

JAIRO Cloud 導入前後における
機関リポジトリのオープンアクセスコンテンツ
利用

筑波大学
図書館情報メディア研究科
2019 年 3 月
小林 俊貴

目次

1. 序論	1
1.1 研究背景.....	1
1.1.1 OpenAccess.....	1
1.1.2 Green OA.....	1
1.1.3 Gold OA	2
1.1.4 BOAI10 と OA に向けた日本での活動.....	3
1.1.5 BOAI15 と OA の実現における問題	3
1.1.6 日本での Green OA 実現の問題点.....	5
1.1.7 歴史的背景.....	7
1.1.8 世界の機関リポジトリ数と日本の機関リポジトリ数.....	10
1.1.9 日本の機関リポジトリ	13
1.1.10 日本国内における機関リポジトリに関する政策	14
1.1.11 日本国内における機関リポジトリに影響する法制度	20
1.1.12 日本国内における機関リポジトリの変遷と今後の位置づけ.....	20
1.2 用語の定義	21
1.2.1 Open Access とは.....	21
1.2.2 機関リポジトリとは.....	22
1.3 研究目的と意義.....	23
1.3.1 研究目的	23
1.3.2 研究の意義.....	23
1.4 先行研究.....	23
1.4.1 機関ごとの調査と標準化、ネットワーク化に関する論文.....	24
1.4.2 個別のリポジトリの知見の共有を目的とした論文.....	24
1.4.3 アクセスログを用いた評価に関する論文	25
1.4.4 関連研究	26
2. 調査手法.....	28
2.1 分析対象の概要	28
2.1.1 筑波大学の機関リポジトリ	28
2.1.2 長期的なアクセスログの保管について.....	31
2.1.3 アクセスログデータの概要	31
2.1.4 用語の解説	36
2.1.5 分析対象期間と単位	40
2.2 データの処理.....	40

2.2.1 アクセスログのスクリーニング	41
2.2.2 COUNTER とは	41
2.2.3 本研究におけるフィルタリングの構造	41
2.2.4 フィルタリングの種類と結果の比較	42
2.2.5 データの種別と重み付けについて	43
2.3 データの分析	43
2.3.1 調査対象 1: Host の抽出によるアクセス元の調査	43
2.3.2 調査対象 2: コンテンツの利用の定義と調査	43
2.3.3 調査対象 3: 検索語句のクリーニングと分析	45
2.3.4 検索語のスクリーニング	45
2.3.5 ストップワード	45
2.4 手法のまとめ	47
2.5 仮説	48
3. 結果	49
3.1 アクセス元の分析	49
3.1.1 アクセス元の調査と内訳の結果	50
3.1.2 トップレベルドメイン別のアクセス元の内訳	52
3.1.3 セカンドドメイン別のアクセス元の内訳	54
3.1.4 アクセス元に関する考察	57
3.2 コンテンツの利用に関する分析	57
3.2.1 全期間を対象としたコンテンツの利用実態	58
3.2.2 ダウンロード回数の多いコンテンツの利用実態	66
3.2.3 紀要の利用実態	69
3.2.5 紀要の利用実態に関する考察	77
3.2.6 学術論文の利用実態	79
3.2.7 学術論文の利用実態に関する考察	87
3.2.8 学位論文の利用実態	88
3.2.9 学位論文の利用実態に関する考察	98
3.2.10 利用のなかったコンテンツの調査	99
3.2.11 コンテンツ利用に関する仮説の検証	103
3.3 検索語句に対する調査	105
4. 考察	109
4.1 アクセス元とデータの利用に関する考察	109
4.2 対象機関におけるメタデータの現状	111
4.3 今後求められるデータ管理と図書館員の役割	111
4.4 コンテンツ利用に関する考察	112

4.4.1 紀要に関する考察.....	112
4.4.2 学術論文に関する考察	113
4.4.3 学位論文に関する考察	114
4.4.4 Research Paper に関する考察	114
4.4.5 コンテンツ利用に関するまとめ.....	115
4.5 考察のまとめ	116
5. 結論.....	117
謝辞.....	118
文献リスト.....	119

図表目次

図

図 1 つくばリポジトリにおけるランキング画面	6
図 2 世界の論文件数の変化	8
図 3 学術コミュニケーションのモデル	8
図 4 自然科学系学術雑誌の価格推移	9
図 5 人文・社会科学系学術雑誌の価格推移	9
図 6 新しい学術コミュニケーションのモデル	10
図 7 OpenDOAR に掲載されている機関リポジトリ数	11
図 8 BASE に掲載されているドキュメント数上位 15 ヶ国	11
図 9 BASE に掲載されているリポジトリ数上位 15 ヶ国	12
図 10 日本の機関リポジトリ数と 世界の機関リポジトリ数の推移	13
図 11 つくばリポジトリのコンテンツ数の変化	29
図 12 2011/01/02 http 通信のログより抽出	32
図 13 2011/01/02 https 通信のログより抽出	32
図 14 筑波大学附属図書館の robots.txt	34
図 15 COUNTER	42
図 16 機関リポジトリに登録されたコンテンツの利用回数の経年変化	58
図 17 コンテンツ登録数の経年変化	59
図 18 利用回数上位 1,611 件における学位論文の経年変化(登録数)	60
図 19 利用回数上位 1,611 件における学位論文の経年変化(利用回数)	61
図 20 コンテンツの登録年度別利用回数の変化	62
図 21 コンテンツ登録年度割合	62
図 22 コンテンツ記述言語割合	63
図 23 登録コンテンツの type 別割合	64
図 24 コンテンツ数の登録年度推移	67
図 25 コンテンツの言語比率(N=1,611)	68
図 26 コンテンツ type 種別比率	68
図 27 登録年と登録コンテンツ数	71
図 28 登録コンテンツの言語比率(N=17,950)	72
図 29 紀要利用数の経年変化	74
図 30 登録コンテンツ数の推移	80
図 31 登録コンテンツの言語比率(N=6,993)	80
図 32 雑誌論文の収録コンテンツ数ランキング	83
図 33 登録コンテンツ数の推移	89
図 34 登録コンテンツ数の言語比率	90

図 35 学位論文の利用回数の経年変化	91
図 36 利用回数ゼロのコンテンツの type 別割合	99
図 37 利用回数ゼロのコンテンツの登録年度	100
図 38 利用回数ゼロのコンテンツの言語比率	100

表

表 1 BASE に掲載されていたうち、OAmark 率が 100%のリポジトリ	12
表 2 日本の機関リポジトリ数と世界の機関リポジトリ数の推移.....	13
表 3 学術情報流通と機関リポジトリに関する政策資料	14
表 4 つくばリポジトリの運用形態の変化とシステム管理者	28
表 5 内訳(コンテンツ登録件数 Top3)	29
表 6 内訳(コンテンツ登録件数 Top3 以外).....	29
表 7 2012.10 時点で本文有コンテンツ数のランキング	30
表 8 DSpace のアクセスログの要素と色分け.....	33
表 9 Jairo Cloud のアクセスログデータのメタデータ(利用統計)	35
表 10 JAIRO Cloud のアクセスログデータのメタデータ(サイトライセンス).....	36
表 11 WEKO で取得可能な Junii2 形式のメタデータ	37
表 12 Oracle の提供する stopword.....	46
表 13 NII の提供するストップワード.....	47
表 14 各年度の IP アドレス(2010 年 3 月-2014 年 3 月)	49
表 15 各年度の IP アドレス(2014 年 5 月-2017 年 12 月)	49
表 16 2010 年 3 月から 2014 年 3 月における http 通信のアクセス元の内訳	50
表 17 2010 年 3 月から 2014 年 3 月における https 通信のアクセス元の内訳	51
表 18 2010 年 3 月から 2014 年 3 月におけるアクセス元の内訳.....	51
表 19 http 通信の Host Top10 (トップレベルドメイン)	53
表 20 https 通信の HostTop10 (トップレベルドメイン).....	53
表 21 https 通信 Top10 の Host (トップレベルドメイン).....	54
表 22 http 通信の Host Top10 (セカンドレベルドメイン).....	55
表 23 https 通信の IP アドレス Top10(セカンドレベルドメイン).....	55
表 24 https 通信の IP アドレス Top10(セカンドレベルドメイン).....	56
表 25 機関リポジトリに登録されたコンテンツ利用回数の経年変化.....	58
表 26 機関リポジトリに登録されたコンテンツの登録数推移.....	59
表 27 type 種別別の経年変化(登録数)	60
表 28 利用回数上位 1,611 件における学位論文の経年変化(利用回数).....	61
表 29 機関リポジトリに登録されたコンテンツの記述言語数.....	63
表 30 機関リポジトリに登録されたコンテンツの type 種別コンテンツ数.....	64

表 31	コンテンツ利用上位 15 件	65
表 32	登録年度とコンテンツ数(N=1,611).....	66
表 33	コンテンツの言語比率(N=1,611)	67
表 34	利用回数 520 以上のコンテンツの type 種別(N=1,611)	68
表 35	登録年と登録コンテンツ数(N=17,950).....	70
表 36	コンテンツ記述言語と登録コンテンツ数(N=17,950)	71
表 37	紀要別の収録コンテンツ数ランキング	72
表 38	紀要別のコンテンツ利用数ランキング	73
表 39	紀要利用数の経年変化.....	74
表 40	コンテンツの登録年度と初回利用までの年数	74
表 41	紀要コンテンツ利用上位 15 件.....	75
表 42	利用年度と登録年度のクロス集計(2010 年から 2017 年まで)	76
表 43	登録年と登録コンテンツ数(N=6,993).....	79
表 44	コンテンツ記述言語と登録コンテンツ数(N=6,993)	80
表 45	学術論文の収録雑誌別コンテンツ数ランキング	81
表 46	雑誌論文の総利用回数ランキング.....	82
表 47	年度別コンテンツ利用回数.....	82
表 48	分野別のコンテンツ数.....	83
表 49	学術論文で利用の多いコンテンツ	84
表 50	コンテンツの登録年度と初回利用までの年数	85
表 51	利用年度と登録年度のクロス集計(2010 年から 2017 年まで)	86
表 52	登録年度とコンテンツ数(N=13,131)	89
表 53	コンテンツ記述言語と登録コンテンツ数(N=13,131)	90
表 54	学位論文の利用回数の経年変化	90
表 55	NII Subject の数と登録学位数のランキング比較	91
表 56	学問分野別利用回数ランキング(上位 15 件)	92
表 57	学位別利用回数ランキング(上位 15 件)	93
表 58	学位論文の利用上位 15 件	94
表 59	分野別のオルトメトリクス合計	95
表 60	コンテンツの登録年度と初回利用までの年数	96
表 61	コンテンツの登録年度と初回利用までの年数	96
表 62	利用年度と登録年度のクロス集計(2010 年から 2017 年まで)	97
表 63	利用回数ゼロのコンテンツの実態(N=1,193).....	101
表 64	Research Paper の年度別公開数(N=2,818).....	102
表 65	Research Paper の年度別公開数(公開日 2010 年以前)	102
表 66	Research Paper の年度別公開数(公開日 2010 年以降).....	103

表 67 平均反応時間と平均値まとめ	104
表 68 検索語句と年度	105
表 69 利用者が検索した回数の多い語句(上位 10 件)	106
表 70 利用者が検索した回数の多い語句	107
表 71 各年度に占める検索語句の言語	107

1. 序論

1.1 研究背景

1.1.1 OpenAccess

オープンアクセス(以下:OA)の定義は、学術的資料へのアクセスがインターネットを介して誰でも無料で可能であること、資料を利用することが可能であることである[1]。OA は理念の提唱で終わることなく、実現することが重要なため、実現するための具体的な手法が提案された。それが 2002 年 Budapest Open Access Initiative (以下: BOAI)で示された 2 つの手法である。以下にて詳細な説明を行う Green Road (Green OA)と Gold Road (Gold OA)である[1]。BOAI15 における報告書では、Gold OA と Green OA では、Gold OA に期待が集まっていたとされ、Green OA には期待されていなかったことが示されていた。だが、結果としては予想を裏切る形で、特に日本では大学や研究機関の図書館が中心となり Green OA が進んでいき、その中でも機関リポジトリが予想外に普及したことが示された[2]。

1.1.2 Green OA

Green Road (Green OA)とは、リポジトリを通した OA 実現の手段である。リポジトリとは、デポジトリと近い意味を持ち、保存庫や倉庫、金庫などを表している。デポジトリとリポジトリの違いは、利用に即した形での保存が行われているか否かであると言える[3]。Green OA がリポジトリとして定義されている理由は、OA の定義に従い、インターネットを通しての研究成果を利用可能な形で保存、提供するためである。以下に示す BOAI(2002)における定義では、利用者に負担のない識別子[4]とインターネットの利用が OA の実現に結びついている[1]。

I. Self-Archiving: First, scholars need the tools and assistance to deposit their refereed journal articles in open electronic archives, a practice commonly called, self-archiving. When these archives conform to standards created by the Open Archives Initiative, then search engines and other tools can treat the separate archives as one. Users then need not know which archives exist or where they are located in order to find and make use of their contents.

前述した通り、機関リポジトリは、BOAI においては示された OA 実現に向けた 2 つの手法(Green OA と Gold OA)における Green OA の中の 1 つの手段であった[1]。倉田(2007)は、Green OA における 4 つの代表的な手段を以下の様に紹介した[5]。

①著者のウェブサイト

- ②物理学分野の ArXive を代表とする分野別のプレプリント・サーバ¹
- ③NIH の PubMed Central に代表される政府主導の分野別アーカイブ²
- ④各大学・研究所が設置している機関リポジトリ

Green OA の実現に向けた手段は複数ある。その多くの手段において、中心となる役割を果たしているのは図書館である。図書館と他の組織の連携によって、リポジトリによる OA が進められてきた。一方で Green OA の実現における中心的な組織の機能不全が指摘されており、図書館がアカデミアよりも"象牙の塔"で自覚を持つべきだという指摘もある[2]。図書館を中心として行われてきた Green OA ではあるが、OA の更なる実現には図書館が自らの体質を自覚し、他の組織や研究者との連携を強化することが求められている。

1.1.3 Gold OA

Gold Road(Gold OA)とは、学術雑誌を通した OA 実現の手段である。高額な購読料を支払うことでしか、学術的生産物にアクセスが出来ない状況[13]を改善するためにも、学術論文へのアクセスを保証することを目的とした OA である[1]。その手法には、三根(2007)によると"完全無料型、著者支払い・読者無料型、ハイブリッド型、一定期間後無料公開型、電子版のみ無料公開型"[14]などが存在している。以下に、BOAI(2002)における定義を記載する。[1]

II. Open-access Journals: Second, scholars need the means to launch a new generation of journals committed to open access, and to help existing journals that elect to make the transition to open access. Because journal articles should be disseminated as widely as possible, these new journals will no longer invoke copyright to restrict access to and use of the material they publish. Instead they will use copyright and other tools to ensure permanent open access to all the articles they publish. Because price is a barrier to access, these new journals will not charge subscription or access fees, and will turn to other methods for covering their expenses. There are many alternative sources of funds for this purpose, including the foundations and governments that fund research, the universities and laboratories that employ researchers,

¹ 現在は、ArXive(<https://arxiv.org>)[6]には 8 つの大項目があり、物理は 13 の小項目に分類されている。また、2019 年には運営が移管する予定である。過去には、Los Alamos National Laboratory[7]から the Cornell University Library[8](現在)に移管。次は Cornell Computing and Information Science[9]になる予定である

² PubMed Central は現在の PMC(<https://www.ncbi.nlm.nih.gov/pmc/>)[10]であり National Library of Medicine[11]と National Center for Biotechnology Information[12]が運営している。

endowments set up by discipline or institution, friends of the cause of open access, profits from the sale of add-ons to the basic texts, funds freed up by the demise or cancellation of journals charging traditional subscription or access fees, or even contributions from the researchers themselves. There is no need to favor one of these solutions over the others for all disciplines or nations, and no need to stop looking for other, creative alternatives.

BOAI の以降では、それ以前と比較しても多くの OA 実現のための仕組みや行動が各所で取られた。そして、OA のポリシーに従った行動の多くは良い結果を伴った。だが、2015 年に至るまでの期間において、OA の実現を掲げて実施された活動の問題点も指摘されるようになった。Gold OA の問題として、指摘されたことは、APC-Gold³が新規のビジネスモデルとして運用され、OA のコンセプトから離れた出版社による営利活動になった点である。Gold OA は BOAI の当初は最も期待されていたが、実現には予想された点と異なる経済的な問題が伴っていた[2]。

1.1.4 BOAI10 と OA に向けた日本での活動

OA が、学術情報流通における喫緊の課題であることは、BOAI10(2012)の時点でも強く主張がなされていた[15]。この時における提言で、2018 年時点において日本でも導入されている仕組みやポリシーが示されていた。具体的には、博士論文が印刷公表からインターネットを利用した公表になったことなどが例として挙げられる[16]。他にも、科学研究費助成事業における OA の推進[17]や、一部ではあるが、機関リポジトリの機関別認証評価[18]や大学の評価・自己点検項目への導入[19]などが実際に行われたことが報告されていた。

1.1.5 BOAI15 と OA の実現における問題

BOAI15(2017)の報告書では、2010 年の BOAI10 と同様に OA が未だに喫緊の課題であることが示されていた。だが、BOAI10 とは異なる点として、BOAI の 15 年の振り返りや当時の予測と現在の状況の違い、問題点などが挙げられた。OA が喫緊の課題であることには変わりがないが、OA という概念の認知が確立された現在において、必要な批判がなされた形であった。以下に、BOAI15(2017)におけるレポートから一部を紹介する[2]。

Much has happened, and much of it is positive, but taking stock of what has been achieved has also become an urgent task, if only to get a clear sense of our bearings: while Open Access is now here to stay, it also displays a variety of forms that do not all conform with the project of distributed human

³ APC-Gold(Article Processing Charge)の分類における著者支払い型を指す。

intelligence with which it is associated. Lesser, degraded, forms of Open Access have also and gradually emerged, sometimes as the result of power plays by powerful actors, sometimes out of compromises proposed by people of good will.

BOAI15において、指摘された点でもある Gold OA の問題は、BOAI15(2017)において、狂気と表現されるほどに複雑な学術雑誌を扱う市場と独自の文化にあった[2]。この市場の特徴は、生産者の評価と生産物の公表が結びついており、ヒエラルキーともいえる構造が確立している点である。出版社や図書館、研究者や大学などの利害関係者が深く関わっている。利害関係者が、互いに依存しつつも、評価基準を共有していない点があるため、価格や需要と供給の関係が複雑な関係となっている。一般的に言われる、研究者評価や研究のインパクトを図る際に、「コア」ジャーナルの存在が、研究の世界に与えている影響が大きいという問題はその一つである[20][21][22][23]。以下に詳細を述べるが、研究者評価と学術雑誌の関係が、不当な形での Gold OA を引き起こす原因にもなっている。

学術情報流通と研究者評価という観点では、研究者が高い評価を得るためには、世界的に評価の高い学術雑誌に論文を投稿することが求められることが知られている。そして、投稿した論文の引用文献には、投稿先と同様に信頼性の高い学術雑誌⁴に掲載された論文を記載することが通例である。研究者が高い評価を得るためには、「コア」ジャーナルを読み、引用、論文を投稿することが求められていることになる。その結果として、「コア」ジャーナルとして扱われない中小規模の学術雑誌や、国際的には知られていない国内誌は、参考文献として無視されることもある。「コア」ジャーナルを出版している様な大手の学術出版は、OA が進められても非弾性的な需要を保ち続けているため、影響を受けなかったが、中小規模の出版社は、OA で読める論文が多く存在することによって、大きな影響を受けた。出版社が生き残るための対策として OA のモデルの 1 つである APC⁵を生存戦略として位置付けたという背景がある。研究者は、評価や成果公表という観点から、国際誌として発行された学術雑誌へ料金を支払ってでも論文を載せることを望んでおり、出版社のビジネスと合致した結果、このような問題となった。

以上のように、OA のための手段が新たなビジネスモデルへの転換を果たした。APC を生存戦略として位置付ける中小規模の学術雑誌の出版社の中には、ハゲタカ出版社と呼ばれるものも出現した。現在は、研究者への注意喚起も必要となるほどに、ハゲタカ出版社が出てきており、研究者が騙されないための、リスト化などもなされている。これらの出版社の学術雑誌は、ピアレビューをまともに実施していないなど、論文の質保証の問題などを引き起こしており、OA ジャーナル全体の問題でもある[24][25][26][27][28]。

⁴ 学術雑誌において、評価の高いは信頼性の高いと近似することができる。著名な論文を引用していることが、論文の信頼性に繋がることもある。

⁵ Article Processing Charge の略であり、[14]で紹介した著者支払い型を指す。

BOAI15(2017)でもこの事実が問題視されており、以下に引用して示す様に、Gold OA 実現のための手段でなく、出版社による新たなビジネスモデルであると評した[2]。

In particular, the publishers' promotion of Open Access, presented as Gold, but limiting itself to APC-Gold, represents a clear indication that the battle around Open Access has been reshaped in a new way: does OA deal with a communication system that can help science, or is it a business model masquerading as OA designed to reinforce the position of some publishers.

まとめると、Gold OA は、BOAI の段階から出版社の協力を得て、ビジネスの機会として扱われていたことは、事実である。だが、これほど問題となっているのは、ハゲタカ出版社による新規の OA ジャーナルの乱立や、査読が行われないことによる論文の信頼性の低下が背景にある。新規の学術雑誌などは、研究者間での評価なども存在しないため、より「コア」ジャーナルへの信頼が高まり、OA ジャーナルの信頼性を証明するのが困難になっている。これは、一部の出版への依存を避けるための活動の結果としては、悪い結果といえるので、今後も注意喚起と正しい理念の下での OA ジャーナルの運用が望まれている。

1.1.6 日本での Green OA 実現の問題点

Green OA の問題を考える際に、考慮する必要があるのは、上記でも示したように Green OA のための取り組みの中心的な組織である図書館である。機関リポジトリの定義にて詳細を述べるが、arXiv[6]や PMC[10]に代表される機関リポジトリ以外のリポジトリに関しても運営元である図書館の影響は考慮が必要である。だが、運営元である図書館は、"象牙の塔"であり、教員や大学との連携がうまくなされておらず、科学・学術コミュニケーションの促進のための機能を発揮できていないことが指摘されていることを上記でも示している[2]。

OA 全体としての問題である大学図書館と研究者のコミュニケーションなどは日本においても問題とされている。日本の Green OA に向けた問題としては、山地(2017)は、"ポリシーが先行しており、内実が付いて行っていない"[29]ことなどをワークショップで指摘していた。

また、学術情報流通に関わる中心的な組織の一つである大学図書館が、大学という組織の一部であることから、個別の大学という単位の影響を受けている点も問題とされている。具体的には、大学図書館は、大学評価の項目になっている部分もあり、それぞれの大学内での位置づけや、運営方針が存在している。そのため、個別の大学評価のために大学図書館の差別化が意識されており、大学間のネットワーク化や協働が図られていないことも問題とされている[2]。日本の大学図書館も大学評価に図書の数などが関わっているため

に、電子化や除籍などが適切に行われていないという問題や増え続ける資料による空間の狭隘化の問題を抱えている現状がある[30][31][32][33]。

リポジトリが研究におけるインフラストラクチャとしての機能を発揮するためには、大学の組織としての位置付けや研究者との協働、コミュニティとしての役割などを果たすことが必要になる。そして、インフラストラクチャとしてあり続けるには、現状を評価し役割を明確にするためには評価が欠かせない。リポジトリの評価という点において注意すべきところが、研究者評価に用いられたインパクトファクターの誤用の様に数字のみを見て、本質的な視点を見失う危険性があることである。リポジトリの図書館における評価は必要であるが、それは利用者の関心を見つけ、改善するためのものであり、文献の評価とは異なる文脈であるということに留意が必要である。

一見すると文献の評価として判断される危険性がある事例としては、JAIRO Cloud を利用した機関リポジトリにおけるランキング表示が挙げられる。ランキング表示は、リポジトリのトップページの横に存在している。

最も閲覧されたアイテム	
3591	<原著>ローゼンバーグ自尊感情尺度日本語版の検討
2997	現代人における年中行事と見出される意味：東方巻を事例として（〈特集2〉現代社会と民俗）
2505	幕末から明治前期の油画技法材料に関する研究：油画媒剤を中心に
最もダウンロードされたアイテム	
42866	<原著>ローゼンバーグ自尊感情尺度日本語版の検討(9.pdf)
18520	水中に懸濁した微粒子の凝集分散-基礎理論とその適用性(JJCTA_45-11.pdf)
17565	直腸がん前方切除術後の排便障害を軽減する看護支援に関する研究(DB02696.pdf)

図 1 つくばリポジトリにおけるランキング画面(accessed-2018-12-25)[34]

上記の図を見ると、リポジトリに実装されているのは、コンテンツの利用ランキングに見える画面である。だが、このランキングは、資料のタイプやクラスタを分類したあとのコンテンツのランキングを示すものでない。あくまで、期間内における文献の利用動向を示す指標である。一見するとリポジトリにおける文献の評価を表す指標に見えるが、利用動向を示す指標であることを踏まえてみる必要がある。各機関におけるリポジトリの評価や特徴を表す様な指標は見られない。

本研究では、このような現状を踏まえ、次世代に求められる OA の手段としてのリポジトリの推進のために、現状のリポジトリのコンテンツ利用実態をアクセスログから分析する。本研究におけるコンテンツの扱いに関しては、直接的な評価対象ではなく、評価のための利用実態を明らかにするための材料である。リポジトリの利用統計をコンテンツの評価という指標だけでなく、コンテンツの所属しているクラスタや機関リポジトリの評価の

一助となるように、研究を進めた。

1.1.7 歴史的背景

BOAI(2002)において OA 実現の目的は、4 点挙げられていた[1]。佐藤ら(2010)によると "1)研究の加速, 2)教育の質向上, 3)富める者と貧しい者の学問の共有, 4)人類共通の知的対話と知識探究の場の実現"[35]である。

Suber(2009)によると OA 運動は 1960 年代から活動が行われてきた[36]。OA 運動の歴史の中で BOAI(2002)は、OA の重要な転機として知られている[15][37]。1960 年代から行われてきた諸々の運動を 1 つにまとめたことが BOAI の功績とも呼べる点だった。

OA が推進されることとなった背景の部分における代表的な例は、2 つ挙げられる。

1. 学術雑誌価格高騰

2. 研究成果の迅速な共有(査読や公開にかかる時間の問題)

1.における問題は、学術論文数の増加、学術雑誌の評価による「コア」ジャーナルの需要の増加、非弾性的な需要など要因が影響したと考えられる。特に、自然科学系のその研究分野においては、必読ともいえ、投稿先としても評価の高い学術雑誌が存在する。大学や研究所にある分野において必読の学術雑誌をそれぞれの組織が購読することで、需要が増加する。そして、その需要は安定しているので、出版社が価格を向上させても購入を続ける必要がある。一部の学術雑誌の価格が上昇すると適正価格も変化する。学術雑誌の価格向上は現在も続いており、問題となっている。2002 年周辺では、発展途上国に対する支援を出版社などがすることになったが、根本的な学術雑誌の価格向上という問題は現在も続いている。

2. の研究成果の共有は、学術雑誌の公開にかかる時間と査読の問題があった。学術コミュニケーションにおいて、フォーマルと呼ばれるコミュニケーションは、学術雑誌に掲載された論文を中心に行われる[38]。だが、学術論文数の増加や、査読などがあるので、学術雑誌に論文が投稿されてから、掲載に数ヶ月から 1 年程度⁶の期間が必要となる。

