学習成果をあげることができるかに不安を感じるようになった。

コロナウイルス感染症の影響で、大学や図書館が閉校、臨時休館になった。それに即した学習支援は、郵送貸出、電子ジャーナル・電子書籍・データベースへの自宅など学外からのアクセス、電子書籍の整備、オンラインツールを活用した講習会などがある。従来、日本の大学図書館は留学生に対して、母国語のガイド案内や留学生用のインターナショナルエリアなど、様々な支援を実施してきたが、具体的な実践や要望調査の蓄積が乏しいことが指摘されていた。平時でさえ留学生は様々な原因で学習に困難を感じる場合が多いため、コロナ禍の特殊状況下においては、支援を強化する必要がある。中でも授業外学習を支援する留学生向けの図書館サービスを検討すべきであろう。

以上を踏まえ、本研究の目的は、コロナ禍における授業外学習の実態と留学生が直面している問題点を明らかにし、大学図書館による留学生向け学習支援のサービスを改善する提案を出すことである。そのため、まず2022年1月18日から2月1日の間、九州大学の修士一年の学生33人にフォトダイアリー調査を行った。本研究で用いたフォトダイアリー調査では、LINEを使って、実験参加者に写真とそれに関する説明事項を一緒に送信することを要求した。撮影する写真は、学習時の目の前の風景(学習時の状況が分かる)写真とした。説明事項は、記入の順に、①学習時間、②場所、③内容、④リソース、⑤共同学習者の人数、⑥感想、⑦備考とした。フォトダイアリー調査後、2回の半構造化インタビューを行った。第1次インタビューでは、フォトダイアリー調査に参加した学生22人を対象に、2022年2月から3月の間、5回に分けて調査を行った。第2次インタビューは2022年8月14日に行い、調査

対象は第1次インタビューに参加し、かつ参加時点で渡日しておらず、8月までに渡日できた学生4名とした。インタビューの質問内容は主に、1)フォトダイアリー調査投稿結果の確認(フォトダイアリー調査への感想、学習内容、学習場所、学習人数、喜怒哀楽などの傾向の確認)、2)フォトダイアリー投稿結果に応じた困難点や行動理由の質問、3)大学図書館の利用状況・困難点・論文探しの方法の確認、4)オンライン授業下・コロナ・入学以来の様々な問題の質問、の4点である。

フォトダイアリー調査の投稿結果により、修士一年の留学生の授業外学習像が構築できる。留学生は2週間、平均毎日1回から2回授業外学習を行い、一日の授業外学習時間は平均約64分であった。また、授業外学習をする場所は圧倒的に自宅が多く、自宅で授業外学習する理由は新型コロナウイルス感染症への心配だった。留学生は、様々なリソースを活用し、コロナ禍の影響下で学習を続けていた。彼らは研究や授業を重視し、授業外で研究や授業に関する学習に力を入れることがわかった。また、留学生が困難を感じたこととして、インタビュー調査やフォトダイアリー調査の両方で、授業への理解や授業課題に関する困難への言及が多かった。特に研究と授業両方で質問に関する困難があることがわかった。留学生は入国後、人間関係の構築に伴い質問に関する困難を解決していた。図書館利用に関する問題は、1)利用不便、2)サービス不明、3)文献検索困難、4)資料問題があった。留学生がよく利用した図書館サービスは「データベースでの論文検索」であり、海外にいる留学生は、図書館に関する情報入手は大学メールに頼る傾向があった。大学図書館の利用に関して、海外にいる留学生は図書館に関する情報の不足を示した。以上の結果をもとに、1)図書館利用に関する情報発信の方法や頻度の増加、2)オンラインリソースを増加する、3)コミュニケーション機会の創造、4)母国語による図書館ガイドの作成、5)日本語の支援だけでなく、英語の学習も支援する、という5点を提案した。

高等教育のグローバル化により、留学生人材を重視している現在、留学生の困難に目を向ける必要がある。本研究は、新型コロナウイルス感染症の影響により、大学が遠隔授業を続けていた特殊な状況下で、日本に居住している留学生と母国にいる留学生の学習実態や困難を詳細に分析した。その調査分析結果は、日本の大学と大学図書館に留学生支援のための基礎的なデータや知見を与えるものであり、大学図書館は本研究の提案を実践することで、留学生の困難により良く対応することができるようになるであろう。