これらの問題に共通する部分は、学術論文数の増加と学術コミュニケーションにある。学術論文の数は年々増加を続けている。以下に、論文量の変化のグラフを示す[39]。

⁶ 投稿する学術雑誌によって、掲載までにかかる時間には差がある。査読の有無や学術雑誌の出版頻度によって、大きく異なる。

【図表4-1-1】 全世界の論文量の変化

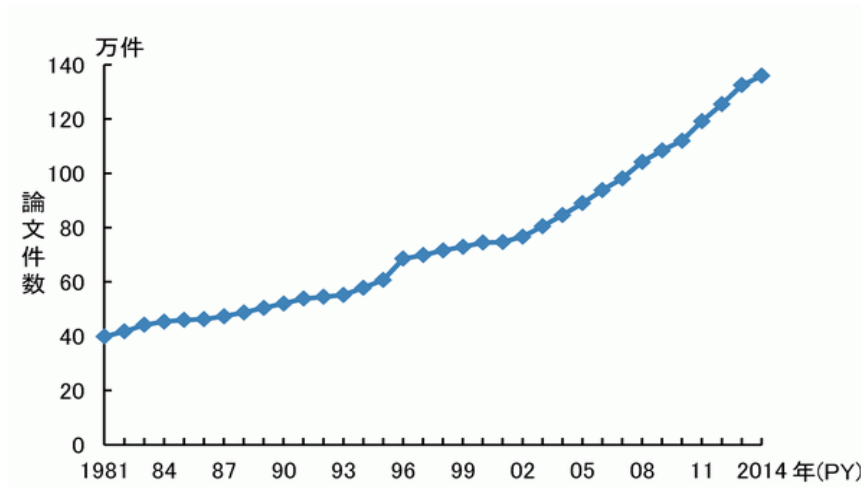


図 2 世界の論文件数の変化(accessed2018-18-18)[39]

そして、学術情報流通におけるフォーマルコミュニケーションの「場」[40]とも呼べる学術雑誌は、非弾性的な需要を持つ。そして、学術コミュニケーションの問題として、「トップ」ジャーナルと呼ばれる雑誌の存在がある。学術雑誌には、一種にヒエラルキーが存在することが知られている[40][41]。そして、評価の高い学術雑誌に論文を投稿することが研究者の評価に繋がっている。大学図書館によって、購入がなされる学術雑誌も、全ての学術雑誌を購入することはできないため、重要性和予算によって購入される雑誌が決まっていた。その際に、参考となる様に開発された指標がインパクトファクターであり、Journal Citation Report[42]という形で公開、利用されてきた[43]。

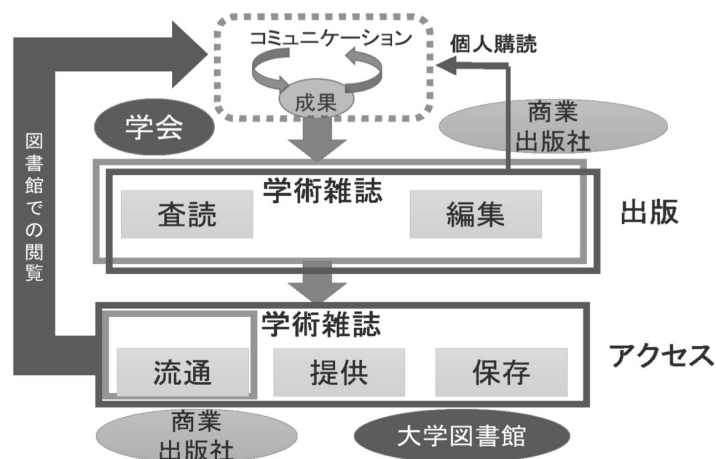


図 3 学術コミュニケーションのモデル(access2018-12-18)[40]

学術雑誌は、閉じたコミュニティにおけるフォーマルコミュニケーションの「場」であり、大学や研究所が購入する雑誌の重要度を求めるための指標が存在している。そして、雑誌として評価の高い論文誌が、「コア」となるジャーナルと呼ばれている。評価の高い雑誌に関する需要と供給の関係や、一部の出版社による寡占状態などを背景にして、雑誌の価格の高騰が続き、シリアルズクライシスと呼ばれる問題となった[13]。これが、1.雑誌価格高騰の背景にあった問題である。この雑誌の価格の高騰は現在も続いている。そして、インターネットが普及した 1990 年以降は、学術雑誌の優位性は変わらずとも、学術雑誌の提供としての電子ジャーナルの普及や学術コミュニケーションの「場」が広がった。(もともと電子ジャーナルは 1970 年に試行はされていた[40][43][44]、学術コミュニケーションの「場」として手紙が知られている。チャールズ・ダーウィン[46]の例が有名である。)

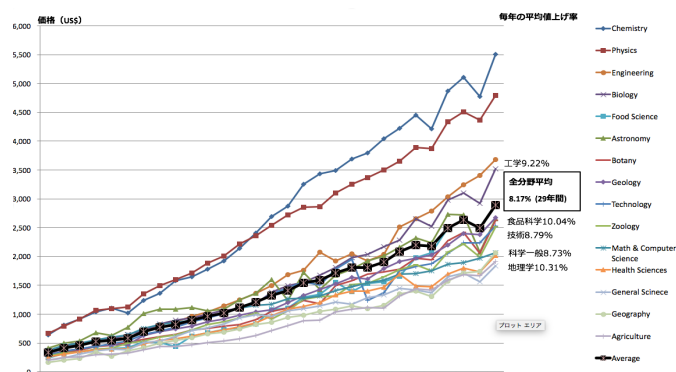


図 4 自然科学系学術雑誌の価格推移(accessed 2018-12-18) [47]

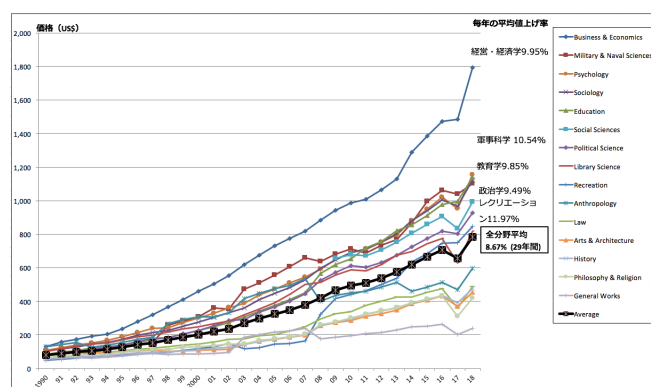


図 5 人文・社会科学系学術雑誌の価格推移(accessed 2018-12-18) [47]

インターネットの普及によって、インフォーマルコミュニケーションの「場」が広がり、研究者同士のコミュニケーションが活発になっても、フォーマルコミュニケーション

の「場」は出版社によって、囲い込みがなされている状況は変化していない。そこで、BOAI では、OA の定義と目的が示されることとなった。

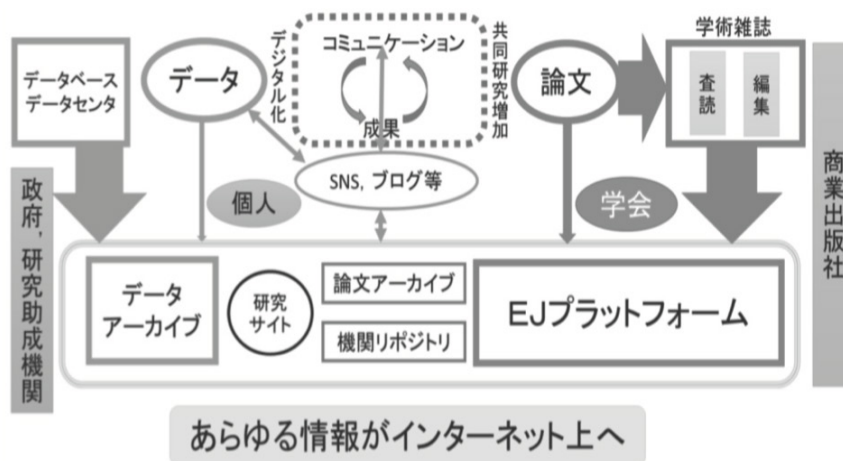


図 6 新しい学術コミュニケーションのモデル(accessed 2018-12-18)[40]

1.1.8 世界の機関リポジトリ数と日本の機関リポジトリ数

本研究では、OA 実現の手段の中でも機関リポジトリに着目している。日本の機関リポジトリは、OpenDOAR(Open Directory of Open Access Repositories)に掲載されているリポジトリ数が 222 と国別に見ると世界で 4 番目に多い[48]。(登録リポジトリ数 3,803)Bielefeld Academic Search Engine(以下 BASE)を参考にすると機関リポジトリ数は、567 となり国別に見ると世界で 3 番目に多い。また、BASE に登録されているコンテンツ数は、2,623,872 と国別に見ると世界で 11 番目に多い[49]。(登録リポジトリ数 6,726、登録コンテンツ数 131,704,743) 国立情報学研究所(National Institute of Informatics 以下:NII)にハーベスティングを行なっているリポジトリ数が 691 存在しており、登録されているコンテンツ数が 2,915,982 であることから日本の機関リポジトリ構築数が世界有数であり、コンテンツも充実している部類に分類することが可能であると言える[50]。BASE 内で、日本の機関リポジトリを確認するとコンテンツが OA であることを明確にしている機関は、非常に少ない。567 機関のうち 76⁷機関のみが OA であると明確に表記されているコンテンツを所持していることになる。これは、OA であることの表記の問題であると考えられる。

⁷ リポジトリの数は 77 件ヒットするがメタデータの付与が間違っていることが確認できた 1 件を筆者が除外した。

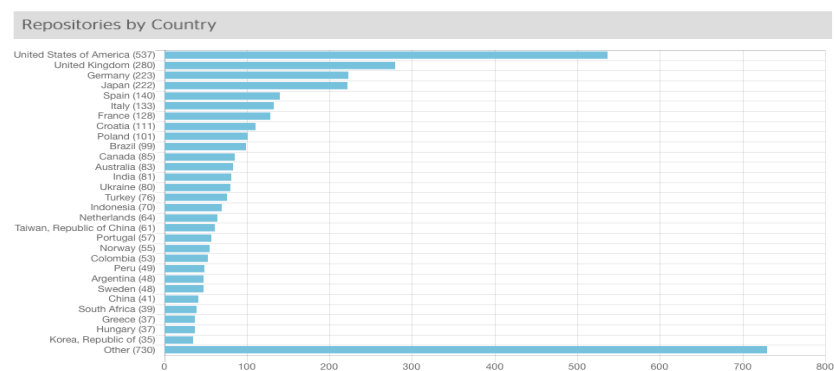


図 7 OpenDOAR に掲載されている機関リポジトリ数(accessed 2018-12-19)[48]

Country ▲ ▼	Content sources ▲ ▼	Documents ▲ ▼	Feed
United States of America	1,091	45,216,703	RSS ATOM
Germany	394	20,953,204	RSS ATOM
France	195	10,266,557	RSS ATOM
International Organisations	44	7,168,310	RSS ATOM
Spain	252	5,212,301	RSS ATOM
United Kingdom	317	5,116,871	RSS ATOM
Italy	160	4,248,646	RSS ATOM
Poland	112	3,400,219	RSS ATOM
Australia	89	2,913,660	RSS ATOM
Switzerland	30	2,785,095	RSS ATOM
Japan	567	2,623,872	RSS ATOM
Brazil	307	2,389,335	RSS ATOM
Russia	183	2,255,439	RSS ATOM
Indonesia	750	1,747,005	RSS ATOM
Belgium	23	1,737,310	RSS ATOM

図 8 BASE に掲載されているドキュメント数上位 15 ヶ国(accessed 2018-12-19)[49]

Country ▲▼	Content sources ▲▼	Documents ▲▼	Feed
United States of America	1,091	45,216,703	RSS ATOM
Indonesia	750	1,747,005	RSS ATOM
Japan	567	2,623,872	RSS ATOM
Germany	394	20,953,204	RSS ATOM
United Kingdom	317	5,116,871	RSS ATOM
Brazil	307	2,389,335	RSS ATOM
Spain	252	5,212,301	RSS ATOM
France	195	10,266,557	RSS ATOM
Russia	183	2,255,439	RSS ATOM
Canada	172	1,723,420	RSS ATOM
Italy	160	4,248,646	RSS ATOM
India	149	622,529	RSS ATOM
Colombia	119	471,911	RSS ATOM
Ukraine	118	719,024	RSS ATOM
Peru	116	203,629	RSS ATOM

図 9 BASE に掲載されているリポジトリ数上位 15 ヶ国(accessed2018-12-19)[49]

表 1 BASE に掲載されていたうち、OAmark 率が 100%のリポジトリ 20181219[49]

リポジトリ名(日本語表記抜粋)	OAmark 率	登録コンテンツ数
愛知淑徳大学 知のアーカイブ(リポジトリ)	[100%]	2,243
愛知工業大学学術情報リポジトリ	[100%]	2,786
福岡教育大学図書館	[100%]	1,885
自治医科大学機関リポジトリ	[100%]	404
京都女子大学学術情報リポジトリ (京女 AIR)	[100%]	2,277
長岡技術科学大学学術情報リポジトリ	[100%]	734
大手前大学大手前短期大学リポジトリ	[100%]	1,759
熊本県立大学学術リポジトリ	[100%]	1,129
立正大学学術機関リポジトリ	[100%]	5,628
成蹊大学学術情報リポジトリ	[100%]	922
滋賀大学学術情報リポジトリ	[100%]	12,949
日本歯科大学学術機関リポジトリ	[100%]	787
筑波技術大学機関リポジトリ	[100%]	1,246
和歌山大学学術リポジトリ	[100%]	3,091

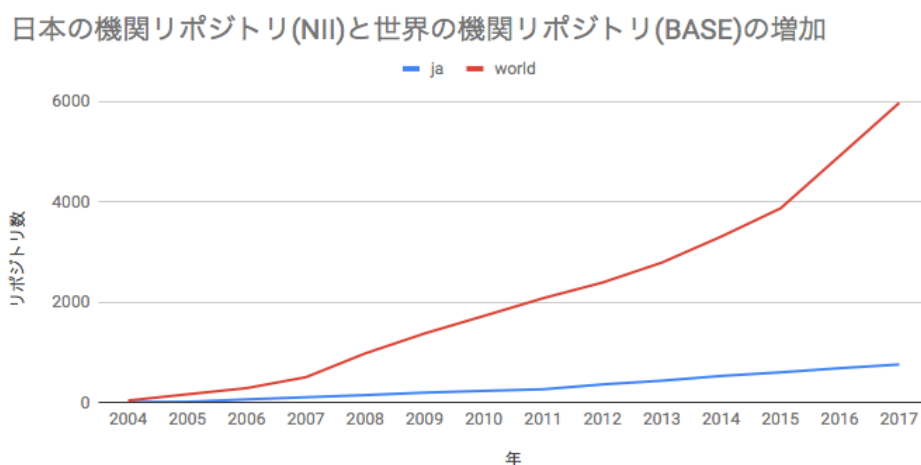


図 10 日本の機関リポジトリ数と世界の機関リポジトリ数の推移(NII と BASE を参照) [49][51]

表 2 日本の機関リポジトリ数と世界の機関リポジトリ数の推移(NII と BASE を参照)20181219 [49][51]

year	2004	2005	2006	2007	2008	2009	2010
ja	2	10	58	101	144	193	228
world	35	160	283	502	977	1,374	1,724
year	2011	2012	2013	2014	2015	2016	2017
ja	260	357	431	526	598	681	754
world	2,078	2,390	2,788	3,307	3,870	4,920	5,974

1.1.9 日本の機関リポジトリ

日本の機関リポジトリのコンテンツの特徴として挙げられるのは、紀要である。大学の紀要は、活動報告や手記、学生の論文から、査読付き論文まで紀要という種別の中でも、雑誌ごとに大きく特徴が異なる。そのため、紀要という名称を持っても、ひとまとめにして、断定することが困難である[52][53]。そのため、日本の機関リポジトリに登録されているコンテンツ種別で一番数が多いとされているが、紀要をリポジトリに登録し、公開することは、OA 本来の目的である学术论文の公開として一概に分類することができないものとなっている。NII がハーベスティングしたコンテンツの内訳を見ることができる IRDB を参考にすると、日本の機関リポジトリに掲載されているコンテンツの 53.2%が紀要である[50]。また、学术论文は、13.7%と割合が低く、日本の機関リポジトリを揶揄して、紀要リポジトリと表現されることもある[29]。

紀要の掲載は、BOAI で示された学術論文の OA という側面とは少し異なるが、良い面もある。論文として発信されにくい分野における研究の社会への公開や、税金を用いた研究の成果公表の義務を果たす媒体としての側面がある。また、灰色文献でもあることから大学としての特徴が出る部分でもある。特に人文系の分野など論文として成果が評価しにくいとされている分野では、その意義が大きい[54][55]。研究者のアウトリーチ活動としての説明責任を果たす「場」としての役割を果たしているとも言えるためである。

1.1.10 日本国内における機関リポジトリに関する政策

部科学省の web サイトにおいて、“学術情報流通” OR “機関リポジトリ”で検索し、適合した資料 115 件のうち関係する 5 件を引用し、報告する。検索結果の表示がコンテンツページ毎で表示されたことや、議事録が多く存在したことから、この件数となった。2009 年以前のものに関しては富本(2009)を参考に 3 件の文献を得た[56]。以下に、学術情報流通と機関リポジトリに関わる政策を表の形式で掲載する。

表 3 学術情報流通と機関リポジトリに関する政策資料

時期	文部科学省のサイトに掲載されている資料
2002.03	学術情報の流通基盤の充実について(審議まとめ)[57]
2003.03	学術情報発信に向けた大学図書館機能の改善について(報告書)[58]
2006.03	学術情報基盤の今後のあり方について(報告)[59]
2009.03	オープンアクセスに関する声明[60]
2010.02	第 4 期科学技術基本計画における情報通信分野の重点事項について[61][62]
2010.07	平成 21 年度文部科学白書[63]
2010.12	大学図書館の整備について(審議まとめ)-変革する大学にあって求められる大学図書館像-[64]
2014.08	大学等におけるジャーナル環境の整備と我が国のジャーナル発信力強化の在り方について[65]

1. 2009 年 3 月 オープンアクセスに関する声明[60]

1990 年代からのインターネットの普及と BOAI を背景にした国立大学図書館協会による呼びかけであり、大学図書館のみならず、その他の利害関係者への呼びかけが含まれている。以下に引用するのは、大学図書館への声明である。

大学図書館へ

1. 図書館利用者やその他の関係者に対し、オープンアクセスへの支持と協力を呼びかけると共に、教員や研究者と協力してオープンアクセスを推進する。
2. オープンアクセス及び大学からの研究成果の発信源として、学術機関リポジトリの整備に努める。

2. 2010 年 2 月 第 4 期科学技術基本計画における情報通信分野の重点事項について[61]

内閣府の政策である第 4 期科学技術基本計画[62]の策定のための審議における重点領域を紹介している。機関リポジトリに関する内容は、2 分野において登場した。以下に、機関リポジトリに関する内容のみを抜粋した。

多様な分野の研究や産業競争力の基盤

《解決すべき課題（社会的・技術的ニーズ等）》

・我が国の研究開発を一層活発なものとするとともに、国際的にもより開かれたものとする観点から、人類にとっての共通の知的財産である学術研究成果の内容を、必要とする全ての人がアクセスできるようなオープン・アクセスを一層推進させることが必要。

《重点的に推進すべき研究領域・課題等》

- ・最先端学術研究を支える学術情報ネットワークなどの情報通信基盤の整備・充実や、それらを維持する体制の強化
- ・研究に不可欠なサイエンスデータベースや学術コンテンツ等の整備や電子化
- ・学術研究成果に関するオープン・アクセスの推進や機関リポジトリの充実

多様な分野の学術研究や産業競争力の基盤となる情報科学技術の研究開発の促進

- ・e - サイエンスを支える最先端の研究情報基盤の構築技術、基盤の整備

《解決すべき課題（社会的・技術的ニーズ等）》

我が国の研究開発を一層活発なものとするとともに、国際的にもより開かれたものとする観点から、人類にとっての共通の知的財産である学術研究成果の内容を、必要とする全ての人がアクセスできるようなオープン・アクセスを一層推進させることが必要。

《国際動向》

・欧米においては、E - サイエンスが研究開発方法論の大変革をもたらすものとして重要視され、2005 年以降、国家レベルの最先端の CSI^{8*} (Cyber Science Infrastructure) 構築の計画が開始。欧州では EU がリーダーシップを発揮し、E - サイエンス向けのグリッド基盤を構築。米国では、米国科学財団 (NSF) が中心となって TeraGrid プロジェクトを推進。わが国においては、国立情報学研究所 (NII) を中心とした CSI 構想が推進されており、NAREGI 等の実運用に基づく研究グリッドなどが推進されており、今後さらなる応用研究が必要。

・学術論文のオープン・アクセス運動については、欧米では政府、研究助成機関等が直接的に学術情報流通に関与をはじめたことを契機に関心が集まりつつある。国内では、欧米の学術出版社からライセンスを受けた電子ジャーナルの価格高騰問題が顕在化しており、欧米の学術情報の確保という観点からも、解決に向けた早急な検討が必要。さらに、文献情報のオープン・アクセス化という国際的潮流を踏まえた対応も喫緊の課題。

《重点的に推進すべき研究領域・課題等》

・コミュニティの形成や実験プロセスの支援など、より広い範囲でサイエンスのアクティビティを支援するグリッド関連技術をはじめ、クラウドのようなシステムに対応する仮想化、分散化などのアーキテクチャ検討

・e - サイエンスや最先端学術研究を支える学術情報ネットワークなどの情報通信基盤の整備・充実や、それらを維持する体制の強化

・研究に不可欠なサイエンスデータベースや学術コンテンツ等の整備や電子化、オープン・アクセスの推進及びそのための国レベルの総合的な体制作り

・機関リポジトリの充実を通じた学術情報の速報的な発信力の強化

3. 2010 年 7 月平成 21 年度文部科学白書[63]

第 2 部文教・科学技術施策の動向と展開の第 5 章科学技術・学術政策の総合的推進の一項目として挙げられている。機関リポジトリの定義について、保存書庫として紹介されているのが特徴である。

知識基盤の整備

・研究情報の発信・流通の促進(デジタルコンテンツの整備)

科学技術振興機構では、学協会の学会誌・論文誌における論文の投稿から査読、審査、公開までの工程を電子化して行う科学技術情報発信・流通総合システム

(J-STAGE) を運用しています(参照：<http://www.jstage.jst.go.jp/browse/-char/ja>)。また、重要な学協誌の電子アーカイブ化(紙媒体の情報の電子的な記

⁸ 機関リポジトリに関する CSI 事業は、NII を中心に 2005 年から実施されていた[61][62]。

録・保存)を行い、インターネットを通じて広く世界に発信・流通しています
(参照：<http://www.journalarchive.jst.go.jp/japanese/>)。 情報・システム研究機構国立情報学研究所では、多種多様な学術情報を統合的に利用できる「学術コンテンツ・ポータル (GeNii)」⁹の運用や我が国の学術雑誌の国際流通を促進する「国際学術情報流通基盤整備事業」の実施などにより、学術情報流通基盤の整備を推進しています(参照：<http://www.nii.ac.jp/services/service-j.shtml#06>)。また、各大学における機関リポジトリ (※33) の構築と連携を支援するとともに、我が国の学術機関リポジトリに蓄積された情報を検索できるポータルサイト JAIRO を提供しています(参照：<http://jairo.nii.ac.jp/>)。※33 機関リポジトリ 大学及び研究機関等における教育研究活動によって生産された電子的な知的生産物を保存し、原則的に無償で発信するためのインターネット上の保存書庫。

4. 2010 年 12 月 大学図書館の整備について(審議まとめ)-変革する大学にあって求められる大学図書館像-[64]

大学図書館の機能・役割及び戦略的な位置付けにおいて、大学図書館の基本的な機能を確認すると共に、変化する学術情報流通における学術情報流通への対応と、研究支援の文脈において、機関リポジトリが位置付けられた。また、大学図書館だけでなく、大学図書館員の育成についても述べられていた。大学のインフラとしての構築から運用へと意識が向いている。

電子化の進展と学術情報流通の変化

大学図書館では、従来、所蔵する図書、雑誌等に関する情報を OPAC (オンライン蔵書目録) として組織化してきた。また、NII と協同して NACSIS - CAT / ILL を構築し、活用することによって、自館にない資料でも効率よく探すことのできる仕組みを実現し、学術情報基盤として運用してきた。しかしながら、大学図書館以外の学内施設が所蔵する資料、機関リポジトリのデータ、また、インターネット上の学術情報などについては必ずしも統合的に大学図書館が扱っているわけではない。これらの多様な媒体や形式で提供されつつあり、大学図書館がこれらの学術情報の収集、蓄積、提供に適切に対応していくことが課題となっている。

一方、サーチエンジンをはじめ、学術論文に関するデータベースや主題等に特化した書誌データベースなど、学術情報を検索し、アクセスを支援するためのサービスがインターネット上に数多く存在しているが、その収録範囲、提供される

⁹ Ginii(<https://ge.nii.ac.jp/genii/jsp/index.jsp>)は、現在、アクセスを試みるとブラウザから警告がなされる。完全に停止している。[68][69]

情報の質や種類は多様であり、これらを大学図書館機能の中にいかに組み入れていくかが重要な課題となる。

さらに、NII が学術系コンテンツサービスの強化を図ってきた結果、論文情報ナビゲータ (CiNii) は、日本語学術文献の検索、フルテキストデータの提供には不可欠のものとなっており、近年になってから提供されている電子ジャーナルリポジトリ (NII - REO)、学術機関リポジトリポータル (JAIRO) なども着実に普及している。これらはサイバー・サイエンス・インフラストラクチャ (CSI) 構想の一環でもあり、大学図書館における新たな課題も踏まえて、各種事業等のさらなる展開の検討が期待される。

研究活動に即した支援と知の生産への貢献

大学等において構築されている機関リポジトリは、研究者自らが論文等を掲載していくことにより学術情報流通を改革するとともに、その公開の迅速性を確保するものである。それと同時に、大学等における教育研究成果の発信を実現し、社会に対する教育研究活動に関する説明責任の保証や、知的生産物の長期保存などを図る上でも、大きな役割を果たすものである。

我が国においては、現在、NII と大学等との連携により、130 件を超える機関リポジトリが構築されている。国立大学の 8 割を超える機関リポジトリを構築していることになり、収録コンテンツ数 (全文情報) は全体で 70 万件を超えている。現状において、大学内で刊行されている紀要の電子化を実現している例が多いが、それ以外にも機関リポジトリの展開には次のような方策が考えられる。即ち、1. 大学で使われる教科書をオープンアクセスとして提供する、2. 学位論文の収集と電子的な公開のためのプラットフォームとして活用する、3. 研究者の研究データの蓄積、共有システムとして活用する、などである。

今後、各大学等において構築したリポジトリを継続して運営していくためには、大学全体におけるリポジトリ事業の位置付けの明確化、大学図書館業務としての定着、システムの構築と維持体制の整備などが課題である。

さらに、電子ジャーナルの導入や機関リポジトリの整備などが進む中で、論文などの学術研究成果にオンラインにより制約なくアクセスできることを 理念とするオープンアクセスを推進する必要がある。

大学図書館の現場における育成

大学図書館職員が教育課程の企画・実施を行う際には、教員はもとより大学図書館以外の部署との連絡と調整を密に行う必要がある。したがって、その育成のためには、大学図書館から外に出て、教員や他の部署と円滑にコミュニケーションをとることができるような環境作りが不可欠である。例えば、教育課程担当の教

員と意見交換を行ったり、研究科長や学部長と定期的に会合を持ったり、学生との懇談会を学生がよく集まる場所で実施したり、機関リポジトリの学内への周知のために教員研究室を訪問するなどが挙げられる。

5. 2014 年 8 月 大学等におけるジャーナル環境の整備と我が国のジャーナル発信力強化の在り方について[65]

研究成果の発信の場が海外のコアジャーナルに依存している現状や、流通における税など予算に直結する問題からの危機意識を基に、日本国内における学術論文の利用環境整備に向けた現状の確認と提言のための資料である。BOAI15 においても問題視される APC-Gold や、オープンデータについての記載がある。

ジャーナルの利用状況と価格情報への対応 (オープンアクセス化の動き)

特に、ジャーナル契約見直しのセーフティネットとして、論文を機関リポジトリで公開するオープンアクセスを全国的に推進し、大学等が知識を共有できる環境を創出する意義は大きい。各図書館は、機関リポジトリの運用のみならず、登載するデータの生成、蓄積、提供に必要な環境の整備や学術情報流通に携わる人材育成等、積極的に関与することが望まれる

オープンアクセスの推進 (オープンアクセスの拡充方法)

- ・オープンアクセスを促進するための具体的な方策としては、各大学等における機関リポジトリ構築を更に拡充するとともに、オープンアクセスジャーナルの育成にも努める方法が妥当である。
- ・オープンアクセスジャーナルの育成に関しては、信頼に足る査読制度の構築、科研費における研究成果公開促進費の充実、我が国の公的支援による出版プラットフォームである J-STAGE の強化により、学協会等が協同して、質が高く、また、研究者が負担する APC を低額に抑えることが可能なオープンアクセスジャーナルを構築することが望ましい。

上記に、政策面を確認した。日本における政策的な影響は 2002 年頃から意識されている。海外の事例を意識した CSI 事業などは、海外において政府の強い影響を受けて成功した事例[70]として紹介される様に、OA の考え方は、普及している。だが、国内での学術情報流通の問題点として、論文の生産や研究発表の場に関してはコアジャーナルを依存しており、独自のマーケットを築くことができていないことなどを含めた研究の発信力の問題を抱えている。

1.1.11 日本国内における機関リポジトリに影響する法制度

日本だけに限った問題ではないが、機関リポジトリや OA においては、著作権の処理に関する議論が存在する。学術雑誌に論文として掲載されたものを OA 化する際には、出版社との著作権の確認作業が必要になる。プレプリントやポストプリントなどと呼ばれる論文の版の違いで権利が変わる場合もあり、各出版社が権利処理に関する表記を行っている場合も多い[71][72][73][74]。また、研究が市民の税金を利用した行為であり、論文は成果物であるといった観点から、研究成果の還元[75][76]といった問題も存在している。2000 年以降に改定し、今後も改定する予定となっている 2 つ法制度について述べる。

1. 学位規則の改正[77]

平成 25 年に改正となった学位規則とは、学校教育法の第 68 条の 2 に基づいて定められた文部科学省令である。平成 25 年 3 月 11 日に公布され、平成 25 年 4 月 1 日から施行された。博士論文やその要旨の印刷公表が義務付けられていたのに代えて、インターネットを利用した公表が義務付けられた。文部科学省の想定しているインターネットを利用した公表とは、"ここにおけるインターネット公表とは、各大学の機関リポジトリによる公表を基本とする。"[78]という文言からもわかる通り、機関リポジトリによる公開義務化とも言える形式が取られた。

2. 著作権法の一部改正[79]

著作権法とは、著作権に関する日本の法律であり、第一章から第八章と附則が存在している。平成 30 年法律第 30 号が"著作権の一部を改正する法律"として公布され、平成 31 年 1 月 1 日が施行期日となっている[80]。OA に関係する内容としては、"第 30 条の 4、第 47 条の 4、第 47 条の 5 等関係 ①デジタル化・ネットワーク化の進展に対応した柔軟な権利制限規定の整備"や"②第 35 条等関係 教育の情報化に対応した権利制限規定等の整備"や、"第 31 条、第 47 条、第 67 条等関係 ④アーカイブの利活用促進に関する権利制限規定の整備等"¹⁰が関係すると考えられている。①や②に関しては、機関リポジトリに掲載可能なコンテンツの充実や情報源としての学術情報資源の活用に繋がると考えられる。④に関しては、広く図書館という観点で見た際の、コンテンツの充実や、Hathi Trust[82]などの電子図書館事業に寄与すると考えられる。

1.1.12 日本国内における機関リポジトリの変遷と今後の位置づけ

日本における最初の機関リポジトリは、2002 年の千葉大学が構築し運用したことが知られている[83]。その後、NII による「次世代学術コンテンツ基盤共同構築事業」[84]として、複数の大学への導入が進んだ。2000 年代前半は、NII の支援を受け各大学が独自に機関リ

¹⁰ ①②④の番号の付与は文部科学省による、著作権法の一部を改正する法律の概要に準じた[81]。

ポジトリを構築していった。デジタルリポジトリ連合(DRF)という組織が設立され、機関リポジトリ構築や運営に関する知識を共有していた[85][86]。

日本の機関リポジトリの転機は、2012 年から NII によってサービスを開始した JAIRO Cloud の稼働にある。JAIRO Cloud は、当時としては最新のクラウドベースのサービスであり、予算や技術的な問題で機関リポジトリを独自に構築できない大学でもサービスの利用が可能なものであった[57][87]。また、2012 年 5 月からは、筑波大学における実験運用を経て、既存の機関リポジトリからの移行[88]などもなされ、利用可能な組織も増加した。

2018 年時点で、日本の大学数は、783 大学ある[89]。そのうち、機関リポジトリを構築している大学数は、536 大学である[89]。(2010 年 3 月の機関リポジトリ構築数は 193)日本において機関リポジトリはインフラストラクチャとしての立ち位置にある。そのため、今後は OA からオープンデータに対応したデータリポジトリへの機能拡張などが可能性として示唆されている[90][91]。

1.2 用語の定義

1.2.1 Open Access とは

Open Access(以下: OA)の定義として広く知られているものは、2002 年に行われた Budapest Open Access Initiative(以下: BOAI)で公開された以下のような定義である[1]。

公衆に開かれたインターネット上において無料で利用可能であり、閲覧、ダウンロード、コピー、配布、印刷、検索、論文フルテキストへのリンク、インデクシングのためのクローリング、ソフトウェアヘデータとして取り込み、その他合法的目的のための利用が、インターネット自体へのアクセスと不可分の障壁以外の、財政的、法的また技術的障壁なしに、誰にでも許可されることを意味する。

上記の定義の重要な部分は、学術的資料へのアクセスがインターネットを介して誰でも無料で可能であり、かつ、資料を利用することが可能であることと言える。この定義の背景には、学術雑誌価格の上昇や発展途上国の研究支援などが挙げられている。論文や学術的生産物へのアクセスを必要としているにも関わらず、論文や学術的生産物にアクセスできない事態への危惧から OA の推進が求められている。

1970 年代からの学術雑誌の価格高騰をシリアルズクライシス[13]と呼び、1990 年代からは大学図書館を中心として、雑誌の購読価格交渉に関する活動や電子ジャーナルの導入などが行われてきた[92][93]。雑誌価格の高騰への対応としては、包括契約[94]などが行われてきており、日本にも JUSTICE[95]と呼ばれるコンソーシアムがある。雑誌の価格高騰と研究資源の提供、科学の公衆利用の環境整備[96][97]などに関するバラバラに行われてきた活動を、整理し、方針を定めてきたのが BOAI からなる OA に向けた一連の活動である。

1.2.2 機関リポジトリとは

機関リポジトリの定義として代表的なものには、Lynch(2003)[98]と Crow(2002)[99][100]による定義が挙げられる。これらの定義では、範囲を大学や大学図書館、学術機関としており、コンテンツを構成員による成果物としていた。BOAI で示された OA の定義[1]では、その範囲はインターネット上で講習による利用が可能な形で提供される学術的資料である。これらを踏まえると、「大学と大学の構成員らが創造した資料や知的資産を保管し、インターネットを介して誰もが利用可能な形で提供するウェブサービスと、それに関連して行われる電子化やメタデータの付与などの一連のサービス」が機関リポジトリの役割であると言える。

日本における機関リポジトリの特徴として、国家レベルでネットワーク化された機関リポジトリが運用されている点にある。従来、日本においても機関リポジトリは、各大学が個別にシステムを構築しサービスを提供していた。だが、近年は、NII の提供するサービスである JAIRO Cloud を利用する機関が増加し、ネットワーク化が顕著に進んでいる[29]。 JAIRO Cloud は、Stanford Prize for Innovation in Research Libraries(SPIRL)[101]の功労賞を受賞したこともあるサービスである。その受賞利用としても挙げられている特徴として、Niggemann(2014)は

JAIRO is a political success story that could well be copied if known in other parts of the world. Our recognition may help and support the idea of open access repositories for research publications. The technology seems to be state-of-the-art but the political driving force is powerful and admirable.

と評した[102]。Okerson(2014)は、

The innovation is applying a recently emerged model of software-as-service and cloud to the problem of IRs and doing what looks like a very good job of making it a lot easier and cheaper for everybody to have one. It's a solution to a real problem (how do we do IRs) and it looks like it's working very well.

と評した[102]。評価された点をまとめると、当時としては最先端なクラウドサービスが使われている点、政治・政策と結びついていることで影響力を持っている点、サービスが安価に提供できている点であった。OA は、政策と結びつけて考えられることが多い。だが、単位としては、組織や連合体を単位とした事例が諸外国に多い。そういった中で、国家の単位で同一のシステムを構築し、推進、整備を進めていった点が特徴的であった。

評価された点として挙げられていた技術、コスト、法制度(政策)という観点は、BOAI(2002)においてオープンアクセスの障壁として挙げられていた"財政、法的また技術的障壁"[1]の3点にあたる。

JAIRO Cloud に代表される機関リポジトリは、日本におけるオープンアクセスやオープンデータに関する政策の基盤としての位置付けがなされている[103][104][105]。Lynch の挙げたインフラストラクチャとしての側面と学術コミュニケーションにおける変革が、Corw の挙げたサービスを伴って実現に向けて進んでいると言えよう[106]。

1.3 研究目的と意義

1.3.1 研究目的

日本における機関リポジトリは、各大学が構築し、管理運用していた。だが、JAIRO Cloud の導入によって、機関リポジトリを構築する大学数も増加し、導入後は各大学として管理する負担が減った部分も存在する。そして、同一の基準による資料の利用実態の提供も同様のサービスによって行われる様になった。本研究では、サーバに蓄積された長期的なログを分析し、インフラストラクチャとしての機関リポジトリの役割を確認するためにも、機関リポジトリに登録されたコンテンツの利用実態を明らかにする。

1.3.2 研究の意義

日本における機関リポジトリの研究では、大学横断の研究や、単一の大学における1~3年程度の期間におけるアクセスログの分析が行われてきた。だが、機関リポジトリがインフラストラクチャとして長期的に運用されてきた利用実態を明らかにするためには、長期的な期間における分析が必要である。アクセスログは、機関リポジトリにアクセスした利用者の行動がすべて記録されているため、利用実態を明らかにするためには適当である。本研究における事例で得られた知見を共有することは、日本における機関リポジトリの長期的な分析に有用であり、一般化可能な方法論の適用事例としても価値があると考えられる。

1.4 先行研究

本研究では、以下の手順にしたがって文献の収集を行なった。まず、データベースである Library & Information Science Abstracts を用いて、査読付き論文の検索を行なった。検索に使用したクエリは、("access logs" or "access log") and ("institutional repository" or "institutional repositories")とした。検索を行なった結果得られた11件の論文のうち、11件の論文に重複が存在したので、本研究における実際の対象は9件を収集した。

次に、日本国内の研究を調査するため、CiNiiにおいて、"機関リポジトリ"AND "アクセスログ"という検索クエリを用いての検索を行なった結果として、5件の論文を収集した。そのうちの3件の実践事例を対象とする。また、図書館等の行なった実践事例や調査を収集することを目的とし追加で、「機関リポジトリ」「Institutional repository(ies)」「アクセ

スログ」「access log(s)」に関連するものを調査した。

1.4.1 機関ごとの調査と標準化、ネットワーク化に関する論文

Mark ら(2006)によると Canadian Association of Research Libraries[107](以下：CARL)における OAI-PMH[108]によって収集されたメタデータに関する調査が行われた。CARL は、29 の大学図書館と 2 つの政府機関によって運営されている。(2018 時点)彼らは、共通のインターフェイスを google に変わるツールとして活用することを目標として、メタデータの重要性を示し、その内訳を統計的に比較した。結果として、"data"、"type"、"title"、"description"、"identifier"が含まれており、"coverage"、"relation"、"creator"などが含まれていなかったことが示された[109]。

Susanne ら(2006)は、ドイツにおける機関リポジトリの認証と認証機関に関する調査を行なった。対象は DINI[110]である。ドイツにおけるリポジトリの相互運用や質保証の観点から評価することで OA への貢献の可能性を示したものである。大学間の相互運用可能なサービスの構造化と標準化を定めた。監査の基準は、大きく 6 項目あり、"visibility and server policy"、"author support"、"legal issues"、"authenticity and integrity"、"indexing (subject indexing, metadata, interfaces)"、"visibility/impact/access statistics"、" long-term availability"である。統計的な分析の際には、COUTNER[103]に従ったサーバログの利用がなされている。機関リポジトリのサービスの可視化を行うと共に、インパクトファクターなどの学術評価との連携を提言している[112]。

1.4.2 個別のリポジトリの知見の共有を目的とした論文

Lam ら(2007)は、Hong Kong University of Science and Technology(以下：HKUST)[113]における機関リポジトリ構築のために得た知見を共有している。リポジトリに存在する文章が 42 のコミュニティと 139 のコレクショングループ化が行われていることを示している。そして、リポジトリの利用には、コンピュータサイエンスやエンジニアリング部門などの分野の貢献が大きいことが示された。HKUST では、アクセス解析の結果として毎月 top20 をリストとして提供しており、著者へのインセンティブとして提供している。アジアにおける機関リポジトリの事例として、ネットワーク化の必要性和政府による支援を推進している[114]。

Francis ら(2008)は、National Coronial Information System(以下：NCIS)[115]における機関リポジトリ構築において得た知見を共有している。使用されているシステムは、ePrint@IISc[116]である。相互運用可能なリポジトリのシステムとしての優れた点を示し、インドにおいて開発されたシステムを例にすることで、発展途上国における研究への貢献を示している。この論文の目的は、OA への貢献である[117]。

Fei(2010)は、大学図書館における電子資料の評価に関する支援を目的としてレビューと評価システムの選択基準を示した。大学図書館の予算が縮小される中、電子ジャーナルを

はじめとした電子資源に関するコストが増大していることを指摘した。従来の指標として、引用の分析と利用の分析という観点があることを示した。利用統計が様々な電子資源の分析に利用できることを示した。利用統計の標準化の一つである COUNTER[111]は大学図書館員に多大な手間と時間がかかるので、SUSHI[116]というプロトコルを紹介した。また、"ScholarlyStats"[118]、"Journal Usage Reports"¹¹、"360 Counter"[119]、"Scholarly iQ"[120]4つのシステムを紹介している。電子資源に関する統計の収集は、意思決定に役立つ情報であることを示すと共に、高価な価格や図書館員の手間がかかり複雑であることを述べた[121]。

1.4.3 アクセスログを用いた評価に関する論文

Gemma ら(2006)は、The National electronic Library of Infection(以下：NeLI)[122]の電子図書館にアクセスログ分析、フリーテキスト検索クエリ分析、オンラインユーザ調査の3手法を用いて、定量的評価と定性的評価を組み合わせた評価を実施した。NeLIの特徴は、文献の品質評価が可能な機能があることである。アクセスログ分析の期間は、2002年1月から2005年7月の期間である。ここでは、3つの手法の組み合わせが有効であること、ユーザは新しい文章へのアクセスが多いことが示された。だが、ユーザ調査のサンプルサイズの小ささや、アクセスログのスクリーニング処理の問題点などが残った[123]。

Bonilla-Caleo(2008)は、the institutional repository of the University of Strathclyde(以下：Strathprints)[124]に収録されている物理学分野の5年分論文(2000-2005)を2007年までのアクセスログを利用して分析した。最も引用された論文は、最も古いコンテンツであった。各年度における利用統計や、CAI¹²、IC¹³との比較、月辺りの利用元国家、他の機関での公開を調査していた。利用としては、学術論文が一番多く、ArXive[6]やWeb of Science [125]に掲載されていない論文の利用が多くなっているが、引用が多い論文としての相関がないことが示された[127]。

Christine ら(2008)は、JISC Usage Statistics Project[128]のためにアイテム単位における利用評価のための集計基準を示した。そのためのレビューとして、IRStats[129][130]プロジェクトを通したAWStats[131]の利用、LogEc[132]内でのrobots.txt[133]の利用、DINI[110]やBEST[134]、MESUR[135][136]、COUNTER[111]、PRIUS[137]、IFABC[138]を紹介していた。これらの組織が利害関係者として参加して示されたのは、人間によるアクセスとロボットによるアクセスの区別し、クッキーの利用、技術的には

¹¹ 本研究においては、所在を得ることができなかった。Thomson Reuters(現在、Clarivate Analytics)の提供していた課金制サービス。

¹² co-authorship index：文書あたりの平均著者数として定義された共著者指数

¹³ collaboration index：センター間コラボレーション指数、

Fractional Counting(FC)：上記のカウントについては分数カウントが用いられている[126]。

DOI[139]の付与、OAI-PMH[108]の利用、OpenURL[140]の利用が推奨された。そして、国際標準の導入や COUNTER の方針が示された[141]。

Brown(2014)は、大学図書館員による研究者の支援のうち、研究成果の証拠や教員の研究分野への影響を提示するための手段を示した。計量書誌学的方法としての引用検索と近年開発された指標を提示した。インパクトファクターの利用から誤用までの注意を示した。Webmetrics 分野の発達と共に、リンク分析や web 引用分析などの有用な分析手法を示し、ジャーナルインパクトファクター¹⁴から Web インパクトファクター¹⁵を紹介した。また、Altmetrics[144][145]や Webometrics を紹介し、伝統的な計量書誌学として受け入れられるには、時間がかかるが強力な道具になり得ることを示した[146]。

1.4.4 関連研究

佐藤ら(2011)は、2008 年から 2010 年にかけて、機関リポジトリ登録論文の被引用数が電子ジャーナルのアクセス数に与える影響を測定するための実験を行なった。概要は、北海道大学学術成果コレクション[147]と京都大学学術情報リポジトリ[148]に登録し、Science Citation Index[149]との比較を行うものであった。COUNTER[111]に基づく処理を行なったアクセスログを用いた利用者の属性を解析した。リファラの解析によって、機関リポジトリの利用が、電子ジャーナルに匹敵することを明らかにしたが、その利用が民間プロバイダからのアクセスが多数であり、専門領域の研究者以外によるアクセスであることから、被引用数の増加には繋がらなかったことを示した。今後、同様の実験を行う場合は、サンプル数や掲載する論文による影響を考える必要性を示した[150]。

佐藤(2013)は、北海道大学学術成果コレクション、京都大学学術情報リポジトリ、つくばリポジトリ[151]、一橋大学機関リポジトリ[152]、アジア経済研究所学術研究リポジトリ[153]における 2008 年の 1 年分のアクセスログ分析を行なった。機関リポジトリの統計データ、アクセス先、アクセス元のコンテンツ分析を実施した[154]。

馬場ら(2011)は、2008 年から 2009 年にかけての九州大学学術情報リポジトリ[155]のアクセスログを用いてのアクセスログ分析を実施した。馬場らは、アクセスログから機械的なアクセスを除去するために"bot"の文字列を含むログの削除の実施をし、同日の同一アドレスからのアクセスを同一利用者と定義した。そのデータを用いて共起解析を実施した[156]。

木下ら(2011)は、2008 年から 2010 年にかけての九州大学学術情報リポジトリのアクセスログの中にある検索語を用いた分析を実施した。検索語と論文のタイトルの合致度を計

¹⁴ Journal Impact Factor : Clarivate Analytics によると "the impact factor of a journal is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years" である[142]。

¹⁵ Web Impact Factor :

Thelwall によると "show how many times a page has been visited" and non-link invocations, "where research has been mentioned online without a link" である[143]。

算し、hits 率を求めた。結果として、月や年ごとによる変化はあまり見られなかった。検索語の情報を元に 3 つの推薦手法を用いた結果、論文タイトルのキーワード推薦とアクセスログの検索語を用いた推薦では効果的であり、共著者推薦では、良い結果が得られなかった。その原因として、直接意味を持たない語の除去などが適切に行われていなかったことを示した[157]。

吉川(2016)は、2014 年から 2015 年にかけての JaLC DOI[158]のアクセスログ分析を実施した。アクセス数の多いコンテンツ、アクセスの多い参照元、利用者の属性を明らかにした。結果として、アクセスの多い上位 15 のコンテンツに関しては、全て本文の公開がされていることが示された。また、分野に関しては、医学・医療分野へのアクセスが多い。また、アクセス元に関しては、サイトによって、大学等の機関や民間プロバイダからのアクセスなどに違いが出たことを結果として示した[159]。

高久ら(2018)は、図書館資料を対象とした検索結果のランキング手法の改善に、2014 年から 2018 年における筑波大学附属図書館 OPAC[160]の利用ログを用いた。書誌情報の各項目のテキスト類似度による BM25F リランキング手法、書誌情報の閲覧回数の重み付けによるリランキング手法、クエリログを用いた検索手法の 3 つの重み付け手法を提案、評価すると共に、検索ニーズの検討や、検索サービスやアクセスログの今後の利用に関する必要性を述べた[161]。

2. 調査手法

本研究において用いた方法は、アクセスログ分析である。本章では、分析の対象と、分析方法に関して述べる。本研究の対象となる機関リポジトリは筑波大学の機関リポジトリ「つくばリポジトリ」のみであるが、利用しているシステムが途中で変化している。そのためアクセスログの出力データの形式も異なる。これを踏まえて、対象を2つに大別する。1. 2010年3月から2014年3月までのアクセスログデータ(DSpace[162]のApacheログデータ)、2. 2014年4月から2017年3月までのアクセスログデータ(JAIRO Cloud[163]の出力データ)以下では、それぞれのデータの形式について述べ、そのデータからわかる情報をまとめる。

2.1 分析対象の概要

本研究における分析対象は、筑波大学の機関リポジトリに登録されたコンテンツの、2010年3月から2017年12月までのアクセスログである。詳細として、2010年3月から2014年3月までのアクセスログデータは、筑波大学がDSpaceを利用していた期間にあたるため、サーバに記録された生のアクセスログデータを用いる。2014年4月から2017年12月までのアクセスログデータは、筑波大学がJAIRO Cloudを導入してからのアクセスログデータのため、JAIRO Cloudの出力形式でのデータを用いる。

2.1.1 筑波大学の機関リポジトリ

本研究における対象は、筑波大学の機関リポジトリ「つくばリポジトリ」[151]に登録されたコンテンツへの2010年3月から2017年12月までのアクセスログである。筑波大学の機関リポジトリに関するデータとリポジトリとしての特徴、本研究における調査対象としての選出理由について以下に記述する。

表 4 つくばリポジトリの運用形態の変化¹⁶とシステム管理者

システム	正式稼働開始	URL	システム管理
DSpace	2007年3月	http://www.tulips.tsukuba.ac.jp/dspace/	筑波大学附属図書館
JAIRO Cloud	2014年5月	https://tsukuba.repo.nii.ac.jp	筑波大学附属図書館 国立情報学研究所

¹⁶ DSpace と JAIRO Cloud の間で大きく違う点は、システムの管理とメンテナンスを各大学が個別にする必要がなくなった点にある[164]。DSpace を利用していた大学では、ソフトウェアのアップデートが進んでも、大学のシステムのリプレースの関係でアップデートが遅れ、セキュリティに問題があったという指摘もある[87]。

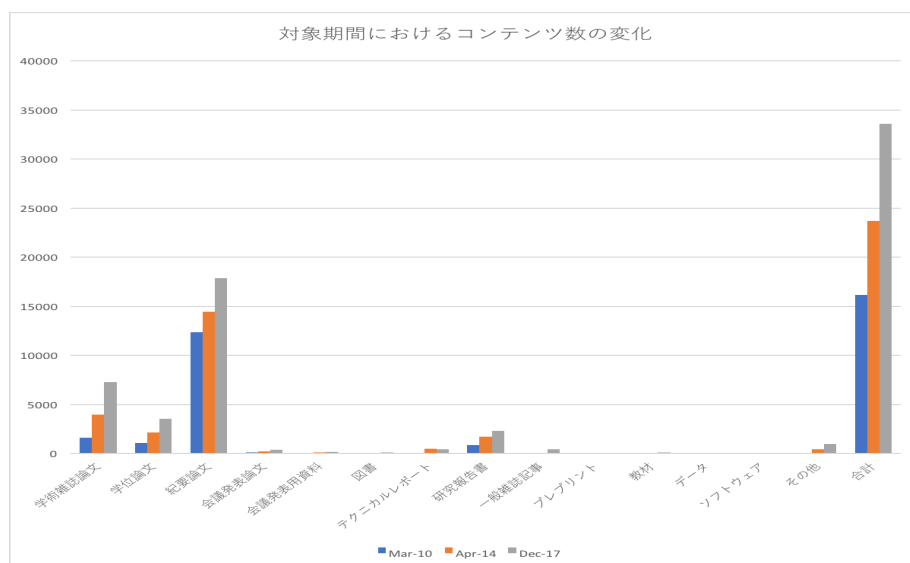


図 11 つくばリポジトリのコンテンツ数の変化(IRDB より作成)[165]

表 5 内訳(コンテンツ登録件数 Top3)

	学術雑誌論文	学位論文	紀要
Mar-2010	1,598	1,079	12,340
Apr-2014	3,983	2,152	14,459
Dec-2017	7,300	3,536	17,867

表 6 内訳(コンテンツ登録件数 Top3 以外)

	会議発表論文	会議発表用資料	図書	テクニカルレポート	研究報告書	一般雑誌記事	プレプリント	教材	データ	ソフトウェア	その他
Mar-2010	129	11	34	27	861	8	48	7	1	0	15
Apr-2014	210	109	78	460	1,690	54	50	28	3	0	404
Dec-2017	365	137	103	454	2,306	422	53	96	0	0	955

筑波大学附属図書館の特徴は、佐藤(2009)の研究においては、以下の様に紹介されていた[166]。

筑波大学(Tulips-R)：1990年代から電子図書館プロジェクトに取り組んできた大学の一つであり学術論文の登録がとりわけ多い。その特徴の分析から、電子図書館の流れを組むものとしての機関リポジトリについて検討できると考えられる。

その他の特徴として、稲永(2012)[167]の報告によると、NII の提供する IRDB (Institutional Repositories Data Base)¹⁷[168]のデータから日本の機関リポジトリの現状を分析した[167]。その中で本文有コンテンツ数のランキングで上位にあるのも特徴であった。

表 7 2012.10 時点で本文有コンテンツ数のランキング ¹⁸[167]