---

# 論文の記述を用いたオープンデータの検索支援に関する研究

A Research on Open Data Retrieval Support Using Research Paper Descriptions

---

2FS21203S 渡邊 康平 WATANABE Kohei

---

データセットとは、表やグラフ、文書、試料、アンケート結果など、特定の目的のためにまとめられた一定の形式を持つデータの集合である。データセットは近年、機械学習や人工知能の分野を筆頭に様々な分野でその利用価値が高まっており、入手が容易になりつつある。データセットのうち、条件を満たせば自由に利用や加工ができるデータはオープンデータと呼ばれる。オープンデータの利活用については世界中で議論されており、データセットが作成された分野とは異なる分野の研究者が利用することで、研究や活動にイノベーションを起こすことが期待されている。

データセットは圧縮されていたり、データセットの内容が数値情報などであることが多かったりすることからその内容は検索の対象にはされず、データセットに付与されたメタデータやデータセットの配布ページにある記述を対象として検索が行われる。オープンサイエンスに期待されるイノベーションを起こすためには、データセットが作成された分野と異なる分野の研究者が検索を行ってデータセットを発見できなければならないが、データセットが作成された分野とは異なる専門分野の検索者がメタデータとして付与されている情報をクエリとして思いつくとは限らない。データセット作成者が付与したメタデータが検索者にとってどれほど有用であり、実際に検索可能かという議論はほとんどなされておらず、データセットに付与されたメタデータやデータの配布ページ以外の記述からもデータセットを検索できる可能性について議論する必要があると考える。

データセット検索のための支援を行うシステムは研究が行われているが、部分的な分野の支援にとどまっており全ての分野のデータセットが支援の対象となっているわけではない。また、オープンデータを検索するツールが研究機関などによって公開され、オープンサイエンスへの一助となっているが、検索ツールによっては有料であるため利用できる研究者が限られていたり、また機能不足な部分があったり、登録される情報に制限があり検索対象にならない場合もあるため改善が必要とされる。

そこで本論文では、データセットが配布されているページに登録されているメタデータ以外に、データセットを使用した論文にかかっている、データセットに関する記述を利用することを考えた。論文の記述をメタデータとして利用した際のデータセット検索の有効性を検討するために、データセット名をクエリとして Google Scholar で論文の検索を行い、ヒットした論文からデータが配布されていた

ウェブページにはない新たな記述をどの程度発見できるか調査した。また、それら新たな記述がどのような情報であり、メタデータ（キーワード）として利用できるか否かという観点で、分類し検討した。

インターネット上で公開されている機械学習向けの、合計100件のデータセットとそれを利用している864件の論文を対象とし調査を行った。対象となった論文から、検索クエリとなったデータセットについて論述されている部分を抜き出し、データセットの配布ページを基準として、配布ページとは異なった記述がなされていたものを新たな記述であるとした。結果249件の論文の記述にデータセットを配布しているウェブページには存在していなかった新たな記述が発見された。また発見された記述を「データセットの使用目的についての新たな記述」、「データセットのデータの属性名についての新たな記述」、「データセットの作成者側から得られるはずの新たな記述」、「その他の記述」の4種類の記述に分類した。さらに検索者がクエリを作成し、現在の検索エンジンを用いてAND検索を行ったと想定した際に、発見された記述を、データセットにメタデータとして追加し利用することができるかを検討した。その結果、「データセットの使用目的についての新たな記述」と「データセットのデータの属性名についての新たな記述」に該当する78件の記述について、メタデータとして利用できる可能性があるとした。また、「データセットの作成者側から得られるはずの新たな記述」と分類された記述に関しては、現在の検索システムにおいては有用ではないが、後の発展次第では検索に有用に利用できる可能性がある」と判断した。

本研究の課題としては、まず対象とした100件のデータセットは機械学習向けのものであることや調査結果の規模からみて、行った分類がすべてのデータセットに対して有効であるとは限らず、より多くの分野において大規模な調査が必要なが挙げられる。また、本研究では論文の記述をメタデータとして利用することで検索することのできるデータセットの範囲を拡大することのできる可能性があることは主張したが、得られた結果が特に非専門家に対して効果的であるかを主張することができなかった。

本研究はデータセットに対するメタデータの追加につながっていくと考え、今後は論文からデータセットに関する記述を機械的に抽出することを目指し、またその記述をメタデータとして追加することによって実際に非専門家に対する支援となるか調査を行う必要があると考える。