順位	大学名(大学種別)	本文有コンテンツ数
1	京都大学(国立大学)	112,645
2	千葉大学(国立大学)	88,416
3	東北大学(国立大学)	40,037
4	北海道大学(国立大学)	37,978
5	お茶の水女子大学(国立大学)	35,554
6	岡山大学(国立大学)	30,648
7	金沢大学(国立大学)	28,085
8	東京大学(国立大学)	27,420
9	筑波大学(国立大学)	27,229
10	広島大学(国立大学)	24,896

筑波大学のコンテンツの掲載数とコンテンツ中の学術論文の割合に特徴があることが先行研究によって示された。そして、現在のデータを見ても同様のことが言える。また、筑波大学の大きな特徴として、DSpace による独自の機関リポジトリを運営した大学で、最初に JAIRO Cloud にシステムを移行した大学である点も挙げられる[170][171]。

筑波大学は、機関リポジトリを DSpace から長期間運用していた実績があり、コンテンツも充実している。コンテンツ内訳に関しても、日本の機関リポジトリの中では、学術論文の割合が多く OA のポリシーに適合したサービスとなっている。そして、JAIRO Cloud の移行実験から参加している大学である。つまり、DSpace と JAIRO Cloud のどちらも導入期から長期的な運用を続けてきた大学であるといえる。上記の様な特徴から、日本における機関リポジトリに登録されたコンテンツと機関リポジトリに関する長期的な分析に適した大学であると判断した。

¹⁷ 2018 年 12 月 20 日からシステムリプレイスのために停止中
移行先は、<https://dev.irdb.nii.ac.jp/>[161](2018 年 1 月時点)

¹⁸ 大学種別に関しては、著者が付与

2.1.2 長期的なアクセスログの保管について

本研究の対象である筑波大学は、DSpace のアクセスログデータを 2010 年の 3 月分から 2014 年 3 月分まで保管していた。これは、2007 年の国立情報学研究所における「機関リポジトリアウトプット評価ワークショップ」[172]の事例を参考にすると非常に稀有な例であると言える。ワークショップに関する報告が掲載されている報告書では、14 の大学が参加したうちのデータ提供を得られた大学は 11 大学、そのうち全期間のデータを保持していたのは 2 大学であった。国立情報学研究所による学術機関リポジトリ構築連携支援事業では、2005 年に 19 の大学への支援を行っていた。2002 年に千葉大学がサービスを開始したことを考えても、このワークショップが開催された際に、長期間¹⁹のアクセスログを保持していた大学があったとは考えにくい。当時から図書館においてログデータが軽視されていることが、懸念されていた。

本研究における対象である筑波大学は、サービスを独自構築していた DSpace から国立情報学研究所の運営する JAIRO Cloud に以降して 4 年以上の期間が経過した 2018 年においても、2010 年からのアクセスログデータを保持していた。そういった意味で長期的なアクセスログデータを保持していた貴重なサンプルである。

2.1.3 アクセスログデータの概要

本研究で得た 2010 年 3 月から 2014 年 3 月までのアクセスログの形式は、DSpace で使用している Apache のログである。http 通信のログは、Apache の共通ログ形式(IP アドレス、識別情報、認証ユーザ、タイムスタンプ、リクエスト、ステータス、バイト数)にリファラとユーザエージェントが加わった拡張形式である。https 通信のログは、リファラやユーザエージェントを記録するための拡張を行っていない標準形式である。

¹⁹ この場合における長期間とは、総務省のデジタルデータの長期保存について、通常利用可能と定義されている 3 年を越えるものとする[173]。

```

119.63.195.33 - - [02/Jan/2011:00:05:16 +0900] GET
/dspace/browse?type=author&sort_by=1&order=ASC&rpp=20&etal=-
1&value=Kagaya%2C+Atsuko&offset=0 HTTP/1.1 200 3286 - DoCoMo/2.0
P05A(c100;TB;W24H15) (compatible;
BaiduMobaidier/1.0;+http://www.baidu.jp/spider/) 8063
125.54.243.215 - - [02/Jan/2011:00:05:16 +0900] GET /dspace/handle/2241/249
HTTP/1.1 200 2399
http://www.google.co.jp/search?hl=ja&q=%E7%AD%91%E6%B3%A2%E5%A4%A7%E
5%AD%A6%E3%80%80%E6%AD%B4%E5%8F%B2%E5%9C%B0%E7%90%86&aq=f
&aqi=&aql=&oq=&gs_rfai= Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.0;
SLCC1; .NET CLR 2.0.50727; Media Center PC 5.0; .NET CLR 3.5.30729; .NET CLR
3.0.30618; OfficeLiveConnector.1.3; OfficeLivePatch.0.0; .NET4.0C) 6430

```

図 12 2011/01/02 http 通信のログより抽出(着色は著者による)

```

119.235.237.19 - - [02/Jan/2011:00:00:48 +0900] GET
/dspace/browse?type=author&value=Heo,%20Myeongja&value_lang=ja HTTP/1.1
200 31264
67.195.111.160 - - [02/Jan/2011:00:00:49 +0900] GET /dspace/handle/2241/5644
HTTP/1.1 200 2691

```

図 13 2011/01/02 https 通信のログより抽出(着色は著者による)

図.10 と図.11 は本研究の対象となる機関リポジトリのアクセスログをそれぞれの形式別に抽出して示したものである。以下に、上記のアクセスログに著者が行った要素別の色分けの詳細を表形式にまとめたものを記載する。

表 8 DSpace のアクセスログの要素と色分け

データ	色	共通	備考
IP アドレス	黄緑	共通	
識別情報	無着色	共通	
認証ユーザ	無着色	共通	
タイムスタンプ	水色	共通	
リクエスト	黄色+オレンジ+無着色	共通	詳細にて説明
ステータス	緑	共通	
バイト数	無着色	共通	
リファラ	赤	http のみ	
ユーザエージェント	紫	http のみ	

上記の表における各要素のうち、実際に色分けした要素に関する説明を行う。アクセスログデータにおける 1 行は、利用者が対象の Web サイト上で行った行動の 1 つを表現している。上記の図では、それぞれ 2 行ずつが示されているため、Web サイト上に記録された 2 つの行動の記録であることがわかる。

以下では、1 行、つまり、利用者が行った行動の 1 つの内訳を確認する。1 行を要素ごとに分割し、必要な要素を確認することで、利用者の行動別にアクセスログを読み取ることが可能である。

まずは、利用者を特定する要素として、IP アドレスが存在する。厳密に個人を IP アドレスから特定するという意味合いではない。IP アドレスは固定と自動割り当てが存在する。例えば、筑波大学であれば"「130.158.0.0/16」「133.51.0.0/16」"である[174]。この IP アドレスから逆引きし、Host をリクエストすることで、ある程度のアクセス元まで特定することができる。具体的には、利用者の所属の国家や大学などがアクセス元として挙げることでできる要素だ。自動割り当てに関しては、必ず同じアクセス元であるとは限らない。

タイムスタンプは、利用者が行動を起こした日時を詳細に示している。時間は、連続アクセスなどを判別するのに利用可能である。web サイトにおけるロボットによるアクセスの回数が大半を占めている現在において、bot と人間によるアクセスを区別するのに重要な要素である[175][176]。

リクエストは、複数の色で着色を行った。黄色の部分は利用者の行動を表す。また、オレンジ色の部分はアクセスをしたファイルの位置を示す。利用者の行動が 1 行のログとして出力されているアクセスログにおいて、利用者がアクセスを試みたのが、サーバのどのディレクトリのどのファイルに対しての行動であるかを示している。つまり、利用者が対象のサイトで行ったことを表す重要な要素である。本研究では、特に 2 つの文字列に着目してデータを扱う。1. handle 2. bitstream である。これらの用語の詳細は次節にて説明する。

リファラとユーザエージェントは、Apache の標準形式だと出力されない。そのため

か、本研究で扱う **https** 通信のデータに関してはこれらの要素が記録されていない。また、**https** では、これらのデータを取得することは、セキュリティ上望ましいとはあまり言えないので、データの取得をすることも可能ではあるが、データを取得していないのは、一般的である。リファラに関しては、W3C の **Referrer Policy** に詳細が記述されている[177]。

リファラとは、参照元の **web** ページの **URL** を示すものである。そのため、利用者が他の **web** サイトからの遷移で該当する **web** ページにアクセスしたのか、また、その元はどこであるかを示す。

ユーザエージェントは、使用しているソフトウェアやブラウザの環境を示す項目である。また、機械によるアクセスはユーザエージェントにその旨を記述していることが多い。例としては、**bot**, **spider**, **crawler**, **agent**, **robot** などの文字列を含むものが挙げられる。

これらの機械によるデータの収集に関する決まりは、それぞれの **web** サイトによって定義されている。日本の図書館では、過去にクローラによる機械的なアクセスが原因でユーザが逮捕されている事例[178][179]もあるため、ユーザエージェントはルールに従って運用する必要がある。国立国会図書館の資料では、“**http://www.example.ac.jp/** における **robots.txt** は「**http://www.example.ac.jp/robots.txt**」で確認できる。”とある。[180]筑波大学附属図書館における **robots.txt** は「**http://www.tulips.tsukuba.ac.jp/robots.txt**」にアクセスすることで取得可能である[181]。

```
User-agent:*

Disallow: /myassist/
Disallow: /admin/
Disallow: /help/
Disallow: /cgi-bin/limedia
Disallow: /manuals/
Disallow: /toukei/
Disallow: /icons/
Disallow: /gifs/
Disallow: /internal.html
Disallow: /jouhou_beya/
Disallow: /pub/new_books/
Disallow: /pub/form/
Disallow: /dspace/browse-title
Disallow: /dspace/browse-author
Disallow: /dspace/items-by-author
Disallow: /mylimedio/
Disallow: /metalib
Disallow: /V
Disallow: /M
Disallow: /sfx
Disallow: /metalib/
Disallow: /V/
Disallow: /M/
Disallow: /sfx/

Sitemap: http://www.tulips.tsukuba.ac.jp/dspace/sitemap
```

図 14 筑波大学附属図書館の **robots.txt**(accessed 2018-11-26)[181]

アクセスログデータの概要に関しては、以上である。上記において筆者がアクセスログ着色したアクセスログにおける 6 要素を抽出することで、ユーザのアクションから必要なデータを抽出し、実際の利用行動の分析に必要な情報を得ることが可能であることを示した。

アクセスログデータは、利用者が対象の Web サイト(本研究の対象では機関リポジトリ)にアクセスし行った 1 つ 1 つの行動がそれぞれ 1 行のログとして保存されている。つまり、利用者がサイトやファイルに対して行った行動の全てがサーバへの通信という形式で保存されている。つまり、アクセスログとは利用者が、機関リポジトリ内で行った行動のすべてが記録されているデータである。

本研究における表記が先行研究において異なる点として、**http** による通信と **https** による通信が別のログとして保存されていた点に対して、先行研究においては、特筆して示されていないなかったものの、本研究で扱う期間に関しては、**http** による通信と **https** による通信が別々のログファイルとして存在していたことを表記している点にある。

本研究では、どちらの通信携帯によるデータも対象とする。一部のデータのメタデータについて、**https** による通信のログデータを区別するために **ssl** と表記している。なお、色分けに使用した規則は、同一のため、同色の部分に関しては、**http** と **https** に共通の要素である。なお、データの色分けに関しては、実際に研究に利用したデータを示すために恣意的に着色と無着色を区別した。

JAIRO Cloud のログデータは、WEKO[182]のの機能で出力されている。本研究に用いるデータは、12 種類提供されている定型レポートのうちの 7 種類の月毎のレポート、4 種類の利用統計レポートのデータを用いた。WEKO から出力可能なデータのメタデータに関しては、以下の表の通りである。以下に示すのは、代表的なメタデータである。

表 9 Jairo Cloud のアクセスログデータのメタデータ(利用統計)[184]

DetailView	タイトル	登録インデックス名	閲覧回数	未ログインユーザ
File Download	ファイル名登録 インデックス名	ファイル ダウンロード 回数	未ログイン ユーザ	ログイン ユーザ
File View	ファイル名登録 インデックス名	ファイル 再生回数	未ログイン ユーザ	ログイン ユーザ
Host Access	ホスト	IP アドレス	WEKO トップ ページアクセス数	
IndexAccess	インデックス	閲覧回数		
SearchCount	検索キーワード	検索回数		
Site Access	WEKO トップページアクセス数	検索回数	閲覧回数	ファイル ダウンロード回数

表 10 JAIRO Cloud のアクセスログデータのメタデータ(サイトライセンス)[184]

名称/メタデータ			
ダウンロード回数 (アイテム) ²⁰ (月ごと)	アイテム ID	アイテム名	ダウンロード回数
ダウンロード回数(日ごと)	期間	ダウンロード回数	
閲覧回数(アイテム)(月ごと)	アイテム ID	アイテム名	閲覧回数
閲覧回数(日ごと)	期間	閲覧回数	

2.1.4 用語の解説

上記では、データの詳細ということで、ユーザのアクションに関わる部分を説明した。それらに関連していくつかのシステムの説明を補足として記述する。対象は主に 3 つである。

1.handle

handle は、Corporation for National Research Initiatives(CNRI)[185]によって提供されている永続的識別子の名称である。永続的識別子には、handle の他に Digital Object Identifier (DOI)[186]や persistent identifier (PID)[187]といったものが知られている。永続的識別子は、電子情報資源への永続的なアクセスを保証する Uniform Resource Identifier(URI)を提供している。これは、資料が保存されている場所(Location)²¹に依存しないため、研究者の所属大学が切り替わり、著作物を利用可能な形で保管するリポジトリが切り替わっても、同一の URI によって、インターネット上でのアクセスを保証するシステムである。

2.bitstream

bitstream は、DSpace 1.2 から永続的識別としての機能を持つようになった。Handle システムに依存しないように、ファイルの保存先が変更してもアクセスが可能になっている。DSpace の説明のシステム説明書を参考によると、“DSpace は、bitstream をそのままの形でアップロード・ダウンロードする機能をサポートしているだけです。”[188]とあるので、DSpace でのダウンロードやアップロードは bitstream を追うことで行動を追跡できる。

3.OAI-PMH

OAI-PMH とは、Open Archives Initiative Protocol for Metadata Harvesting[100]の略であるが、一般的に OAI-PMH という省略された形で使用される。図書館情報学用語辞典における定義は以下に示す通りである。

²⁰ 出力単位に関しては、ユーザがある程度変更することができる。

²¹ Uniform Resource Locator(URL)の場合サイトの構造や、保管するフォルダなどの場所に影響を受けてアクセスができない場合があるが、永続的識別子を用いて、ユニークな Uniform Resource Name(URN)を持つ URI なら、場所に依存せずにアクセスを保証することができる[189][190]。

2001 年に Open Archive Initiative (OAI) が策定した、インターネット上のメタデータ自動収集のためのプロトコル。リポジトリを維持しメタデータを開示するデータプロバイダと、開示されたメタデータをハーベスタと呼ぶアプリケーションにより収集して加工・提供するサービスプロバイダから構成される。収集に使用されるリクエストは 6 種類のみ。多くの機関リポジトリで採用され、ミシガン大学が運営しているサービスプロバイダ OAIster は全世界の数百のリポジトリを対象にハーベスティングを行っている。

日本における事例で考えると、JAIRO Cloud を運営している NII がサービスプロバイダであり、各大学がデータプロバイダに当たる。データは、誰にでも開かれた形で提供されている[191][192]。利用者が API を使用することで、サービスプロバイダから、データプロバイダが提供しているメタデータを xml 形式で取得することができる。形式としては、"OAI-PMH, Dublin Core, Junii2, LOM, LIDO, SPASE"などがある。(WEKO で利用可能な形式)[184]本研究で使用する筑波大学附属図書館のメタデータは、Junii2 形式で取得した。

表 11 WEKO で取得可能な Junii2 形式のメタデータ[192]

No .	タグ名		属性	内容
	第 1 階層	第 2 階層		
1	junii2		xsi:schemaLocation	文書のルート要素
2		title		アイテムタイトル
3			lang	アイテムタイトルの言語
4		alternative		別言語のアイテムタイトル
5			lang	別言語のアイテムタイトルの言語
6		creator		著者
7			lang	著者の言語
8			id	著者の研究者リゾルバーID
9		subject		キーワード
10		NIIsubject		国立情報学研究所メタデータ主題語彙集
11		NDC		日本十進分類法
12		NDLC		国立国会図書館分類表
13		BSH		日本件名標目表

14		NDSLH		国立国会図書館件名標目表
15		MeSH		医学件名標目表
16		DDC		デューイ十進分類表
17		LCC		米国議会図書館分類表
18		UDC		国際十進分類表
19		LCSH		米国議会図書館件名標目表
20		description		抄録
21		publisher		出版者
22			lang	出版者の言語
23			id	出版者の研究者リゾルバーID
24		contributor		寄与者
25			lang	寄与者の言語
26			id	寄与者の研究者リゾルバーID
27		date		日付
28		type		資源タイプ
29		NIItype		国立情報学研究所メタデータ主題語彙集（資源タイプ）
30		format		フォーマット
31		identifier		その他の資源識別子
32		URI		資源識別子 URI（アイテム表示画面へのリンク）
33		fullTextURL		本文フルテキストへのリンク
34		selfDOI		selfDOI
35		isbn		ISBN
36		issn		ISSN
37		NCID		書誌レコード ID（総合目録 DB）
38		jtitle		雑誌名
39			lang	雑誌名の言語
40		volume		巻

41		issue		号
42		spage		開始ページ
43		epage		終了ページ
44		dateofissued		発行年月日
45		source		情報源
46		language		言語
47		relation		他の情報源との関係
48		pmid		PubMed 番号
49		doi		DOI
50		NAID		NII 論文 ID
51		ichushi		医中誌 ID
52		isVersionOf		異版である
53		hasVersionOf		異版あり
54		isReplacedBy		置換される
55		replaces		置換する
56		isRequiredBy		要件とされる
57		requires		要件とする
58		isPartOf		部分である
59		hasPart		部分を持つ
60		isReferencedBy		参照される
61		references		参照する
62		isFormatOf		別フォーマットである
63		hasFormat		別フォーマットあり
64		coverage		範囲
65		spatial		空間的範囲
66		NIIspatial		国立情報学研究所メタデータ主題語彙集（地域）
67		temporal		時間的範囲
68		NIItemporal		国立情報学研究所メタデータ主題語彙集（時

				代)
69		rights		権利
70		textversion		著者版フラグ
71		grantid		学位授与番号
72		dateofgranted		学位授与年月日
73		degreename		学位名
74		grantor		学位授与機関

2.1.5 分析対象期間と単位

先行研究によると、機関リポジトリのアクセスログを対象とした研究においては、対象とする期間が1年から2年のものが多い。5年を超える様な長期間のアクセスログを対象とした研究はほとんど見られない。機関リポジトリは、Crow(2002)の定義において、"(1)学術コミュニケーションの変革を促進し大学と図書館を構成要素として位置付け、(2)学術機関の構成員の知的資産を保管し、学術機関の認知度と地位の向上を図るもの。"として記述されている[99]。そこでは、保存という観点から機関リポジトリに関して長期的な視野での観察が必要であることが示されていた。

本研究では、対象を2010年3月から2017年12月と長期間に設定することによって、アクセスログという観点から機関リポジトリの実態を調査した。約8年分のデータを対象とすることで、筑波大学が採用した各システムにおける長期的な分析と、1つの機関における長期的な分析を実現した。

本研究におけるアクセスログの分析期間の区切りは、基本的に年度を採用する。年度の定義は、1月1日から12月31日までを1年度とした。なお、3月から始まっているデータや、5月から始まっているデータに関しては、同様の年度のデータの一番古い日付から12月31日までの期間を1年度として扱っている。

2.2 データの処理

本研究では、期間(DSpace と JARIO Cloud)によって異なる形式をもったデータ2種類のデータに対して、それぞれ別々の処理を行っている。基本的な考え方は、COUNTERに従うため、基準としては同一の基準をもって処理を行ったことになるので、データの比較が可能である。

アクセスログには、webサイトを構築する画像の呼び出しやJava scriptでの処理、コンテンツと直接関係の薄いページへのアクセスなどのログからシステムの更新記録などのユーザのコンテンツの利用とは全く関係ないログも記録されている。本研究では、ユーザのアクションを中心にコンテンツの利用実態を明らかにしたいと考えていたため、不要なログ

を除外し、必要なログのみの抽出を行った。

2.2.1 アクセスログのスクリーニング

コンテンツの利用を表すログを、**bitstream** の文字列を含む行と定義した。だが、上記の文字列が含まれている行を全て抽出した結果は、ユーザの利用行動を表すアクセスログであるとは言えない。その理由は、前項で示した **bot** によるアクセスの問題がある。前述のユーザエージェントの説明において、アクセスログには機械的なアクセスが多く含まれていることを示した。その機械の多くは、**Robots.txt** に従い処理を行うことになっている。逆説的に言えば、それぞれの **web** サイトにおいて **Robots.txt** を記述する必要がある程度には、機械によるアクセスの数が多いことになる。筑波大学附属図書館のアクセスログを利用した高久(2018)の研究では、アクセスログデータを扱う際に、全データから機械などを除いた実データを母数と定義しなおして分析を実施していた[162]。

本研究でも、機関リポジトリへのアクセスログを適切に把握することを目的として、**COUNTER Code for Practice** に準拠したカウント方法を採用した。また、この方法を採用する目的は、2014 年以降のデータである **JAIR O Cloud** のアクセスログを分析する際の齟齬を減らす意味もある。前田ら(2016)は、**JAIR O Cloud** のアクセスログ分析の高度化のために行なっていることが、"CSI 委託事業の **ROAT** プロジェクトを踏まえたアクセス統計の客観的評価である。"と記していた[87]。**ROAT** の報告書にも示されているが、アクセスログは、処理の際の考え方や基準次第で、内容が大きく異なるものである[193]。そのことも踏まえ、長期的なログの分析を行う際にも近い手法を採用する必要があった。

2.2.2 COUNTER とは

COUNTER(<https://www.projectcounter.org>)[110]は、"Counting Online Usage of NeTworked Electronic Resources."を示す非営利団体である。名称からも読み取れることができるが、図書館や出版社などが中心となって、電子的な資料の扱いに関する行動規範などについて、ワーキング・グループを通して提供している団体である。アクセスログの利用統計の標準化のための活動を行っており、本研究におけるアクセスログのスクリーニング処理は、**COUNTER** を参考に行なった。

2.2.3 本研究におけるフィルタリングの構造

本研究では、先行研究においても用いられている方法によって、データのスクリーニングを行った。下図は、今回用いた佐藤(2008)[193]による提案モデルである。なお、データに関して、自ら処理を行った範囲は、2010 年 3 月から 2014 年の 3 月の期間におけるデータに限る。

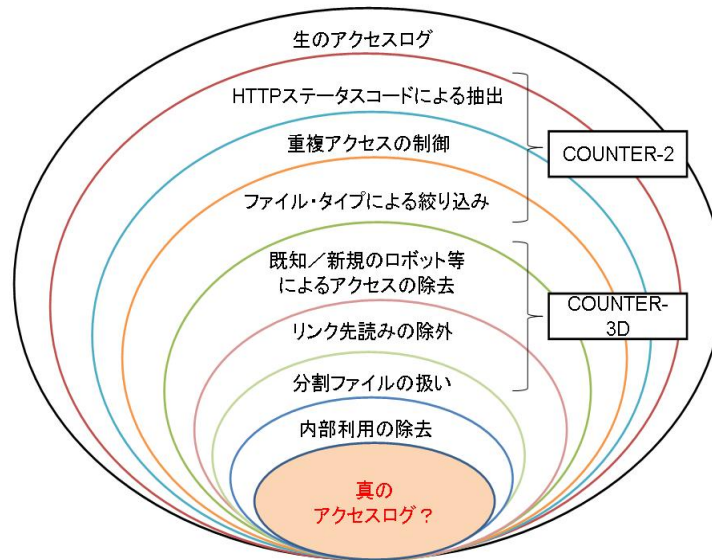


図 15 COUNTER [110](accessed 2018-12-26)

本研究では、佐藤(2013)による、上記のフィルタリングモデルの活用事例を参考にスクリーニングを実施した[154]。ここでは、先行研究と異なる部分のみを記述する。

既知/新規のロボット等によるアクセスの除去では、先行研究では、IP アドレスとユーザエージェントの文字列一致と、同一コンテンツへのアクセス回数により、bot の特定を行っていた。だが、本研究では、NII の提供するロボットリストを参考にしたため、アクセス回数によるロボットの特定は行っていない。

内部利用の除去では、先行研究においては内部利用のリストを利用したとあるが、本研究では得られていない。そのため、逆引きした IP に含まれている Tulips と個人名を持ったものを内部によるアクセスと定義し、一部を除去した。

2.2.4 フィルタリングの種類と結果の比較

本研究においては、アクセスログの解析の手法として、2つの方法の比較を行なった。上記においても基準が異なる場合は、分析対象となるデータが異なることが示されていた。2013年のROAT報告書においては、AWstats[131], Google Analytics[194], WebAlizer[195]等のソフトが具体的に示されていた。中でもAWstatsの利用について、"IRstats"[130]が紹介されていた。

本研究では、AWstats同様にサーバログの形式のデータをローカルな環境で解析可能な、Visitors[196]を用いたアクセスログ解析の結果を用意した。今回の対象となるデータは、10億件近いデータになったため処理が早くローカルで作業可能なものである必要があった。実際にVisitorsによって行なった解析の結果とCOUNTER code for Practiceに沿

った形で解析の結果を表にして付録に示す。

2.2.5 データの種別と重み付けについて

アクセスログデータは、ユーザの行動を示すデータであることは前述している。では、そのデータを測るのに適切な分類はという点に関する議論はあまりなされていない。本研究では、データに関する視点を提示し、分析に用いるデータをどのように位置付けているのかを明確にする。

ユーザの視点から、ファイルのダウンロードを行うという行為をデータで表すと、ファイルをダウンロードする/しないという行為を 1/0 で表現することに等しい。そのため、この場合のデータは名義尺度として扱われている。つまり、この視点におけるデータを統計的に扱うことや演算を直接行うことは意味がない。

だが、視点をファイルのダウンロード回数や利用回数として表現すると、これは比例尺度として判断することができる。ファイルの回数が増えることには意味があり、数が多いとは利用の回数が多いことに等しくなり、順序尺度である。また、利用を%で表現し、期間中に n%上昇したと表記することも可能である。それらの要因を加味して、今回はデータを比例尺度として使用する。

期間を定めることによりデータの利用を増減で評価することできるとした。そのため、期間を設定すると共に、利用に重み付けを行う必要があると考えられる。だが、それはデータの形式的な問題と登録の際の理由等を加味して行わない。

データの形式の観点から判断すると 2010 年 3 月から 2014 年 12 月の期間のデータに関しては、コンテンツが利用された期間と紐づけて計測することが可能である。だが、提供を受けた 2014 年 5 月以降のデータは、個別のコンテンツについて利用された時期と回数が紐づいていないため、それぞれを別に判断する必要がある。そのため、期間を定めて重み付けを行うとデータの期間によって、判断基準が変わる恐れがある。

2.3 データの分析

2.3.1 調査対象 1: Host の抽出によるアクセス元の調査

本研究では、COUNTER に従う形でデータのクリーニングを実施した。その結果で抽出されたアクセスログにおける IP アドレスを対象にして、アクセス元を明らかにするために、Host のリクエストを行った。リクエストして結果が得られた Host をドメイン別に分類した。トップレベルドメインとセカンドレベルドメインを抽出して、アクセス元の国家や組織等の属性を明らかにした。

2.3.2 調査対象 2: コンテンツの利用の定義と調査

本研究におけるコンテンツの利用を定義として、コンテンツが 1 回ダウンロードされたことを 1 回コンテンツが利用されたこととした。コンテンツの本文が含まれるファイルへ

のアクセスログには、"bitstream"の文字列が記録されていた。そのため、アクセスログの全体から"bitstream"の文字列を含む行を抽出した。つまり、COUNTER による抽出後、bitstream の文字列を含む 1 行のログを 1 回の利用として定義し、分析を行った。手法としては、単純集計、オルトメトリクスに関する調査の文字列を含む行を抽出した。つまり、COUNTER による抽出後、bitstream の文字列を含む 1 行のログを 1 回の利用として定義し、分析を行った。手法としては、集計、オルトメトリクスに関する調査、利用減衰に関する調査、平均反応時間の算出を実施した。

オルトメトリクスに関する調査では Ceek.jp において、“筑波大学 AND 学位論文”というクエリを用いて検索を実施した。Ceek.jp を用いた理由としては、オルトメトリクスの特性にある。オルトメトリクスは、SNS 使用割合や言語などの地域性に影響を受けるため、世界的に使われている Altmetrics.com ではなく、日本におけるデータを収集しているものを選択した。結果としては、654 件が該当した。その中からスコアが 2 以上のデータを抽出して 77 件に絞り込みを行い集計した。集計の際には<NIISubject>ではなく、<degreename>と<discription>から抽出した学位名を情報として用いた。

利用減衰に関する調査では、岸田(1994)[197]による図書貸出のオブソレッセンス測定手法を適用した^{[])}。先行研究における期間は 10 年間であり、図書館における図書の利用に着目していた。本研究では、8 年間の機関リポジトリに登録された OA コンテンツの利用を測定した。先行研究における批判として、貸出は利用と同義ではないが、操作的に定義を行わないと測定が不可能であるという指摘がなされていた。本研究における OA コンテンツにおいても、同様のことが言える。論文の利用として明確なものとして用いられてきている指標は、引用であり利用ではないというのが一般的に言える。だが、科研の調査報告書や、学位論文など引用をすることはあまりないが、利用されているコンテンツが機関リポジトリには存在する。本研究の範囲は、学術論文のみならず、機関リポジトリに登録されているコンテンツ全体なので、統一的に測定することが可能な利用という指標を用いている。そのため、利用者の行動という意味での利用ではなく、記録された結果としての行動を基に利用減衰を測定したことになる。OA コンテンツの利用を測る指標として、ダウンロード回数を用いることは一般的であり、本研究でも同様の理由で用いた。本研究では、受け入れ年度と利用年度でのクロス集計を実施することで、共時オブソレッセンスと通時オブソレッセンスの測定を実施した。なお、コンテンツ数による違いがあるため、純粋な数値に加えて、割合を求めて表示する。なお、メタデータの付与のミスであると考えられるデータについてもそのまま表示をしている。だが、分析の際は、外れ値として認識をしている。

平均反応時間の算出では、2 つの手法を適用した。2 つのどちらの手法においても、本研究では、単位を年度で区切っている。手法の 1 つ目は、単純集計から平均値を求める手法である。手法の 2 つ目は、平均反応時間の公式を用いた[198]。この公式は

$$MRT = -\ln(f_0 + f_1e^{-1} + f_2e^{-2} + f_3e^{-3} + f_4e^{-4})$$

で表される。(MRT は、Mean Reaction Time を表し、ln は、自然対数を指す。)
決められた期間内において、何年に最初の引用を受けたか、その件数の割合が各変数に入ることになる。本研究では、期間を 8 年間に拡張して利用しているため、公式に用いる数の数が異なる。そのため、本研究において用いた式は、
$$\text{MRT} = -\ln(f_0 + f_1e^{-1} + f_2e^{-2} + f_3e^{-3} + f_4e^{-4} + f_5e^{-5} + f_6e^{-6} + f_7e^{-7})$$
である。

2.3.3 調査対象 3: 検索語句のクリーニングと分析

本研究の対象は、機関リポジトリである。先行研究において、そのアクセスの大多数を占めるのが、他のサイトからの遷移によるアクセスであることが示されていた。そのような状況の中で、非文献コンテンツの利用にも注目が集まっている。非文献コンテンツの特徴として、図書館が使用しているものと異なるメタデータの付与が必要であり、検索においても全文検索などが難しくメタデータに対する検索を行う必要があることが示されていた。そこで、コンテンツに対する適切なメタデータの付与と利用者の適切な検索語句の入力が求められている。先行研究では、問題点の示唆とメタデータに対する言及はあるが、検索語句に関する言及はなされていない。本研究では、機関リポジトリの内部で検索されたログデータを用いて、検索語に対する分析を行った。

本研究において用いた検索語句は、リファラに付与されている語句ではなく、サーバ内に蓄積された検索語句である。つまり、サイト内検索によって得られた語句を使用している。先行研究で示されていた通り文献コンテンツは、全文検索が可能なため外部からの遷移によるアクセスが可能である。

2.3.4 検索語のスクリーニング

自然言語の解析を行う場合は、ストップワードを定義して、データのスクリーニングを行う必要がある。先行研究では、ストップワードの除去を行わない状態で検索語句の分析がなされていたが、詳細な分析の際に、ノイズになることが知られている。web サイトにおける検索では、html タグや検索クエリなどが含まれているため、機械的な検索語句などが含まれている。それらのノイズをストップワードとして、除去した。本来の自然言語処理では、文章を n-gram に分割して、単語を品詞分類した後に、ストップワードの除去を実施することで、必要な単語を抽出する。だが、今回は検索語句として実際に入力されたものを使用しているため、使用されている語句を名詞として処理する。ノイズとなっているストップワードを除去することで、ユーザによる検索語句と再定義した。

2.3.5 ストップワード

今回使用したストップワードは、以下の表に示す語句に加えて、html のタグなどを含むものである。参考にしたのは、Oracle[199]の提供するストップワードと NII[200]の提供するストップワードである。また、機械的な処理を行った後に、人力でファイルの結果を確認

し、["?.", "*function*", "=*"]などの要素を追加した。

表 12 Oracle の提供する stopword[199]

a	did	In	only	then	where
all	do	Into	onto	there	whether
almost	does	is	or	therefore	which
also	either	it	our	these	while
although	for	its	ours	they	who
an	from	just	s	this	whose
and	had	ll	shall	those	why
any	has	me	she	though	will
are	have	might	should	through	with
as	having	Mr	since	thus	would
at	he	Mrs	so	to	yet
be	her	Ms	some	too	you
because	here	my	still	until	your
been	hers	no	such	ve	yours
both	him	non	t	very	
but	his	nor	than	was	
by	how	not	that	we	
can	however	of	the	were	
could	i	on	their	what	
d	if	one	them	when	

表 13 NII の提供するストップワード[200]

!	[全	『	⊂	∞	☆	ℓ	┘
"	¥	↗	』	⊃	∴	★	♪	└
#]	/	【	⊄	∫	◦	†	┐
\$	^	\	】	⊅	∫	●	‡	┘
%	_	~	+	∧	°	◎	¶	┘
&	'	//	±	∨	'	◇	—	┘
({		×	⊆	"	□		┘
)		...	÷	⇒	°C	■	┘	┘
*	}	..	=	⇔	¥	△	┘	┘
+	—	'	≠	∇	\$	▲	┘	┘
,	。	(≤	∃	¢	▽	┘	┘
.	「)	≥	∠	£	▼	┘	┘
/	」	[∞	⊥	%	※	┘	┘
:	、]	∴	⌒	Å	〒	┘	┘
;	・	{	♂	∂	‰	→	┘	┘
<	˘	}	♀	∇	#	←	┘	┘
=	˙	<	∈	≡	&	↑	—	┘
>	ˆ	>	⊃	≡	*	↓		
?	—	≪	⊆	√	@	=	┘	
@	”	≫	⊇	∞	§	#	┘	

2.4 手法のまとめ

本研究では、アクセスログデータの中から IP アドレス、bitstream、検索語句の 3 要素に着目した。アクセスログデータからは、IP アドレスと bitstream の文字列を含むものを、COUNTER に従う形式でスクリーニングした後のデータを利用する。IP アドレスは、処理後のデータから逆引きを実施し Host を得た。本研究における利用の定義は、bitstream を含むログの 1 行が 1 回の利用となっている。また、weko の出力と合わせて分析対象とする。こちらに関しては、出力形式の都合上、データに処理が行われた後の形式なので、スクリーニング等の処理は実施していない。検索語句に関しては、ストップワードの機械的な除去を行った。上記の様に、データ種別でそれぞれに異なる処理を行い分析ができる形式を整えた。処理を終えたデータを集計し、分析を行うことで機関リポジトリの利用状況を明らかにした。

2.5 仮説

本研究において2つの仮説を立てた。1つめは、機関リポジトリに登録されたコンテンツにも、論文の被引用半減期のように、登録されてからある程度の時間が経過することで利用がされなくなるという仮説である。図書館の資料に関する利用の減衰は、リポジトリの利用を長期的に分析することによってのみ明らかにすることが可能である。過去の研究においては、調査対象期間が短期間であったため調査されていなかったが、本研究では長期間のデータを使用した分析が可能のため、この仮説に対する検証を実施する。

2つめは、機関リポジトリにコンテンツが登録されてから利用されるまでには、1年程度のタイムラグが発生するという仮説である。コンテンツの利用までの速度に関しては、実際にデータをスクリーニングした際の経験から得られた推測である。この仮説を検証するためにも複数年に渡るアクセスログデータが必須であるため、機関リポジトリにおける調査は行われていない。

上記の2つの仮説についてコンテンツ利用の割合の大きかった3type種別(紀要、学術論文、学位論文)を対象として検証を行った。

3. 結果

本研究では、筑波大学の2010年3月から2017年12月までのアクセスログを対象に分析を実施した。そのうち、2010年3月から2014年3月までは、DSpaceのアクセスログである。2014年5月から2017年12月までは、JAIR Cloudのアクセスログである。

この2つの期間におけるアクセスログを合わせて、2010年3月から2017年12月までの長期間のアクセスログを分析することで、機関リポジトリの長期的な利用実態を明らかにした。

3.1 アクセス元の分析

アクセス元に関して、2010年3月から2014年3月までアクセスログのIPアドレスからHostの逆引きを行なった。2014年5月から2017年12月までは、出力結果として、提供を受けたIPアドレスとHostのログを利用した。1月から12月を単位として年度ごとにIPアドレスを求めた結果を以下の表に示す。また、アクセス元の属性を明らかにするために、同一のアドレスからのアクセスをまとめたデータに関しては、ユニークと表記した。各年度に行ったIPアドレスのユニーク化の結果も同時に記載する。以下では、各年度IPアドレスの逆引きを行った際に、以下では、Hostがリクエストに応答した結果を母数として扱う。結果を以下の表に示す。

表 14 各年度のIPアドレス(2010年3月-2014年3月)

IP	2010	2011	2012	2013	2014
http(ユニーク)	255,077	20,668	148,278	152,269	33,027
https(ユニーク)	46,635	63,809	62,047	105,864	624
http(全件)	633,984	111,672	476,052	518,058	39,008
https(全件)	95,626	108,766	95,626	1,864,538	757

DSpaceのアクセスログから年度ごとにアクセス元の属性を明らかにした。

表 15 各年度のIPアドレス(2014年5月-2017年12月)

IP	2014	2015	2016	2017
weko(ユニーク)	13,236	11,356	25,576	39,936
weko(全件)	39,567	16,621	39,575	64,441

WEKOのアクセスログから年度ごとにアクセス元の属性を明らかにした。

本研究では、COUNTERに準じた処理を行った後のアクセスログの中からユニークなIPアドレスを求めて、利用元を明らかにした。IPアドレスをユニーク化したのは、アク

セス元の属性を明らかにするためである。コンテンツの利用状況を明らかにするための基準には次節で述べるダウンロード数(利用数)を用いた。

利用元の属性を明らかにするために、Host の分類を実施した。分類の際の注意事項として、現在 gTLD²² は、大幅に増大している。そのため、ドメインを参考にデータの分類を行ったが、すべての gTLD を網羅し、分類しているわけではなく、代表的なドメインに絞って分類を実施している。分類基準は佐藤(2009)を参考に実施した[176]。国別アクセスに関しては、jp ドメインとそれ以外に分類した。(jp は、日本を表す。)また、属性を表すドメインを 6 つに分類した。分類内容は、民間・個人(ne,net)、民間・団体(or, org)、大学等(ac, edu)、企業(go, gov)、その他である。

3.1.1 アクセス元の調査と内訳の結果

2010 年 3 月から 2014 年 3 月までの期間に関しては、民間・個人に分類した Host からのアクセス割合が 50%以上であった。次点で多いのは、大学等教育機関だが、2011 年の 1 年度を除き、20%以下の割合であった。通信方式を考慮せず期間のみで考えると 10%程度である。https 通信による 2014 年度のデータに関しては、国外からのアクセスと企業からアクセスの割合が大きくなっているが、サンプルサイズが他と比べて小さいため、1 つのアクセス元の影響が大きくなった結果であると考えられる。この期間を除くと 2010 年から 2014 年までの期間では、国内からのアクセスの割合が 70%以上を占めていた。以下に、各期間における Host の割合を示す。

表 16 2010 年 3 月から 2014 年 3 月における http 通信のアクセス元の内訳

	2010	2011	2012	2013	2014
国内	83.00%	87.52%	80.88%	75.70%	72.93%
非国内	17.00%	12.48%	19.12%	24.30%	27.07%
	2010	2011	2012	2013	2014
ne,net	69.73%	54.71%	66.35%	63.90%	56.34%
or,org	8.49%	7.09%	7.14%	5.98%	5.29%
ac,edu	5.38%	22.47%	9.02%	8.80%	14.62%
co,com	2.88%	3.10%	4.00%	9.21%	13.12%
go,gov	0.27%	0.48%	0.32%	0.23%	0.33%
Other	13.25%	12.16%	13.17%	11.89%	10.30%

²² gTLD: Generic Top Level Domain” gTLD は、特定の領域・分野ごとに割り当てられたトップレベルドメインで、一般的に地理的制限なしに世界のどこからでも登録することが可能です。”とされている[199]。

国内からのアクセスが 70%以上を占めている。また、民間・個人のプロバイタからのアクセスが 50%以上を占めている。大学等教育機関は、上位 5 番目には入る集団であるが集団としての母数の問題か数は多くない。

表 17 2010 年 3 月から 2014 年 3 月における https 通信のアクセス元の内訳

	2010ssl	2011ssl	2012ssl	2013ssl	2014ssl
国内	84.54%	86.16%	83.69%	77.63%	25.66%
非国内	15.46%	13.84%	16.31%	22.37%	74.34%
	2010ssl	2011ssl	2012ssl	2013ssl	2014ssl
ne,net	61.66%	62.84%	62.40%	63.53%	8.68%
or,org	7.30%	6.80%	6.20%	5.19%	0.38%
ac,edu	14.11%	13.60%	14.28%	9.80%	11.32%
co,com	5.11%	3.14%	3.79%	10.64%	71.70%
go,gov	0.50%	0.41%	0.44%	0.25%	0.00%
Other	11.32%	13.21%	12.89%	10.59%	7.92%

http の結果と大きな差異は、基本的には見られなかった。2014 年の結果に関しては、他の年度と比較し、サンプルサイズが小さく 1 つの要素の影響を強く反映した結果となった。

表 18 2010 年 3 月から 2014 年 3 月におけるアクセス元の内訳

	2014	2015	2016	2017
国内	69.54%	73.02%	46.75%	26.52%
非国内	30.46%	26.98%	53.25%	73.48%
	2014	2015	2016	2017
ne,net	42.81%	46.62%	41.09%	40.26%
or,org	4.26%	4.94%	3.26%	1.87%
ac,edu	21.74%	21.59%	14.72%	6.93%
co,com	9.40%	9.09%	13.72%	18.22%
go,gov	0.75%	0.70%	0.51%	0.35%
Other	21.03%	17.06%	26.70%	32.37%

2015 年度を境に、アクセス元に大きな変化が現れたことがわかる。2015 年度までは、国内からのアクセスが全体の 70%近くを占めていたが、2016 年度から国内を示すドメイン以外からのアクセスが急激に増加した。2014 年度に関しては、比率が逆転し、国内以外からのアクセスが 70%を超える結果となった。また、民間・個人からのアクセスが 50%を下回

っている点で、2014 年までと差異がみられた。

上記の分析では、2013 年以降の日本国内からのアクセス割合の低下や教育機関や企業からのアクセスの増加の傾向があることを示した。だが、上記の分類では属性以外は不明である。そこで、詳細なアクセス元を明らかにすることを目的として、年ごとに、トップレベルドメイン、セカンドレベルドメインのランキングを 10 件ずつ抽出し、次節以降にて表で示す。

この結果におけるトップレベルドメインやセカンドレベルドメインに関しては、Host に含まれている”.”の数によって、機械的に抽出している。そのため、厳密な定義におけるドメインと異なる結果も含まれている。本研究では、操作的定義として、Host における右から 1 番目の”.”の右側に登場する文字列をトップレベルドメイン、2 番目の”.”右側からトップレベルドメインまでの間の文字列をセカンドレベルドメインとして、分類している。なお、分類使用したデータは、右から 3 番目までの”.”で区切っているため、国別のコードなどを含み、セカンドレベルドメインが右から 3 番目の”.”以降にある場合は集計に反映されている。

3.1.2 トップレベルドメイン別のアクセス元の内訳

2010 年 3 月から 2014 年 3 月までのアクセスをトップレベルドメインで見ると「jp」「net」「com」「cn」「de」が特徴的である。国別のドメインでは、「jp」が日本、「cn」が中国、「de」はドイツである。その他のドメインでは、「net」が一般プロバイダで「com」が商用向けを表すドメインである。

2014 年 5 月から 2017 年 12 月までのアクセスをトップレベルドメインで見ると「jp」「net」「com」「cn」「de」「br」が特徴的である。上記に登場していない国別のドメインでは、「in」がインドを表しており、「br」がブラジルを表している。

2016 年から中国を表す「cn」からのアクセスが減少し、ブラジルを表す「br」からのアクセスが増加している。また、2014 年以降では、「info」や「adsl」などのドメインも登場している。

表 19 http 通信の Host Top10 (トップレベルドメイン)

year	2010		2011		2012		2013		2014	
rank	domain	count	domain	count	domain	count	domain	count	domain	count
1	jp	162,969	jp	12,281	jp	90,514	jp	87,430	jp	18,077
2	net	21,426	net	1,205	net	13,320	net	13,924	net	3,051
3	com	2,382	com	170	com	1,974	cn	5,527	cn	1,829
4	de	971	cn	67	cn	836	com	3,612	com	971
5	cn	916	kz	41	de	555	de	506	de	103
6	tw	544	de	27	tw	330	in	368	tw	82
7	in	441	th	24	ru	318	tw	246	in	50
8	ru	439	tw	22	in	255	edu	216	fr	37
9	it	379	fr	19	fr	220	fr	201	edu	35
10	fr	343	adsl	11	edu	206	br	169	br	34

3.1.1 で見た jp ドメインの割合を示す結果となった。教育機関を表す edu と国別のドメインで、「kz」カザフスタン、「tw」台湾、「ru」ロシア、「in」インド、「de」ドイツ、「fr」フランス、「cn」中国など多様な国や地域からのアクセスがあることがわかる。

表 20 https 通信の HostTop10 (トップレベルドメイン)

year	2010ssl		2011ssl		2012ssl		2013ssl		2014ssl	
rank	domain	count	domain	count	domain	count	domain	count	domain	count
1	jp	27,375	jp	40,308	jp	38,226	jp	63,826	cn	104
2	net	3,032	net	4,382	net	4,775	net	8,082	Com	84
3	com	1,110	com	781	cn	203	in	72	Jp	68
4	cn	140	cn	181	com	967	de	252	Net	2
5	de	90	de	142	bd	2	fr	76	Info	1
6	tw	82	tw	84	sg	20	eu	112	It	1
7	th	34	edu	51	ua	56	com	2,289	Pl	1
8	in	33	in	43	at	10	th	117	us	1
9	edu	27	it	43	vn	20	np	9		
10	it	27	tr	39	tw	100	cn	5,699		

ドメインに 4 文字の文字列が目立つ「adsl」「arpa」「info」などこれらは、国や地域を表すドメインではなく、新しく追加された gTDL であると考えられる。アクセス元として、欧州の国家が多くみられるのが特徴的である。

表 21 https 通信 Top10 の Host (トップレベルドメイン)

year	2014		2015		2016		2017	
rank	domain	count	domain	count	domain	count	domain	count
1	jp	6,653	jp	8,292	jp	12,086	jp	10,593
2	adsl	814	net	972	net	3,488	net	7,247
3	net	757	com	660	com	1,791	com	3,141
4	com	543	cn	548	in	585	br	2,160
5	cn	169	ru	138	br	483	in	1,526
6	de	91	de	68	adsl	327	it	875
7	edu	46	eu	54	pl	320	pl	757
8	eu	33	pl	53	cn	309	adsl	638
9	uk	31	org	36	ru	257	rs	587
10	ru	27	ua	36	rs	246	ro	540

2016 年以降の jp ドメインの割合が減少したことを示した。その理由として、jp ドメインが減少しているのではなく、他のドメインからのアクセスが急増しているのが理由であることが判明した。特に、「br」ブラジル、「in」インド、「rs」セルビアが上位である。

3.1.3 セカンドドメイン別のアクセス元の内訳

2010 年 3 月から 2014 年 3 月までの期間におけるセカンドレベルドメインは、「ne.jp」「ac.jp」「or.jp」「ad.jp」「bbtec.net」が特徴的である。「ne.jp」は、日本の個人または法人の提供するドメイン、「ac.jp」は、日本の大学等の教育機関を表すドメイン、「or.jp」は日本の組織を表すドメイン、「ad.jp」は、JPNIC²³会員となっている組織を表すドメイン、「bbtec.net」はソフトバンク回線を表すドメインである。http 通信と https 通信では、機械によるアクセスの行程に差が出るため、アクセス元の企業に関する違いが懸念されていたが、大きな差異は現れなかった。

2014 年 5 月から 2017 年 12 月の期間におけるセカンドレベルドメインでは、「ne.jp」「ac.jp」「or.jp」「pldt.net」「bbtec.net」「ny.adsl」が特徴的である。「pldt.net」と、「ny.adsl」に関しては、web で詳細を検索した結果スパムとしての報告事例が見つかった。また、「google」「amazon」「naver」などの企業の名称を得ることができた。これらの企業は、web サービスを扱っているため、機械によるアクセスである可能性が高い。

²³ 日本ネットワークインフォメーションセンター

表 22 http 通信の Host Top10 (セカンドレベルドメイン)

year	2010		2011		2012		2013		2014	
rank	domain	count	domain	count	domain	count	domain	count	domain	count
1	ne.jp	114,214	ne.jp	6,431	ne.jp	60,135	ne.jp	59,018	ne.jp	10,774
2	or.jp	16,601	ac.jp	3,117	bbtec.net	9,507	bbtec.net	9,943	ac.jp	3,523
3	bbtec.net	14,786	or.jp	994	ac.jp	9,433	ac.jp	9,559	bbtec.net	1,876
4	ad.jp	12,590	ad.jp	909	or.jp	7,945	or.jp	6,862	com.cn	1,796
5	ac.jp	9,619	bbtec.net	780	ad.jp	5,664	com.cn	5,353	or.jp	1,305
6	hinet.net	1,476	uqwimax.jp	210	uqwimax.jp	2,331	ad.jp	4,791	ad.jp	812
7	co.jp	1,305	co.jp	169	co.jp	905	uqwimax.jp	2,382	msn.com	485
8	bbiq.jp	1,224	bbexcite.jp	84	bbexcite.jp	691	amazonaws.com	1,218	uqwimax.jp	390
9	bbexcite.jp	1,171	asianetcom.net	66	com.cn	651	co.jp	880	co.jp	344
10	commufa.jp	1,132	go.jp	66	mopera.net	636	msn.com	750	hinet.net	330

民間のプロバイダからのアクセスの多さと、企業名がわかるドメインが特徴的である。

表 23 https 通信の IP アドレス Top10(セカンドレベルドメイン)

year	2010ssl		2011ssl		2012ssl		2013ssl		2014ssl	
rank	domain	count	domain	count	domain	count	domain	count	domain	count
1	ne.jp	16,824	ne.jp	24,877	ne.jp	23,542	ne.jp	43,936	com.cn	105
2	ac.jp	4,450	ac.jp	6,198	ac.jp	6,314	ac.jp	7,752	reverse.com	75
3	or.jp	2,353	bbtec.net	3,209	bbtec.net	3,549	bbtec.net	6,193	ac.jp	30
4	ad.jp	2,209	or.jp	3,171	or.jp	2,822	com.cn	5,620	ne.jp	21
5	bbtec.net	2,117	ad.jp	3,087	ad.jp	2,470	or.jp	4,249	naver.jp	13
6	googlebot.com	568	uqwimax.jp	1,206	uqwimax.jp	1,289	ad.jp	3,341	google.com	3
7	co.jp	305	co.jp	343	co.jp	349	uqwimax.jp	1,977	msn.com	3

						jp			
8	hinet.net	281p	bbexcite.j	324.jp	269msn.com	623co.jp			2
9	bbexcite.j	247et	mopera.n	mopera.	254ws.com	570istella.it			1
10	google.co	175hinet.net		amazonaws.co	242co.jp	451s.info	limenetwork		1

http の結果との大きな差異は見られなかった。ともに、特徴的な部分として大学等機関からのアクセスを示す「ac.jp」によるアクセスが上位にきている点である。企業からのアクセスでも、明確な企業名を示したドメインであることが多く、利用者が明確である。サンプルサイズの小さな 2014ssl の結果においても、国別のドメインが付与されているが企業からのアクセスであることが判明した。

表 24 https 通信の IP アドレス Top10(セカンドレベルドメイン)

year	2014		2015		2016		2017	
rank	domain	count	domain	count	domain	count	domain	count
1	ne.jp	3,283	ne.jp	3,283	ne.jp	6,153	ne.jp	5,899
2	ac.jp	1,973	ac.jp	1,973	ac.jp	3,697	ac.jp	2,675
3	ny.adsl	814	or.jp	814	or.jp	721	com.br	1,341
4	or.jp	398	bbtec.net	398	bbtec.net	670	pldt.net	1,260
5	bbtec.net	306	sx.cn	306	pldt.net	400	bbtec.net	847
6	ad.jp	245	ad.jp	245	ad.jp	355	net.br	790
7	co.jp	173	co.jp	173	ny.adsl	325	or.jp	672
8	com.cn	143	amazonaws.com	143	amazonaws.com	296	ny.adsl	638
9	amazonaws.com	116	com.cn	116	com.cn	290	airtelbroadband.in	495
10	hichina.com	102	baidu.com	102	com.br	280	tedata.net	488

2014 年 3 月までのアクセス元では、企業名のドメインが一般的に知られている企業が多くみられた。だが、2014 年以降では、身元の判然としないアクセス元も見られるようになった。2015 年まで上位に来ていたドメインが 2017 年になると大きく変化していた。

3.1.4 アクセス元に関する考察

結果 1 の期間と結果 2 の期間における結果の違いとして、結果 2 には **name** サーバを立てたと考えられるアクセス元が現れた。また、アクセス元の国家として、多く見られたドイツ、中国、ブラジルに関しては、科学技術指標²⁴を参考にするといずれも論文生産数が上位 15 件以内にいる国家である。中国に関しては、近年の研究開発費の増加が著しいことが知られているが、リポジトリへのアクセス数は、減少傾向にあることが判明した。

本研究では、年度ごとに、アクセス元の IP を参考にユニークな **Host** を求めて結果を出している。そのため、多様性は明らかになっても数は明らかにしていない。そのうえで、セカンドレベルドメインまでを抽出した上位の結果に表れるのは、多数のクローラを動かしている結果である。NII の提供するクローラリストは、2016 年に更新が停止しているが、多数の IP を持つ国際的企業による Web 情報収集は行われているため、アクセスログを分析するためにも、クローラのリストの更新は必要である。

特に、2016 年や 2017 年の結果に表れている日本国外からアクセスの割合が高いのは、トップレベルドメインの多様化とともに、機械によるアクセスの割合の増加が背景にあると考えられる。

本研究では、アクセスの回数ではなく、アクセス元の属性を調査した。だが、表.14 や表.15 の結果を参考にしてもわかる通り、アクセス回数を考慮した結果を求めると重複が多数発生するので、結果が異なる可能性がある。

3.2 コンテンツの利用に関する分析

学術情報の利用の定義は複数ある。論文における引用回数は、図書における貸出回数などが計量的な指標として用いられてきた。引用に関しては、信頼性の高い量的指標として軽量書誌学で多く用いられてきている。貸出回数は、貸出という行為を測定することは可能だが、「読んだ」という利用者の行動を表すことはできない操作的定義である。

本研究における利用の定義は、コンテンツが 1 回ダウンロードされたことを 1 回の利用としてカウントする。そのために、「**bitstream**」の文字列を用いる。本研究における定義も操作的定義であるが、ダウンロードや Web ブラウザ上での閲覧は、引用よりも即時性の高い指数として、知られている。また、機関リポジトリは大学所属の研究者の学術的生産物を公開する場としての役割を担っているため、必ずしも利用のあったコンテンツが引用されるとは限らない。そのため、本研究で用いている利用は、引用よりも幅の広いコンテンツに対して適用可能である。

本研究では、今後、機関リポジトリの機能が拡張して、データリポジトリとしての役割を果たす可能性も考慮し、軽量書誌学で伝統的に用いられてきた引用ではなく、利用や、社会的インパクトの指標であるオルトメトリクスに着目して機関リポジトリの利用実態を明らかにした。

²⁴ NISTEP が発表している研究開発に関わる指標である[203]。

3.2.1 全期間を対象としたコンテンツの利用実態

本研究の対象期間におけるすべてのコンテンツの利用に関する利用実態を概観するために、以下に 2010 年 3 月から 2017 年 12 月の間に登録されたコンテンツのメタデータから、コンテンツに関する基本的な情報を抽出しまとめた結果を以下に示す。内容は、コンテンツ全体の利用回数の経年変化、年度ごとのコンテンツ登録数、登録コンテンツの type 種別、コンテンツ記述言語である。これらの割合からアクセス先のコンテンツの特徴を明らかにした。また、利用の多かったコンテンツの上位 15 件を表にまとめた。

表 25 機関リポジトリに登録されたコンテンツ利用回数の経年変化

	Dspace					JAIRO Cloud				
	2010	2011	2012	2013	2014	2014	2015	2016	2017	合計
http	67,270	6,545	63,334	145,882	16,734	201,598	732,852	1,322,034	2,420,423	5,281,592
https	85,345	118,422	61,988	34,644	4,521					
合計	152,615	124,967	125,322	180,526	21,255	201,598	732,852	1,322,034	2,420,423	5,281,592

コンテンツのダウンロード回数を指標に、利用と定義した。利用回数は、2011 年以降は年を重ねるごとに単調増加をしていた。2014 年までは、ほぼ横ばいとも呼べる変化率であったのが、2015 年以降は、指数関数的に増加した。

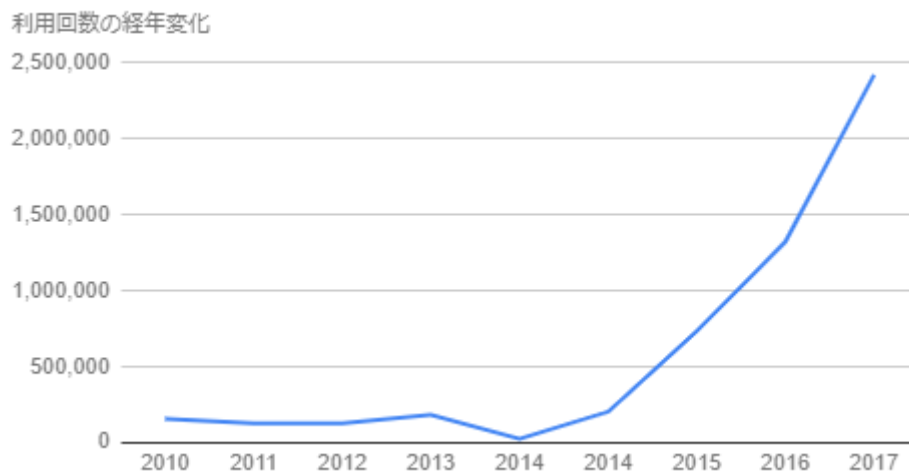


図 16 機関リポジトリに登録されたコンテンツの利用回数の経年変化

表 25 より作成したグラフである。2014 年に一度減少して見えるのは、DSpace の期間と JAIRO Cloud の期間でデータを分割した状態でグラフを作成したためである。

表 26 機関リポジトリに登録されたコンテンツの登録数推移

登録年	登録数	利用回数
2007	15,140	1,778,918
2008	1,932	224,349
2009	1,455	208,768
2010	2,715	487,235
2011	4,578	582,263
2012	1,851	289,902
2013	1,252	200,894
2014	2,574	420,493
2015	4,307	646,043
2016	3,336	296,853
2017	4,356	145,874
総数	43,497	5,281,592

筑波大学の機関リポジトリのメタデータを参照し、年度ごとの登録数と総利用回数の変遷を調査した。2007 年の数が多いのは、それまでに発表され、機関リポジトリの運用開始共に、公開されたコンテンツが存在したためであると考えられる。

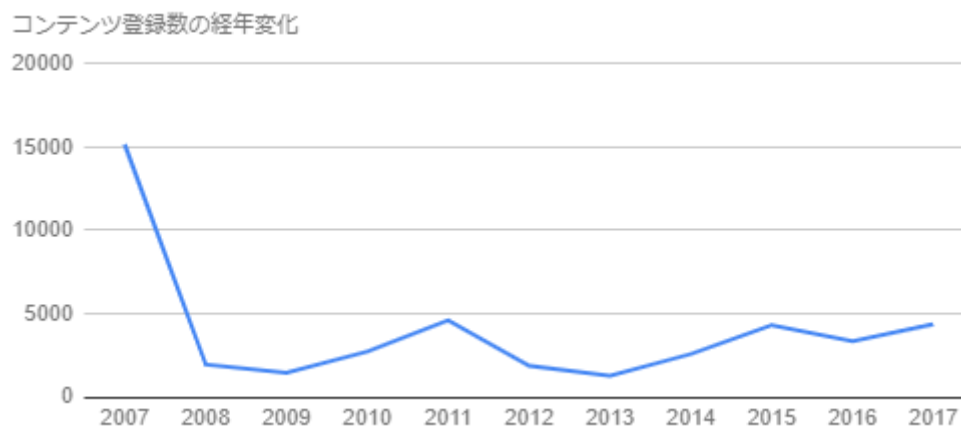


図 17 コンテンツ登録数の経年変化

2007 年以降は、年間で 1,000 件以上 5,000 件以内のコンテンツを毎年登録している。登録数には、機関リポジトリのシステムの変更などの影響はあまりなかったと考えられる。

表 27 type 種別別の経年変化(登録数)

year	2010	2011	2012	2013	2014	2015	2016	2017
Article	26	24	2		17	1	45	97
Book	19	11	6	4	9		4	8
Conference Paper	27	84	8	11	19		52	73
Departmental Bulletin Paper	618	1,946	733	561	651	71	499	1,497
Journal Article	420	798	387	291	370	17	1,214	1,006
Learning Material	1	2	1	12	8	3	28	20
Others	285	33	20	7	90		141	179
Preprint	1	2	8	2	1		2	1
Presentation	11	36	22	9	23	1	5	6
Research Paper	254	181	258	307	338	5	307	433
Technical Report	41	383	25	10	11			2
Thesis or Dissertation	1,012	1,078	381	38	1,037	130	1,039	1,034

分野別の登録コンテンツ数を調査した。機関リポジトリには、毎年コンテンツが登録されているが、2014 年の学位論文の登録数の変化などを調査した。

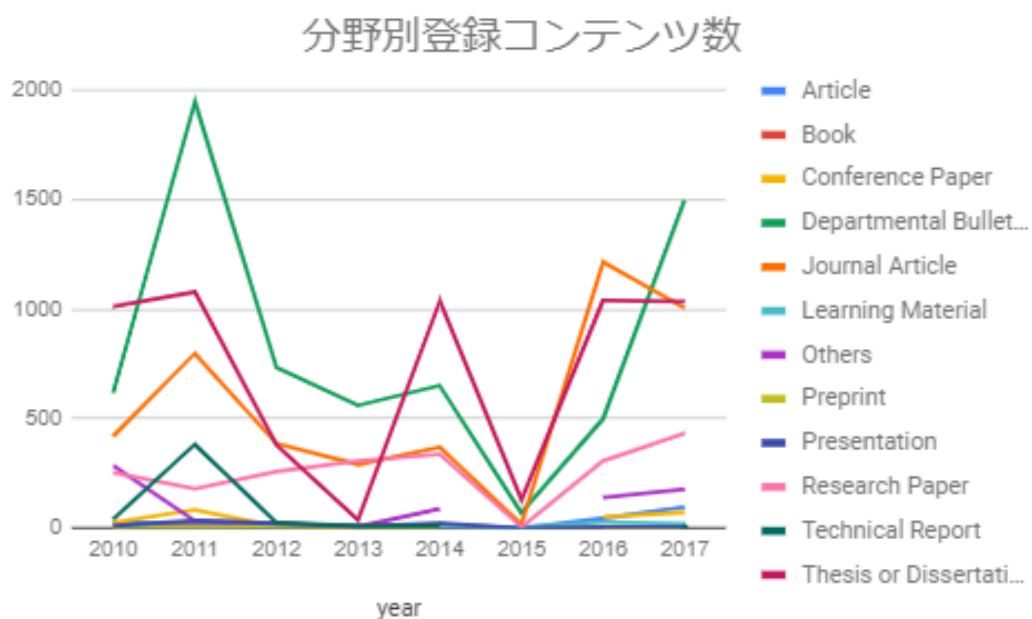


図 18 利用回数上位 1,611 件における学位論文の経年変化(登録数)

表. 27 をグラフで可視化した結果である。

表 28 利用回数上位 1,611 件における学位論文の経年変化(利用回数)

year	2010	2011	2012	2013	2014	2015	2016	2017
Article	1,590	1,659	515		1,523	563	2,354	1,458
Book	11,967	3,737	3,220	1,556	4,625		1,255	2,826
Conference Paper	8,762	10,256	4,698	1,555	3,266		2,372	1,595
Departmental Bulletin Paper	152,926	299,441	158,571	127,453	153,417	81,083	75,764	58,877
Journal Article	87,105	117,634	52,219	30,186	54,851	31,505	85,068	36,905
Learning Material	137	4,469	2,327	3,259	7,392	3,101	6,757	2,749
Others	8,568	2,088	1,224	208	4,679		3,754	1,955
Preprint	56	104	352	155	27		66	3
Presentation	2,219	3,688	2,687	743	1,734	651	1,498	261
Research Paper	22,086	11,512	26,780	24,897	21,960	4,259	27,465	3,508
Technical Report	2,853	22,412	4,207	551	494			81
Thesis or Dissertation	188,966	105,263	33,102	10,331	166,525	207,302	90,500	35,656

機関リポジトリに登録されたコンテンツの利用状況を概観するために、コンテンツ type 種別とコンテンツの登録された年度ごとの利用数を調査した。

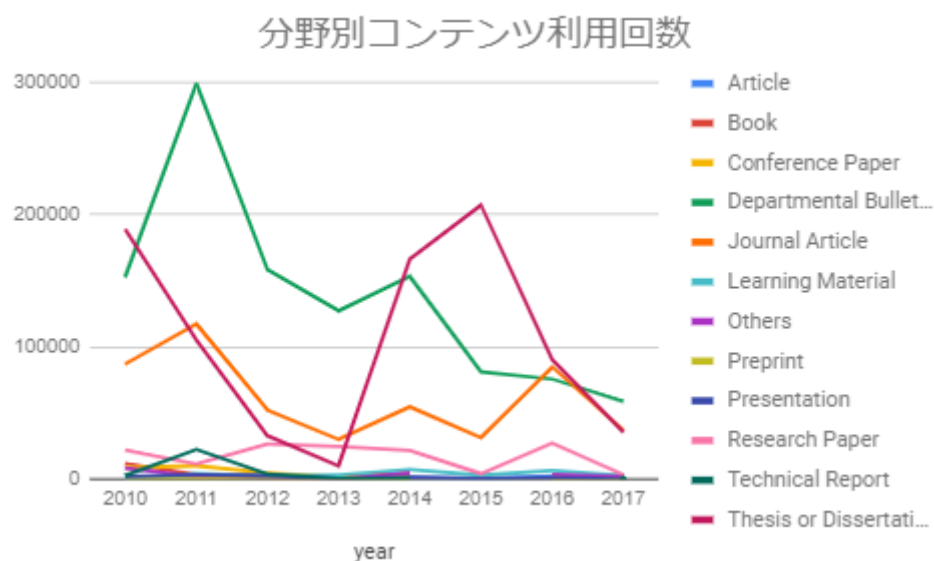


図 19 利用回数上位 1,611 件における学位論文の経年変化(利用回数)

表.28 をグラフで可視化した。

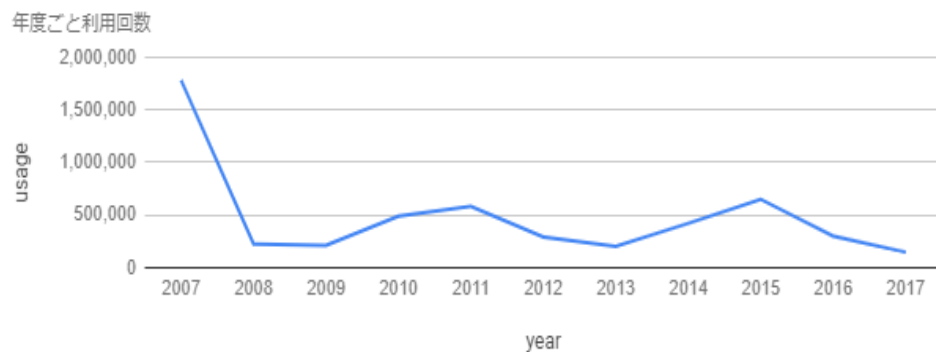


図 20 コンテンツの登録年度別利用回数の変化

登録年度ごとの資料の利用回数では、2007 年に登録された資料数が多い影響もあってか、特出して利用回数が多かった。コンテンツ全体で見ると、登録年度が古いコンテンツの方が、公開されてから時間が経過しているため、利用回数が多くなると考えられたが、コンテンツ利用回数の増加が著しいのは、2015 年からであるためか、登録年数の長さによる影響は見られなかった。

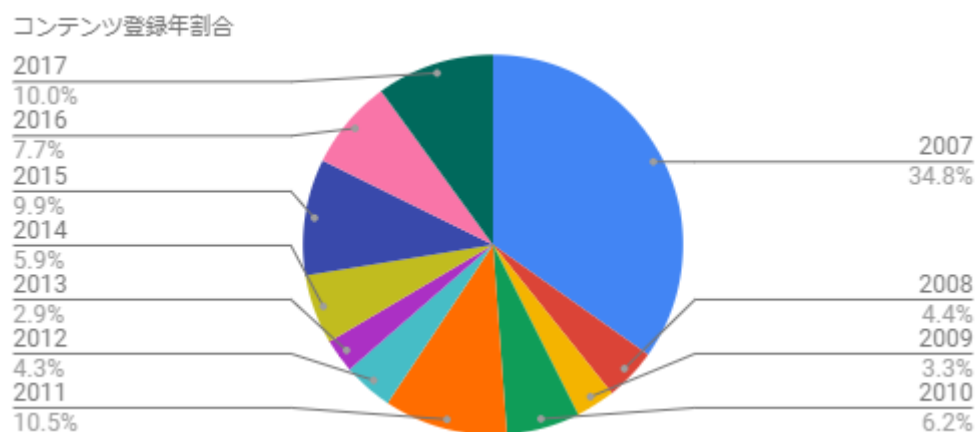


図 21 コンテンツ登録年度割合

コンテンツの年度ごとの登録割合をグラフによって、可視化した。

表 29 機関リポジトリに登録されたコンテンツの記述言語数

lang	コンテンツ数
jpn	35,746
eng	7,647
fre	36
ger	25
rus	16
no-data	14
zho	9
spa	1
Others	1
ita	1

機関リポジトリは、大学の研究者の生産物の保存庫であり、海外のリポジトリネットワークからの検索もコンテンツの検索が可能になっている。そのため、どのような言語のコンテンツがあるのかを調査した。

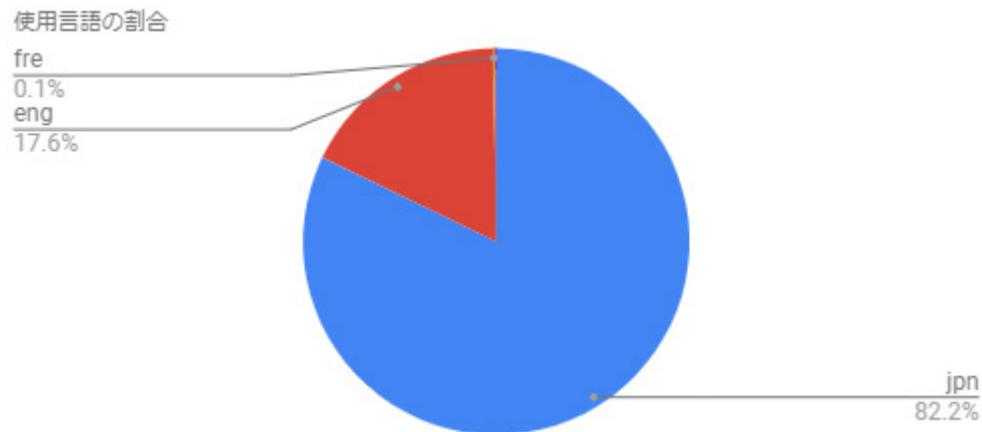


図 22 コンテンツ記述言語割合

調査結果をグラフにして可視化した。機関リポジトリ全体で見ると、日本語コンテンツの割合が 80%以上を占めており、英語と日本語を合わせると 99%以上である。

表 30 機関リポジトリに登録されたコンテンツの type 種別コンテンツ数

type	コンテンツ数
Departmental BulletinPaper	17,950
Thesis or Dissertation	13,131
Journal Article	6,621
Research Paper	2,818
Others	937
Article	794
Technical Report	474
Conference Paper	365
Presentation	137
Book	103
Learning Material	96
no-data	46
Preprint	24

日本の機関リポジトリの特徴として、諸外国と比較して²⁵紀要の割合が多いことが知られている。その中で、本研究の対象の筑波大学は紀要だけでなく、学术论文の割合が高いことが特徴である。

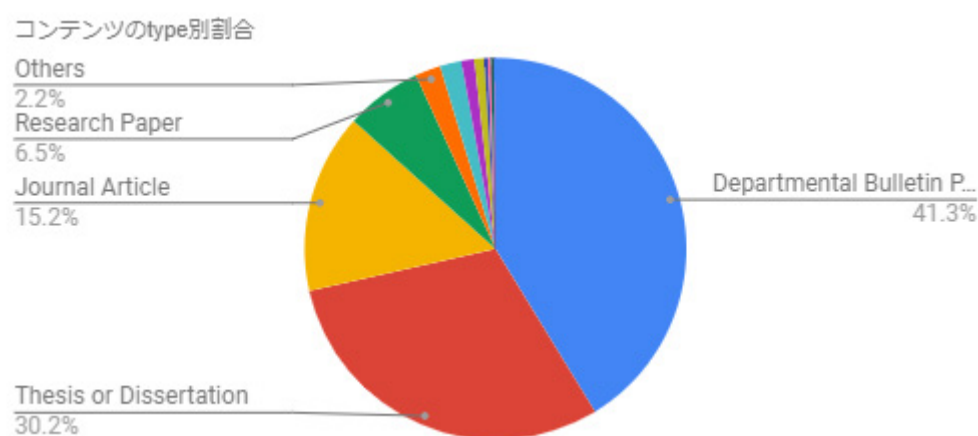


図 23 登録コンテンツの type 別割合²⁶

²⁵ BASE を参考にした割合を国立情報学研究所が発表している[202]。

²⁶ Departmental BulletinPaper は紀要であり、Journal Article は雑誌論文であり、Thesis or Dissertation は学位論文を指すメタデータである。登録年度に関しては、2007年から2106年まで幅広く存在していた。3.2.3、3.2.4、3.2.5において、この3つのtypeに焦点を当てた分析を実施する。

表. 30 をグラフにして可視化した。上位 3 コンテンツで 85%以上を占めている。

表 31 コンテンツ利用上位 15 件

rank	title	type	year	count
1	＜原著＞ローゼンバーグ自尊感情尺度日本語版の検討	Departmental BulletinPaper	2007	34,263
2	Coherent gigahertz phonons in Ge ₂ Sb ₂ Te ₅ phase-change materials	Journal Article	2016	21,535
3	バランス能力の因子構造に及ぼす加齢の影響	Thesis or Dissertation	2011	16,719
4	Lack of Negative Correlation in Glucose Dynamics by Nonexercise Activity Thermogenesis Restriction in Healthy Adults	Journal Article	2014	11,959
5	光トポグラフィによる歩行動作の評価	Journal Article	2010	11,869
6	直腸がん前方切除術後の排便障害を軽減する看護支援に関する研究	Thesis or Dissertation	2015	11,279
7	イメージ操作は脳の中でどのように行われるのか：心的回転の回転方向に関する実験的検討	Thesis or Dissertation	2014	9,991
8	水中に懸濁した微粒子の凝集分散--基礎理論とその適用性	Journal Article	2015	9,783
9	＜翻訳＞ヤング グッドマン ブラウン	Departmental BulletinPaper	2011	9,711
10	環境倫理学の課題 ―世代間倫理をめぐって―	Departmental BulletinPaper	2007	9,255
11	〈Research Reports〉 Mapping the human settlement of South East Asian cities using ALOS AVNIR-2	Departmental BulletinPaper	2012	8,302
12	体育授業における教師の効果的なフィードバック行動に関する検討	Thesis or Dissertation	2010	8,090
13	内発的―外発的動機づけ測定尺度の開発	Departmental BulletinPaper	2011	7,920
14	コントロールテストからみた跳躍競技の種目特性および競技パフォーマンスとの関係	Journal Article	2009	7,491
15	〈特別寄稿〉デンマークの教育制度：国民を幸福にする教育と教育制度の探究（試論）	Departmental BulletinPaper	2012	7,331

紀要、学術論文、学位論文が上位 15 件に表れた。登録年度はあまり影響がなかった。

3.2.2 ダウンロード回数の多いコンテンツの利用実態

表. 31 において、本研究の対象期間において利用の多かったコンテンツの実態を調査した。上位 15 件のデータを見ると、登場したコンテンツ type 種別は、紀要、学术论文、学位論文であった。

本節では、利用の多かったコンテンツの実態をより詳細に知ることが、利用実態を明らかにすることになると考え、利用回数が 520 回より多いコンテンツを対象に、登録年度、登録年度、type 種別、を調査した。利用回数 520 回を基準とした理由は、全期間におけるコンテンツ利用回数の平均値が 121 であり、標準偏差が 399 であるため、標準よりも利用回数の多い群を対象とした。全 43,497 件のコンテンツのうち、上記の基準に適合した 1,611 件を対象が対象である。

利用回数の偏りから判断して、利用回数の多いコンテンツは、コンテンツ利用全体に対して強い影響を与えているといえる。登録コンテンツ全体に対して、利用回数が多いコンテンツの特徴が現れたのは、3 点ある。

1. コンテンツ全体の言語比率に占める日本語で書かれたコンテンツの利用の割合が大きい
2. コンテンツ利用割合に占める紀要の利用割合が大きい
3. 2015 年に登録されたコンテンツの利用が特筆して多い

1.に関しては、アクセス元の Host の調査の結果と一致する。2.に関しては、3.2.3 にて詳細を調査した。3. は、3.2.5 において詳細に調査した。

表 32 登録年度とコンテンツ数 (N=1,610)

登録年度	コンテンツ数
2007	518
2008	52
2009	60
2010	165
2011	154
2012	97
2013	74
2014	141
2015	228
2016	80
2017	41

利用回数 520 回以上のコンテンツの登録年度別コンテンツ数を示した。



図 24 コンテンツ数の登録年度推移

年度別のコンテンツ登録数を可視化した。利用回数が 520 回以上のコンテンツを抽出したが、グラフの傾向はコンテンツ全体で見た際とあまり差異が見られなかった。

表 33 コンテンツの言語比率(N=1,610)

lang	コンテンツ数
jpn	1,489
eng	121

利用回数が 520 回のコンテンツの記述言語を明らかにした。全体の 99%が日本語と英語であることは前述したが、利用数が多いものに関しては、100%となった。

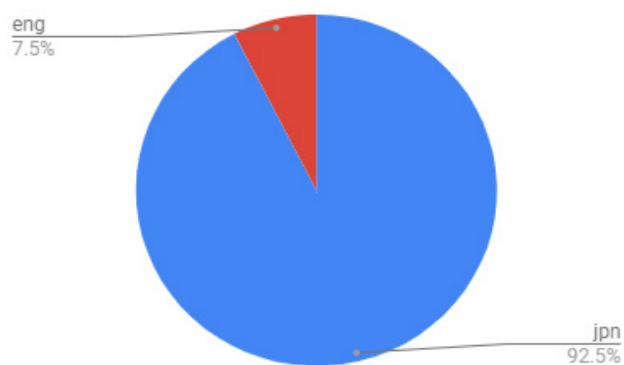


図 25 コンテンツの言語比率(N=1,611)

表.33 の結果をグラフで可視化した。全体と比較し 10%近く日本語の割合が増加した。

表 34 利用回数 520 以上のコンテンツの type 種別(N=1,611)

NIItype	コンテンツ数
Departmental BulletinPaper	925
Thesis or Dissertation	399
Journal Article	182
Research Paper	38
Book	24
Learning Material	16
Conference Paper	14
Technical Report	7
Presentation	3
Others	1
Article	1

利用回数が 520 回のコンテンツについての type 種別を調査した。結果として、紀要、学位論文、学術論文の順に利用が多かった。この結果は、全体の傾向と一致した。

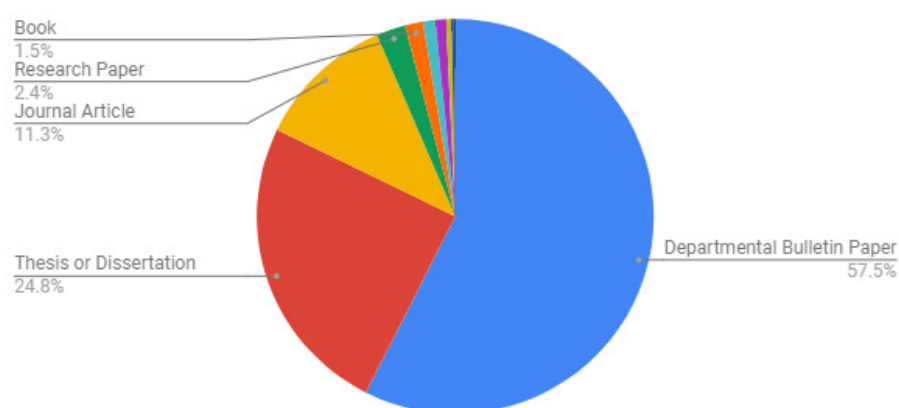


図 26 コンテンツ type 種別比率

表. 34 の結果をグラフで可視化した。全体の傾向では、紀要の比率は 50%以下であったのに対して、利用回数が 520 回以上のコンテンツ占める紀要の割合は 50%以上である。学

位論文の割合は全体と同様の傾向であるが、学術論文の割合が減少し登録数が2桁のコンテンツが含まれなくなった。

以上の結果より、利用回数の多い順に、紀要、学位論文、学術論文であることが判明した。その3type種別のみで、機関リポジトリの利用の多いコンテンツの90%以上を占めていた。このことから、機関リポジトリの利用実態をより詳細に調査するためには、上位3type種別に関して、調査と分析を実施した。

3.2.3 紀要の利用実態

日本の機関リポジトリの特徴として、紀要の登録の多さが知られている。紀要には、査読付きの学術雑誌もあれば、灰色資料に位置づけられる大学における活動報告を記録するための雑誌も存在している。紀要の分類は課題とされつつも、現実的に詳細に分類を行うのが困難な資料とされている。

本研究の対象である筑波大学は、日本の機関リポジトリの中でも、特徴的な部分があり、それが紀要における学術論文の割合が大きい点である。だが、そのような特徴の筑波大学においても、機関リポジトリに登録されているコンテンツの40%を紀要が占めている。

また、日本の大学図書館の空間の狭隘化の解決策として、紀要は大学図書館によるセルフアーカイビングが進んでいる分野でもある。また、近年では紀要を電子版のみで発行する大学もあるので、機関リポジトリにおいて、紀要がどのように利用されているのかを明らかにすることは、大切である。紀要をサンプルにして、図書館によるセルフアーカイビングの資料の利用実態を明らかにすることができれば、資料の電子化やアーカイブの充実のための根拠にもなり得る。世界的に見て、日本語資料の電子化が進んでいないことは、Hathi Trust のデータ[205]でも明らかになっているので、紀要の利用状況を明らかにすることで、図書館による資料の電子化や今後の法改正に伴うアーカイブの充実に向けて参考となるデータを示すことが重要である。

本研究では、機関リポジトリに収録された紀要全17,950件を対象にして、利用実態を明らかにした。なお、紀要のタイトルについては、ハーベスティングしたメタデータ内での表記ゆれが数多く見られた。表記ゆれに関しては、タイトルの文字列一致によって名寄せを行ったが、筑波大学発行の紀要以外のデータも含まれており、完全なリストとの参照が困難であった。そのため、名寄せを行ったが後の紀要のタイトルについても、完全な同定には至っていないことに留意する必要がある。

紀要の名寄せを行い、集計を行った結果として、筑波大学の機関リポジトリに登録されている紀要は、176タイトルに絞ることができた。また、紀要の利用に関して、相関係数は0.319でコンテンツ数と利用数には弱い相関があることが判明した。つまり、コンテンツを充実させることは、コンテンツの利用につながる可能性がある。

紀要が機関リポジトリに登録されてから利用されるまでの平均の年数は、全紀要が対象の場合は、平均 4.0 年かかることが判明した。また、分析の期間を考慮して 2010 年以降に登録されたコンテンツで同様に平均を算出すると登録から 1.4 年かかることが判明した。なお、今回の分析において何月に登録されたかは考慮していないため、コンテンツが登録されてから実際に利用されるまでに平均して 1 年から 2 年かかるといえる。2010 年から 2014 年に登録されたコンテンツの同期間における初回利用と、2014 年から 2017 年に登録されたコンテンツの同期間における初回利用までの期間は、どちらも平均して 0.02 であった。つまり、登録されてかつ利用されるコンテンツに関しては、どちらも登録後 1 年以内に利用がなされていることになる。2010 年から 2014 年の期間と 2010 年から 2017 年の期間における大きな違いは、利用率である。前者の期間においては、利用率が 1.9% であったのが、後者の期間においては、96.0% まで上昇していた。

表 35 登録年と登録コンテンツ数(N=17,950)

登録年	登録コンテンツ数
2007	8,099
2008	1,285
2009	783
2010	618
2011	1,946
2012	733
2013	561
2014	651
2015	1,278
2016	499
2017	1,497

紀要の登録状況を明らかにするために年度別にコンテンツ登録数を明らかにした。

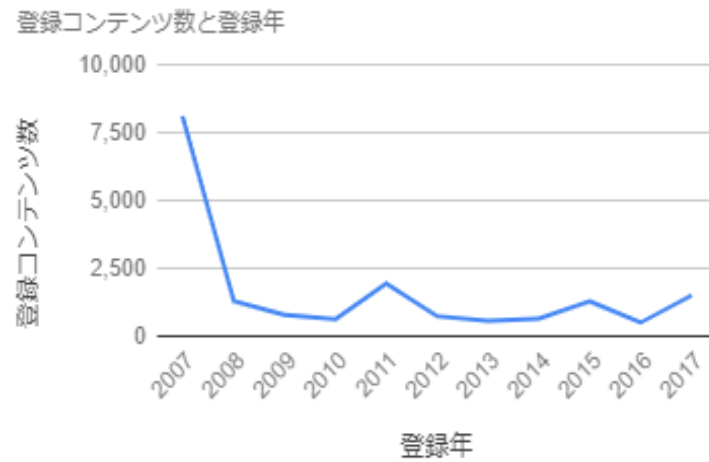


図 27 登録年と登録コンテンツ数

表. 35 の結果をグラフにして可視化した。2011 年度が増えているが 500 弱から 2000 までの間で推移している。

表 36 コンテンツ記述言語と登録コンテンツ数(N=17,950)

lang	登録コンテンツ数
Jpn	16,209
Eng	1,666
Fre	31
Ger	23
Rus	7
Zho	6
no-data	6
Ita	1
Spa	1

紀要は基本的に大学内の組織が発行する雑誌である。紀要がどのような言語で記述されているかは、大学のグローバル化を測ることにつながる可能性があると考えてコンテンツの記述言語を調査した。

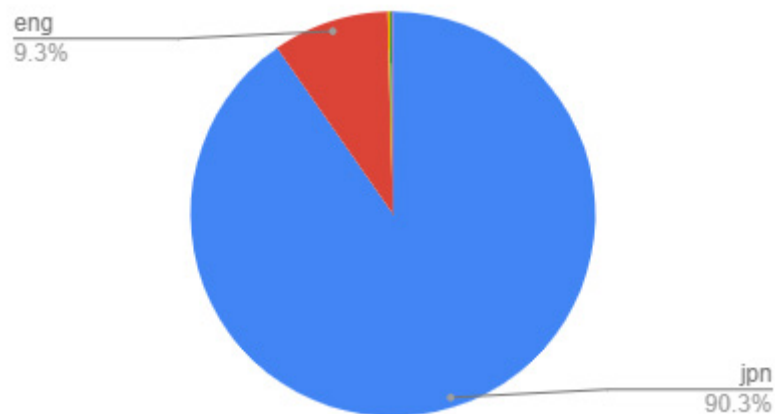


図 28 登録コンテンツの言語比率(N=17,950)

コンテンツの記述言語は、日本語が 90%以上を占めていた。英語の割合が 10%弱であり、その他のコンテンツは 1%以下であった。

表 37 紀要別の収録コンテンツ数ランキング

rank	journal title	収録コンテンツ数
1	Tsukuba Journal of Mathematics	1,005
2	筑波フォーラム	815
3	筑波大学心理学研究	773
4	文藝言語研究	773
5	体育科学系紀要	676
6	心身障害学研究	554
7	Tsukuba English Studies	493
8	比較文化研究	437
9	筑波大学技術報告	428
10	哲学・思想論集	422
11	大学体育研究	405
12	年次研究報告	358
13	人文科教育研究	346
14	文学研究論集	333
15	筑波大学留学生センター日本語教育論集	322

ハーベスティングしたメタデータを基に、紀要別の収録コンテンツ数を調査した。結果として、最もコンテンツ数が多いのは、『Tsukuba Journal of Mathematics』であった。数学

や心理学、言語学、体育科学など上位を占めているのは、研究科や系などが発行している雑誌であった。だが、筑波フォーラムなどのシンポジウムや筑波大学技術報告なども含まれていた。『Tsukuba Journal of Mathematics』は、学術雑誌としてメタデータが付与されているコンテンツも存在したため、上位の結果を見ても紀要を定義することの困難さを示す結果となった。また、一番登録数が多い『Tsukuba Journal of Mathematics』は学術雑誌でもある。そのことを踏まえると、紀要のコンテンツ数が多い分野は、STEM 分野よりも人文・社会科学分野であることが結果として明らかになった。

表 38 紀要別のコンテンツ利用数ランキング

rank	journal title	紀要別コンテンツ利用数
1	筑波大学心理学研究	307,239
2	体育科学系紀要	184,069
3	心身障害学研究	101,867
4	人文科教育研究	96,281
5	筑波大学発達臨床心理学研究	71,816
6	筑波大学体育系紀要	66,414
7	筑波大学留学生センター日本語教育論集	61,594
8	Tsukuba Journal of Mathematics	51,580
9	地域調査報告	45,756
10	大学体育研究	45,116
11	筑波大学人文地理学研究	44,277
12	筑波大学特別支援教育研究：実践と研究	43,568
13	筑波応用言語学研究	40,315
14	筑波大学技術報告	37,758
15	学校経営研究	33,792

紀要別のコンテンツ利用回数を調査し、結果の上位 15 件を表にまとめた。登録コンテンツ数が 3, 5, 6, 13 番目に多かった雑誌が、利用数の上位 4 件までとなった。登録コンテンツ数の段階で STEM 分野よりも人文・社会科学分野のコンテンツの量が多いことが判明した。利用の上位を調査すると、コンテンツ登録数の上位で、研究科等の発行する紀要以外の順位が下がり、研究科等の名称を持つ紀要が上位になった。また、登録数では一番多かった、『Tsukuba Journal of Mathematics』以外は、人文・社会科学分野や学際分野が上位を占める結果となった。

表 39 紀要利用数の経年変化

year	2010	2011	2012	2013	2014	2014	2015	2016	2017
total	2,217	3,063	3,406	6,408	766	123,935	452,405	709,461	1,272,897

紀要の利用状況を明らかにするため、紀要の利用数の経年変化を調査した。紀要の利用回数は、DSpace を使用していた年度と JAIRO Cloud 移行後の年度で明確な差が現れた。

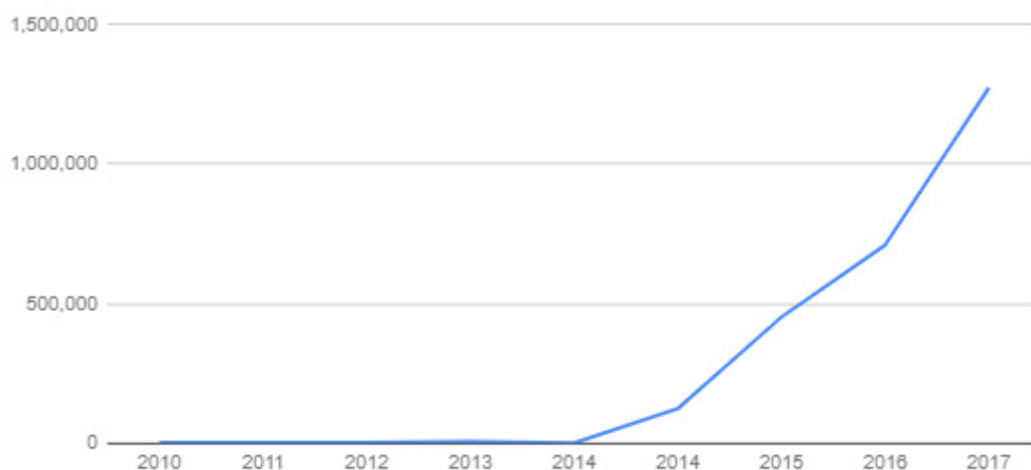


図 29 紀要利用数の経年変化

紀要の利用数の経年変化を可視化すると、2014 年までとそれ以降の利用数の変遷が明確になった。

表 40 コンテンツの登録年度と初回利用までの年数

コンテンツ登録年度	2010-2014	2014-2017	2010-2017	2007-2017
全数	4,915	3,926	7,783	17,950
利用のあったコンテンツ	95	3,769	7,595	17,685
登録後初回利用平均年度数	0.021	0.027	1.429	4.023

上記の DSpace と JAIRO Cloud の期間における違いを調査するために、コンテンツの登録年度を各システムの運用開始年度、アクセスログの存在する年度で区分して、利用のされ方を調査した。結果として、2014 年 3 月までは、限られたコンテンツへのアクセスのみが行われており、利用回数が少なかったことが判明した。2014 年 5 月以降は、アクセスされているコンテンツの種類が増加していた。

表 41 紀要コンテンツ利用上位 15 件

title	journal title	year	lang	NIIsubject	usage
<原著>ローゼンバーグ自尊感情尺度日本語版の検討	筑波大学発達臨床心理学研究	2007	jpn	心理学	34,263
<翻訳>ヤング グッドマン ブラウン	人文科教育研究	2011	jpn		9,711
環境倫理学の課題 ―世代間倫理をめぐって―	倫理学	2007	jpn		9,255
〈Research Reports〉 Mapping the human settlement of South East Asian cities using ALOS AVNIR-2	Tsukuba Geoenvironmental Sciences	2012	eng		8,302
内発的―外発的動機づけ測定尺度の開発	筑波大学心理学研究	2011	jpn		7,920
〈特別寄稿〉デンマークの教育制度：国民を幸福にする教育と教育制度の探究（試論）	教育制度研究紀要	2012	jpn		7,331
民事訴訟記録における個人情報の取扱に関する一考察	筑波法政	2011	jpn	法律・法律学	6,687
Killing vector fields and the holonomy algebra in semiriemannian manifolds	Tsukuba Journal of Mathematics	2008	jpn	数学	6,336
現代人における年中行事と見出される意味：恵方巻を事例として（〈特集2〉現代社会と民俗）	比較民俗研究	2010	jpn		6,198
海外における知能研究と CHC 理論	筑波大学心理学研究	2010	jpn		6,068
日本人の言語行動パターン：ウチ・ソト・ヨソ意識	筑波大学留学生センター 日本語教育論集	2007	jpn		5,626
青年期における情動的共感性の特質	筑波大学心理学研究	2011	jpn		5,560
〈プロジェクト報告〉長距離選手のランニングエコノミーに影響を及ぼす体力および技術的要因の検討	筑波大学体育系紀要	2013	jpn		5,493
〈調査報告〉大学アメリカンフットボールにおける傷害調査：10 年間（1999 年から 2008 年）の傷害報告	体育科学系紀要	2011	jpn		5,452
教育的タクトと授業の展開	教育方法学研究	2011	jpn		5,325

紀要の利用上位 15 件を調査した。STEM 分野のコンテンツが 1 件で残りが学際、人文・社会科学分野のコンテンツとなった。第 1 位『<原著>ローゼンバーグ自尊感情尺度日本語版の検討』の利用が他と比較して多い。また、上位には尺度に関する文献が含まれていた。

表 42 利用年度と登録年度のクロス集計(2010 年から 2017 年まで)²⁷

年度	2010	2011	2012	2013	2014	2015	2016	2017	登録数
								58,877	
2017	0	0	0	0	0	0	0	(5%)	1,497
							22,851	52,913	
2016	0	0	0	0	0	0	(3%)	(4%)	499
						31,614	71,174	104,044	
2015	0	0	0	0	0	(7%)	(10%)	(8%)	1,278
				72*	9,096	30,913	43,206	70,117	
2014	0	0	0	(1%)	(7%)	(7%)	(6%)	(6%)	651
				333	7,963	25,027	35,296	58,804	
2013	0	0	0	(5%)	(6%)	(6%)	(5%)	(5%)	561
			485	832	8,848	28,689	41,255	78,409	
2012	0	0	(14%)	(13%)	(7%)	(6%)	(6%)	(6%)	733
		316	311	1,109	14,911	51,308	82,632	148,775	
2011	0	(10%)	(9%)	(17%)	(12%)	(11%)	(12%)	(12%)	1,946
	1,996	2,623	2,440	3,768	9,650	29,605	37,481	66,788	
2010	(90%)	(86%)	(72%)	(59%)	(8%)	(7%)	(5%)	(5%)	618
	214	107	162	281	4,395	10,946	19,496	29,776	
2009	(10%)	(3%)	(5%)	(4%)	(3%)	(2%)	(3%)	(2%)	783
	7	17	8	13	6,927	18,964	37,394	70,257	
2008	(0%)	(1%)	(0%)	(0%)	(5%)	(4%)	(5%)	(6%)	1,285
	0	0	0	0	64,362	225,339	318,676	534,137	
2007	(0%)	(0%)	(0%)	(0%)	(51%)	(50%)	(45%)	(42%)	8,099
総計	2,217	3,063	3,406	6,408	126,152	452,405	709,461	1,272,897	
	(100%)	(100%)	(100%)	(100%)	(100%)	(100%)	(100%)	(100%)	17,950

紀要の利用実態を明らかにするために、登録年度と利用年度をクロス集計した。縦軸は、通時オブソレッセンスを横軸は共時オブソレッセンスを示している。2014 年までは、利用の減衰ではなく、利用の偏りが生じていたことが判明した。各年度での傾向を見ても紀要は、登録されてからの年数での利用の減衰は観測されず、登録されているコンテンツの数に影響を受けることが判明した。2010 年から 2013 年の期間に登録されたコンテンツでは、2010 年に登録されたコンテンツが多く利用されていた。2014 年から 2017 年に登録されたコンテンツでは、2007 年に登録されたコンテンツが多く利用されていた。

²⁷ *はメタデータの付与ミスであると考えられる。

3.2.5 紀要の利用実態に関する考察

紀要の利用回数を年度ごとに測定したが、利用の減衰は観測されず、コンテンツの数に依存して利用がなされていることが本研究では明らかになった。また、期間によって利用されていたコンテンツが大きく異なっていたことも判明した。特徴的なのは、2014 年以降に利用が多いコンテンツとして、コンテンツ数から判断すると 2010 年のコンテンツが予測されたが、実際に利用が多いのは、2011 年に登録されたコンテンツであった点である。このように、年度によって特徴的な利用傾向を示すことは紀要と後述する学位論文に見られる傾向であった。

また、タイトルに研究科などの名称がついており、分野がタイトルから判断できる雑誌は、大学の発行する雑誌よりもコンテンツ数が少ないにも関わらず利用が多い傾向があった。また、紀要単位で利用回数を測定した結果から、紀要は STEM 分野よりも人文・社会科学分野、学際分野の利用が多くなることが判明した。

今日の学術情報流通の重要な点として、国際的な学術雑誌で論文を出版することが、研究者評価につながっている。また、それにより国際誌の需要の固定化と価格の向上が起きていることは事実である。雑誌の価格向上のグラフ(図.4 図.5)を見た際に、自然科学分野では、人文・社会科学分野の方が比較的、価格が高騰が抑えられていた。

人文・社会科学分野の紀要が機関リポジトリで公開されていることは、自然科学分野の学術雑誌の OA と同様の効果をもたらすと考えられる。人文・社会科学分野の議論としては、エンバーゴ期間の設定に関する議論もなされている。自然科学分野においても同様の議論があるが、人文・社会科学分野は 35 年などの長期での設定も見られるなど、データに基づく適切な期間設定に関する知見が不足している現状がある。本研究における 2007 年の減衰の様な視点で調査を行うことが、適切なエンバーゴ期間設定の一助になるだろう。そのために、更なる知見の蓄積が求められる。

本研究の調査結果から、人文・社会科学分野の紀要に関して、分野の明確な紀要にコンテンツを登録し、機関リポジトリを通して OA にすることは、明確なコンテンツの利用につながる事が分かった。そのため、研究者へのインセンティブを提示する際に、玉石混濁ともいえる紀要というカテゴリではなく、研究分野の明確な名称の紀要の利用が多かったことを示すことができる。これは、紀要を重要な研究発表の場として見ていない自然科学分野の研究者にも示すことのできるメリットである。雑記などではなく、研究に結び付いたコンテンツが利用されている状況といえるためである。

2014 年までの紀要の利用状況を踏まえると、「大学においてセルフアーカイビングしたコンテンツをデポジットしておくだけなので、資料の保存や狭隘化の対策につながっても、コンテンツの著者に対するメリットがない。」という状況であったといえる。だが、2014 年以降の利用状況を踏まえると「ネットワーク化されたりポジトリに保存することで、コンテンツが利用可能になり、著者にもメリットが現れる。”象牙の塔”として図書館

がコンテンツを保存だけでなく、研究者とのコミュニケーションをとりつつ OA を推進する組織になるのに影響を与えることができる。」状況に変化したといえる。

3.2.6 学術論文の利用実態

BOAI で示された OA のポリシーを踏まえると Green OA において、機関リポジトリに求められているのは、出版された学術論文の収集である。また、収集した学術論文を利用可能な形で提供することが必要である。機関リポジトリにおける学術論文の収集の困難な点として、研究者へのメリットの提示が挙げられる[204]。プレプリントサーバにおけるコミュニティ機能などが無い。実際に OA 化することで引用が増えるという結果はあるが、OA 誌でなく大学図書館の機関リポジトリである必要性はない。グッドプラクティスとして、研究費の投資と機関リポジトリにおけるアーカイビングをセットにしている事例もあるが、日本においては一般的でない。

本研究における対象である筑波大学の機関リポジトリの特徴として、学術論文のコンテンツ割合が多いことが事実としてあるので、機関リポジトリにおける学術論文の収集と公開の推進のために利用実態を明らかにした。

結果として、対象の機関リポジトリは、80%以上が日本語によって記述されたコンテンツで構成されているのに対して、学術論文のコンテンツは、言語比率で日本語よりも英語の比率が高く 67%を占めていたことが判明した。これは、学術論文の投稿先として日本の学術雑誌が重要視されていない背景を反映していると考えられる。特に、コンテンツの割合が上位の分野に STEM 分野が多いことが要因として挙げられる。

紀要と学術論文の異なる特徴として、学術論文に関しては、75%弱(74.05%)の資料が利用されていた点が挙げられる。

表 43 登録年と登録コンテンツ数(N=6,993)

登録年	登録コンテンツ数
2007	372
2008	565
2009	579
2010	420
2011	798
2012	387
2013	291
2014	370
2015	991
2016	1,214
2017	1,006

コンテンツの登録年度とコンテンツ数を調査した結果を表.43 に示した。2015 年以降、登録された学術論文の数が大幅に増加したことが読み取れる。



図 30 登録コンテンツ数の推移

表. 43 の結果をグラフで可視化した結果である。学术论文の登録コンテンツ数は年度によってばらつきが大きいといえる。

表 44 コンテンツ記述言語と登録コンテンツ数(N=6,993)

lang	登録コンテンツ数
eng	4,693
jpn	2,293
fre	5
zho	2

学术论文は国際誌での論文の出版が研究者評価につながっている現状があることを踏まえて、機関リポジトリに登録されている学术论文の記述言語を明らかにした。

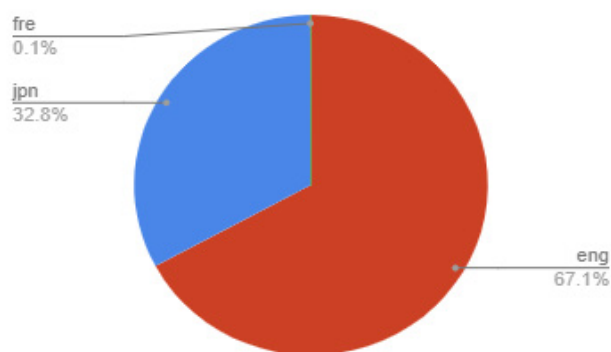


図 31 登録コンテンツの言語比率(N=6,993)

表. 44 をグラフで可視化した。他のコンテンツでは、日本語の比率が 80%~90%を占めていたが、学术论文に関しては比率が逆転し、英語の文献の登録数が多くなっていた。

表 45 学術論文の収録雑誌別コンテンツ数ランキング

rank	journal title	収録論文数	総利用回数
1	体力科学	396	13,822
2	Physical review B	311	30,186
3	デジタル図書館	276	7,850
4	Applied physics letters	184	24,113
5	情報処理学会論文誌	162	8,803
6	Physical Review Letters	159	13,113
7	Scientific Reports	155	6,418
8	Physical review D	154	14,485
9	PLOS ONE	134	11,425
10	年報筑波社会学	117	37,750
11	Journal of applied physics	94	10,057
12	体育学研究	91	10,582
13	Journal of High Energy Physics	83	1,874
14	月刊高校教育	72	12,330
15	Physics Letters B	61	2,321

ハーベスティングしたメタデータを基に、学術雑誌の収録雑誌別コンテンツ数を調査した結果の上位 15 件を表にまとめた。物理学分野の学術雑誌が 7 誌登場するなど STEM 分野のうち、特に物理学分野の学術論文のコンテンツが多いことが判明した。また、図書館情報学を含む情報学分野が 2 誌、体育科学分野が 3 誌存在した。また、OA ジャーナルとして、代表的な 2 誌である ²⁸PLOS ONE と Scientific Reports が見られた。

²⁸ 国立国会図書館の運営するカレントアウェアネスポータルにおいて、OA ジャーナルとして首位であった、PLOS ONE を Scientific Reports が逆転し、論文数世界一になった記事が紹介された[207]。PLOS ONE[208]とは、Public Library of Science の発行する OA ジャーナルであり、Scientific Reports[209]は、Nature が発行する OA ジャーナルであるため、商業出版と学術情報流通の依存という観点でも注目された。

表 46 雑誌論文の総利用回数ランキング

rank	journal title	収録論文数	総利用回数
1	年報筑波社会学	117	37,750
2	Physical Review B	311	30,186
3	スポーツ方法学研究	19	26,173
4	Applied Physics Letters	184	24,113
5	Journal of Physics. Condensed Matter	14	22,435
6	日本バーチャルリアリティ学会論文誌	61	16,224
7	Physical Review D	154	14,485
8	体力科学	396	13,822
9	Physical Review Letters	159	13,113
10	月刊高校教育	72	12,330
11	Medicine and Science in Sports and Exercise	2	12,091
12	PLOS ONE	134	11,425
13	体育学研究	91	10,582
14	Optics Express	44	10,086
15	Journal of Applied Physics	94	10,057

コンテンツ数の多い学術雑誌とは異なり、物理学分野から 6 雑誌が雑誌内の総論文利用回数の多い雑誌となった。また、1 番利用の多かった雑誌は社会学分野で、2011 年に刊行が終了していた。最新の学術雑誌論文だけでなく、アーカイブとしての機能が果たされていることが判明した。コンテンツ数の多い分野は STEM 分野というよりも物理分野に偏っていたが、利用という視点で見ると、STEM 分野として、医学や工学(情報学)の分野の学術雑誌が得られた。また、対象の機関リポジトリにおいても Scientific Reports が、PLOS ONE よりもコンテンツ数で上回ったが、利用数では、PLOS ONE の方が多い結果となった。(Scientific Reports 総利用回数: 6,418)

表 47 年度別コンテンツ利用回数

year	2010	2011	2012	2013	2014	2014	2015	2016	2017
利用総数	39,028	42,976	45,731	67,790	8,742	16,426	60,293	173,495	335,364

機関リポジトリにおける学術論文の利用状況を明らかにするために、年度ごとのコンテンツ利用回数を調査した結果を表に示した。紀要とは異なり、2014 年において利用回数の増加などは見られなかった。だが、2016 年以降に関しては、増加を続けている。

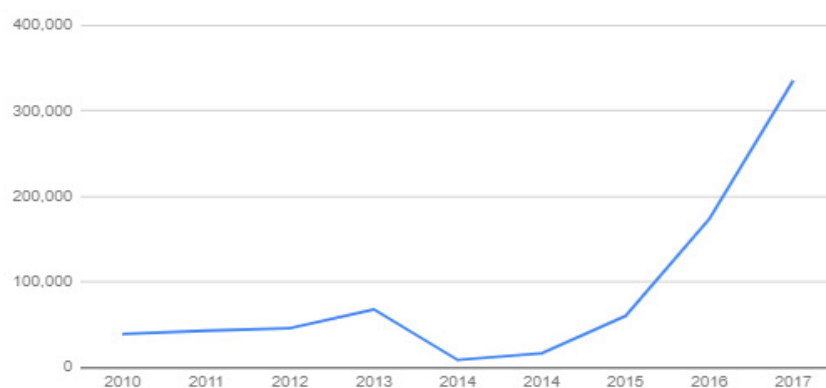


図 32 雑誌論文の収録コンテンツ数ランキング

表.47 をグラフで可視化した。2015 年までのコンテンツ利用回数から 2016 年度以降の利用回数の上昇幅の大きさが特徴である。

表 48 分野別のコンテンツ数

rank	NIIsubject	コンテンツ数
1	no-data	1,990
2	物理学	966
3	医学	840
4	基礎工学・応用物理学	735
5	情報学	413
6	生物学	391
7	数学	211
8	環境学	173
9	健康・スポーツ科学	145
10	社会科学	131
11	法律・法律学	122
12	化学	112
13	教育・教育学	96
14	電気電子工学	79
15	地理学・地誌	66
16	エネルギー工学	41

ハーベスティングしたメタデータを基に、分野別のコンテンツ数を調査した。学術雑誌のタイトルで判断した際に、物理学分野のコンテンツが多いことは判明していたが、上位

に STEM 分野が多いことが、この調査で明らかになった。なお、メタデータが付与されいていないコンテンツが無視できないほど多いため、結果には留意が必要である。

表 49 学術論文で利用の多いコンテンツ

rank	title	Journal Title	year	lang	NIIsubject	usage
1	Coherent gigahertz phonons in Ge ₂ Sb ₂ Te ₅ phase-change materials	Journal of Physics. Condensed Matter	2016	eng	物理学	21,535
2	Lack of Negative Correlation in Glucose Dynamics by Nonexercise Activity Thermogenesis Restriction in Healthy Adults	Medicine and Science in Sports and Exercise	2014	eng	健康・スポーツ科学	11,959
3	光トポグラフィによる歩行動作の評価	日本バーチャルリアリティ学会論文誌	2010	jpn	機械工学	11,869
4	水中に懸濁した微粒子の凝集分散・基礎理論とその適用性	塗装工学	2015	jpn	環境学	9,783
5	コントロールテストからみた跳躍競技の種目特性および競技パフォーマンスとの関係	スポーツ方法学研究	2009	jpn	健康・スポーツ科学	7,491
6	日本的経営と儒教思想：「忠」の哲学・「和」の理念	年報筑波社会学	2011	jpn	社会科学	5,806
7	Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}=13$ TeV	The European Physical Journal. C	2016	eng		4,484
8	《投稿論文》 ノーマライゼーションとはなにか	年報筑波社会学	2011	jpn	社会科学	3,789
9	朝の読書の評価に関するアンケート調査－意義と問題点－	日本生涯教育学会論集	2012	jpn	教育・教育学	3,776
10	学校教育制度におけるア	教育制度学研究	2011	jpn	教育・	3,686

	ティキュレーションの問題 -課題意識の変容と教育課題 -				教育学	
11	Development of RNA-FISH Assay for Detection of Oncogenic FGFR3-TACC3 Fusion Genes in FFPE Samples	PLOS ONE	2017	eng		3,603
12	Carrier injection dynamics in heterojunction solar cells with bipolar molecule	Applied Physics Letters	2015	eng	基礎工学・ 応用物理学	3,144
13	Epidemiological survey of acute low-tone sensorineural hearing loss	Acta Otolaryngologica	2017	eng	医学	3,139
14	「世界遺産白川郷」におけ る観光の現状と課題	日本造園学会誌	2016	jpn		3,038
15	土壌と界面電気現象：3. 界 面動電現象とその利用	日本土壌肥科学雑誌	2014	jpn	環境学	2,995

学術雑誌単位で見た際は、物理学分野の雑誌のコンテンツ数と雑誌別の利用回数が高い結果が得られた。コンテンツ単位で利用数の多い論文を調査した結果として、上位は STEM 分野と社会学・教育学などの論文があることが判明した。また、OA ジャーナルである PLOS に掲載された論文も利用回数の上位のコンテンツとして結果に表れた。

表 50 コンテンツの登録年度と初回利用までの年数

コンテンツ 登録年度	2010-2014	2014-2017	2010-2017	2007-2017
全数	2,266	3,582	5,477	6,993
利用のあったコンテンツ	1,678	3,359	5,398	6,756
登録後初回利用平均年度数	0.034	0.202	0.319	0.666

紀要では、2010 年 3 月から 2014 年 3 月までに登録されたコンテンツが 2014 年 3 月までに利用されている割合は、0.2%であったが、学術論文に関しては、75%弱のコンテンツが利用されていた。学術論文は、紀要と比較し機関リポジトリに登録された場合、利用されるまでの期間はあまり変化が見られないが、利用される割合が高いことが明らかになった。

表 51 利用年度と登録年度のクロス集計(2010 年から 2017 年まで)²⁹

year	2010	2011	2012	2013	2014	2015	2016	2017	利用総数	登録数
2017								36,905 (11%)	36,905	1,006
2016							15,102 (9%)	69,966 (21%)	85,068	1,214
2015						5,060 (8%)	24,859 (14%)	45,465 (14%)	75,384	991
2014				260* (0.39%)	2,826 (11%)	7,871 (13%)	15,215 (9%)	28,679 (9%)	54,851	370
2013	4* (0%)	9* (0%)	5* (0%)	3,032 (4%)	1,984 (8%)	3,662 (6%)	10,664 (6%)	10,826 (3%)	30,186	291
2012			4,047 (9%)	7,875 (12%)	2,798 (11%)	5,737 (10%)	14,851 (9%)	16,911 (5%)	52,219	387
2011		10,774 (25%)	8,169 (18%)	14,313 (21%)	5,388 (21%)	12,866 (21%)	26,672 (15%)	39,452 (12%)	117,634	798
2010	5,245 (13%)	6,524 (15%)	6,708 (15%)	20,505 (30%)	3,623 (14%)	7,061 (12%)	14,615 (8%)	22,824 (7%)	87,105	420
2009	15,028 (39%)	9,268 (22%)	10,451 (23%)	10,602 (16%)	4,415 (18%)	10,764 (18%)	28,112 (16%)	31,629 (9%)	120,269	579
2008	11,326 (29%)	9,866 (23%)	9,823 (21%)	6,709 (10%)	2,864 (11%)	4,561 (8%)	12,789 (7%)	22,328 (7%)	80,266	565
2007	7,425 (19%)	6,535 (15%)	6,528 (14%)	4,494 (7%)	1,270 (5%)	2,711 (4%)	10,616 (6%)	10,379 (3%)	49,958	372
総数	39,024 (100%)	42,967 (100%)	45,726 (100%)	67,530 (100%)	25,168 (100%)	60,293 (100%)	173,495 (100%)	335,364 (100%)	789,845	6,993

学術論文の利用実態を明らかにするために、登録年度と利用年度をクロス集計した。縦軸は共時オブソレッセンスを、横軸は通時オブソレッセンスを示している。学術論文は、被引用半減期を計算することにより公開されてから引用数が減衰するまでの年数を求めることも可能である。対象の機関リポジトリにおいては、同様の傾向が現れると予測したが、結果としては、明確な利用の減衰は得られなかった。

²⁹ *はメタデータの付与ミスであると考えられる。

3.2.7 学術論文の利用実態に関する考察

コンテンツの充実度には、大学の特色が現れているといえる。上位にある体育学、物理学、図書館情報学は、筑波大学の特色として知られている分野である。学術雑誌別の利用回数のランキングにも体育学や物理学が反映されている。

また、筑波大学の機関リポジトリにおける雑誌単位の利用が一番多いコンテンツが『年報筑波社会学』であった。人文・社会科学分野の論文は投稿先や公開場所が限られており、STEM 分野と比較し、論文として公開が困難な面を持っていた。そういった中で、利用数が一番多かったのが、『年報筑波社会学』であった。雑誌自体は 2005 年で刊行を終了しており、後続誌も 2011 年までであった。同一タイトルの雑誌に掲載されている論文の登録年度がすべて 2011 年であることから、セルフアーカイビングとして残された雑誌であると考えられる。

紀要との利用実態の比較として、2010 年から 2014 年までの利用実態と 2014 年から 2017 年の利用実態での差異が異なる形で現れた。言い換えれば、紀要は DSpace を利用していた期間、デポジトリとして保存されていたが、JAIRO Cloud に移行してからコンテンツの利用が増加し、リポジトリの役割を果たすことになった。だが、学術論文の利用に大きな変化が現れたのは 2016 年であった。2016 年は、アクセス回数やダウンロード数といった観点でも急激な増加をしているため、利用の動向に影響を強く受けたととらえることができる。つまり、学術論文はネットワーク化された影響をあまり受けにくい程度に利用がなされていたといえる。

また、結果として Gold OA の学術雑誌である PLOS ONE に掲載されたコンテンツの利用が多かったことは、Gold 論文をリポジトリでも公開することのメリットを示すことにつながると考えられる。論文を OA 化することで引用数が増加することが示唆されていた。本研究では、OA 誌で利用されることが予測された論文が、機関リポジトリにおいても利用されていたことを示すことになった。これは、利用元の分散ととらえることもできる。利用するプラットフォームやアクセスする元の信頼性などを選択した結果のアクセスであると考えることができる。先行研究では学術雑誌のエンバーゴ期間のみで調査をしていた [150] ので、掲載元が OA ジャーナルにおいても利用数が向上するのかを今後精査し、明らかにする必要がある。

3.2.8 学位論文の利用実態

2013 年 4 月からそれまでは、国立国会図書館に紙媒体による納本が義務付けられていた博士論文のインターネットにおける公開が義務付けられた。公開は、基本的に大学の機関リポジトリで行われている。ただし、実際には要旨のみしか公開されないものもある。

博士論文の公開義務化や、CiNii Dissertations のサービス運用開始によって、2013 年以降、博士論文の利用環境は整備されてきている[210]。だが、機関リポジトリで公開されることとなった博士論文の利用実態は明らかになっていない。そこで、本研究では機関リポジトリに登録された博士論文の利用実態を明らかにするために、登録数の推移、記述言語、利用回数、利用の多い学問分野、利用の多い学位、利用の多かった上位 15 件の論文、利用の多かった上位 100 件学位論文のオルトメトリクス、オルトメトリクス 2 以上の学位論文を調査した。ここで、博士論文ではなく学位論文を対象としているのは、<Thesis or Dissertation>というメタデータでは、博士論文と修士論文の区別がつかないためである。

本研究において、オルトメトリクスに着目した理由に関しては、学位論文は学術論文と異なり、引用文献に登場することはあまりない。学位論文の構成論文が引用されても、学位論文の利用ではないので、実際に利用されたかが不透明なコンテンツであるためだ。

結果として、博士論文の公開義務化から 2 年を経た 2015 年に利用が急激に増加したことが判明した。

登録コンテンツ数が多い分野としては、STEM 分野が上位を占める中、言語学や教育学が含まれている。利用数が多い学問分野で見るとコンテンツ数の多い学問分野と大きく異なる結果となった。学術、文学、芸術学、などの人文社会科学系の学位論文が上位に表れた。

学位論文の利用回数上位 100 件のうちの 26 件とおよそ 1/4 がオルトメトリクスのスコアを持っていることに着目し、オルトメトリクスとのランクと利用数のランクの比較を行った。本研究では、オルトメトリクスによる分類を採用した。ceek.jp において、“筑波大学 AND 学位論文”の検索クエリから得られた結果は、654 件であった。本研究は、その中でも合計値でソートして、利用数 2 以上の 77 件を対象とした。利用数が 2 以上のコンテンツに限定した理由は、コンテンツ利用回数上位 100 件のコンテンツをサンプリングして確認した結果、オルトメトリクスが 1 のコンテンツの多くが、附属図書館のツイッターによる情報発信であったためである。また、オルトメトリクスが 2 以上のコンテンツに関して、学位論文の利用数との相関、利用数のランキングに関する相関を求めたがほとんど相関がなかった。だが、利用上位のコンテンツや登録数上位のコンテンツにおける STEM 分野と人文・社会科学系の分野の出現割合が逆転した。

表 52 登録年度とコンテンツ数(N=13,131)

year	登録コンテンツ数
2007	6,224
2008	10
2009	25
2010	1,012
2011	1,078
2012	381
2013	38
2014	1,037
2015	1,253
2016	1,039
2017	1,034

機関リポジトリにおける博士論文の効果義務化の前後において、機関リポジトリで公開される学位論文数に差があるかを調べた。2013 年 4 月 1 日以降に提出された学位論文、つまり 2014 年度からは、毎年 1,000 件以上の学位論文が登録されていることがわかった。

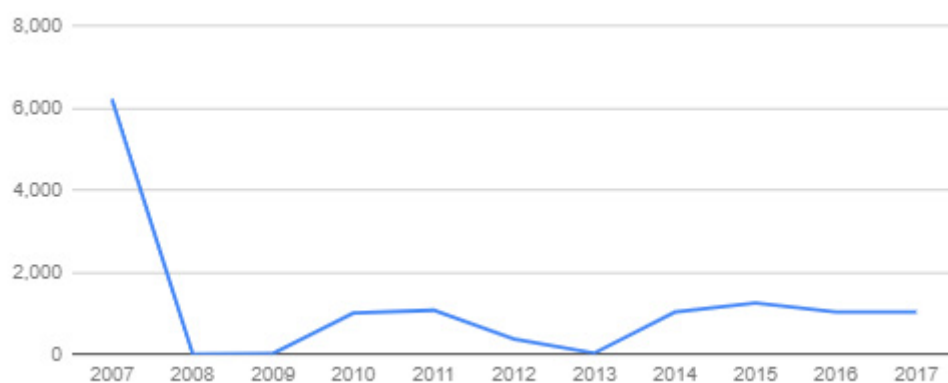


図 33 登録コンテンツ数の推移

表.52 をグラフで可視化した。2014 年度以前は波があったが、それ以降は毎年同程度の数の学位論文が公開されていることがわかる。

表 53 コンテンツ記述言語と登録コンテンツ数(N=13,131)

Lang	登録コンテンツ数
jpn	12,376
eng	753
ger	1
no-data	1

筑波大学における学位論文が記述されている言語を明らかにした。

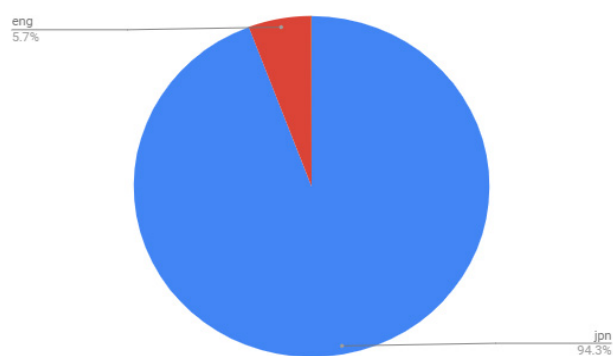


図 34 登録コンテンツ数の言語比率

学位論文の言語比率は、日本語書かれている割合が 95%弱と非常に大きい。

表 54 学位論文の利用回数の経年変化

year	2010	2011	2012	2013	2014	2014	2015	2016	2017
総利用回数	95,541	59,602	50,645	63,471	8,093	45,211	176,588	344,605	647,743

学位論文の利用回数も、学術論文と同様に毎年同程度利用されていた。学位論文に関しては、2015 年度から利用数が顕著に増加したと考えられる。

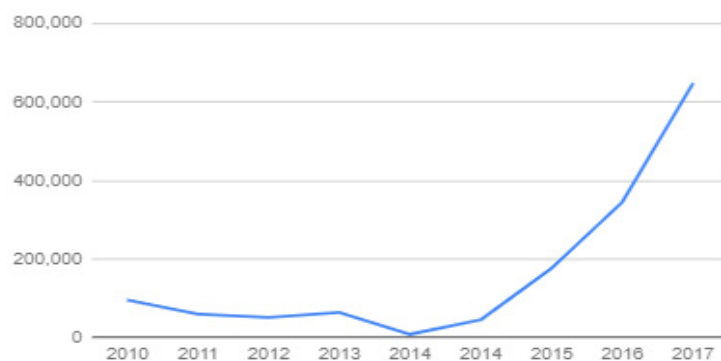


図 35 学位論文の利用回数の経年変化

グラフにて学位論文の利用回数の経年変化を可視化した。

表 55 NII Subject の数と登録学位数のランキング比較 ³⁰

rank	NIIsubject	登録コンテンツ数	学位	登録コンテンツ数
1	no-data	7,420	工学	1,822
2	医学	753	医学	1,458
3	生物学	675	no-data	1,301
4	物理学	482	Science	1,241
5	基礎医学	429	農学	887
6	化学	324	MedicalSciences	614
7	教育・教育学	282	文学	539
8	電気電子工学	233	理学	468
9	数学	232	Engineering	436
10	農学	221	学術	367
11	地理学・地誌	195	教育学	358
12	情報学	172	体育科学	322
13	健康・スポーツ科学	170	言語学	225
14	言語学	122	心理学	224
15	経済学	89	Agriculture	197

ハーベスティングしたメタデータを基に、登録コンテンツ数を分野と学位別に上位 15 件ずつ表にまとめた。学位に関しては、junii2 において<degree name>が義務付けられているが、学位に関するデータの多くは<description>にのみ記載されていた。oai_dc にお

³⁰ 学位の名寄せ(学位授与年度によって学位名が異なるので表記揺れのみを修正)

る<degreename>のダムダウン先が<description>であるためであると考えられる。³¹本研究では、<degreename>と<description>からデータを抜き出し統合した。メタデータを統合することによって、工学に関する学位論文が最もコンテンツ数が多いことが判明した。なお、言語による表記ゆれは統合していないため、農学と Agriculture などが示された。

表 56 学問分野別利用回数ランキング(上位 15 件)

rank	NIIsubject	コンテンツ数	利用回数合計
1	no-data	7,436	962,700
2	医学	753	53,798
3	生物学	677	45,577
4	教育・教育学	282	43,758
5	健康・スポーツ科学	170	40,098
6	基礎医学	429	32,776
7	農学	222	24,864
8	物理学	482	24,624
9	言語学	122	22,430
10	化学	324	18,094
11	電気電子工学	233	17,701
12	心理学	85	17,196
13	数学	232	16,468
14	情報学	172	15,426
15	社会・文化	81	14,538

登録コンテンツ数と合わせて、利用回数の多い学問分野の調査も実施した。結果として、教育学、健康・スポーツ科学、言語学、社会・文化、心理学などの学際、人文・社会科学分野が登録コンテンツ数の順位よりも高い順位となっていることがわかった。

³¹ 学位規則改定後のメタデータフォーマットが原因であると考えられる[211]。

表 57 学位別利用回数ランキング(上位 15 件)

Rank	学位	コンテンツ数	利用回数合計
1	工学	1,822	228,367
2	医学	1,458	128,576
3	no-data	1,301	105,358
4	体育科学	322	104,918
5	農学	887	101,875
6	Science	1,241	71,665
7	学術	367	66,140
8	文学	539	64,142
9	心理学	224	48,271
10	言語学	225	45,007
11	教育学	358	41,695
12	芸術学	180	37,293
13	MedicalSciences	614	34,200
14	スポーツ医学	128	31,818
15	ヒューマン・ケア科学	138	25,990

学位別で利用回数を調査した結果、上位の 3 件は学位論文の登録数と同じ並びとなった。芸術学やスポーツ医学などの学位論文は、論文数では上位 15 件に入っていないが、利用数では上位となった。体育科学の学術論文は、登録コンテンツ数はあまり多くないが、利用回数が上位となった。

表 58 学位論文の利用上位 15 件

rank	title	year	学位	利用	Ceek.jp
1	バランス能力の因子構造に及ぼす加齢の影響	2011	体育科学	16,719	0
2	直腸がん前方切除術後の排便障害を軽減する看護支援に関する研究	2015	看護科学	11,279	0
3	イメージ操作は脳の中でどのように行われるのか：心的回転の回転方向に関する実験的検討	2014		9,991	1
4	体育授業における教師の効果的なフィードバック行動に関する検討	2010	体育科学	8,090	0
5	医療機関における聴覚障害者の手話通訳支援に関する研究	2015	学術	7,106	6
6	蒸気爆発現象のトリガリング過程における蒸気膜崩壊現象に関する研究	2007	工学	6,784	0
7	地域密着型サッカークラブマネジメントにおける成長概念モデル・プロ事業化へ向けた意思決定モデルの構築と分析	2014	マネジメント	5,550	0
8	自己開示における抵抗感の規定因に関する研究	2014	心理学	5,510	1
9	A Study of Influential Factors Surrounding Writing Performances of Japanese EFL Learners: From Refinements of Evaluation Environment Toward Practical Instruction	2015	言語学	5,462	0
10	デカルト哲学における「身体教育」への指標：心身関係論を機軸として	2015	体育科学	5,415	1
11	子どもの読書の意義と効果に関する研究	2014		5,306	1
12	ナノ結晶構造を持つ軟質および硬質磁性材料における保磁力機構の解析	2014	工学	5,234	0
13	Higher education and politics in Malaysia	2011	International Political Economy	5,101	1
14	ベイズ理論に基づく移動ロボットの自己位置推定と地図生成に関する研究	2015	工学	5,087	0
15	TiO ₂ Photocatalytic Degradation of Waste Activated Sludge and Potassium Hydrogen Phthalate in Wastewater for Enhancing Biogas Production	2014	環境学	5,038	0

利用の多い学位論文の上位 15 件を表にまとめた。学位論文の利用が多くなる原因として、SNS などで話題になった可能性を考えて ceek.jp よりオルトメトリクスを取得した。学位名が記載されていないものを調査した結果、2 件は修士論文であった。

表 59 分野別のオルトメトリクス合計³²

rank	学位	コンテンツ数	altmetrics
1	図書館情報学	16(15)	130(49)
2	教育学	5	32
3	学術	6	22
4	文学	8	20
5	工学	8	18
6	体育科学	4	15
7	社会学	1	10
7	医学	5	10
7	情報学	5	10
10	国際政治経済学	2	8
11	言語学	3	6
12	コーチング学	2	5
12	政治学	1	4
12	農学	2	4
12	心理学	2	4
12	システムズ・マネジメント	2	4
17	心身障害学	1	3
18	芸術学	1	2
18	デザイン学	1	2
18	no-data	1	2
18	経営学	1	2

学位論文は学術論文とは異なり、引用数を指標とした分析が困難である。そのため代替の指標として、利用数とオルトメトリクスに着目した。オルトメトリクスに関する調査の結果を表にした。対象となる論文は、オルトメトリクスが2以上の論文であった。オルトメトリクスのスコアが高い分野は、人文・社会科学分野が多い結果となった。

³² 0内の数値は上位1件を外れ値とした場合の数値

表 60 コンテンツの登録年度と初回利用までの年数

title	利用回数	利用回数 rank	学位論文内部 利用 rank	登録年
サッカーのゲームパフォーマンス尺度と因果構造	1943	269	82	2007
臨床看護師の倫理的感受性尺度の開発	568	1458	372	2017
インターネット利用が社会性および攻撃性に及ぼす影響：インターネット行動尺度による分析を通して	258	3606	843	2010
中規模病院における看護管理者のコンピテンシー評価尺度の開発	192	5071	1185	2014
発達障害児におけるコミュニケーション構造の発達と支援方法についての検討：評価尺度と支援プログラムの作成を通して	168	5939	1384	2007
心像の鮮明性尺度の作成に関する研究	134	7784	1859	2007
看護学生における実習での学び尺度の作成と学びに対する教育的支援の検討	111	9608	2373	2011
街路景観の構成要因と心理評価構造：多次元尺度構成法と SD 法による定量的比較分析	104	10323	2613	2007
刑事責任能力の客観的評価：筑波式刑事責任能力予測評価尺度開発の試み	28	29345	9563	2007

コンテンツの利用回数の上位には、心理統計などの尺度を含むことが多いことが経験的に知られているため、本研究でも、学位論文に限定して、“尺度”の文字列を抽出して利用回数の位置づけを調査した。結果として、尺度の開発に関する学位論文の利用は、あまり多くないことが判明した。

表 61 コンテンツの登録年度と初回利用までの年数

コンテンツ 登録年度	2010-2014	2014-2017	2010-2017	2007-2017
全数	3,546	4,363	6,872	6,993
利用のあったコンテンツ	1,766	4,311	6,814	6,756
登録後初回利用平均年度数	0.541	0.254	0.799	0.666

紀要、学術論文では、2010 年から 2014 年の間に登録された資料の利用に明確な違いが出たため、同様の調査を学位論文にも実施した。結果として学位論文は、紀要と学術論文のほぼ中間の値である約 50%弱が利用されていたことが判明した。

表 62 利用年度と登録年度のクロス集計(2010 年から 2017 年まで)³³

year	2010	2011	2012	2013	2014	2015	2016	2017	総利用数	登録数
2017								35,656 (6%)	35,656	1,034
2016							23,753 (7%)	66,747 (10%)	90,500	1,039
2015						27,377 (16%)	106,988 (31%)	180,504 (28%)	314,869	1,253
2014				214* (0%)	4,663 (9%)	32,082 (18%)	47,799 (14%)	81,767 (13%)	166,525	1,037
2013				1,084 (2%)	488 (1%)	2,055 (1%)	2,804 (1%)	3,900 (1%)	10,331	38
2012			181 (0%)	467 (1%)	1,075 (2%)	4,145 (2%)	10,036 (3%)	17,198 (3%)	33,102	381
2011	3,556* (4%)	6,653 (11%)	5,767 (11%)	5,632 (9%)	5,229 (10%)	12,897 (7%)	21,359 (6%)	44,170 (7%)	105,263	1,078
2010	4,791 (5%)	7,960 (13%)	5,654 (11%)	8,377 (13%)	10,321 (19%)	32,099 (18%)	44,376 (13%)	75,388 (12%)	188,966	1,012
2009	448 (0%)	149 (0%)	161 (0%)	190 (0%)	72 (0%)	210 (0%)	477 (0%)	702 (0%)	2,409	25
2008	69 (0%)	67 (0%)	64 (0%)	67 (0%)	58 (0%)	77 (0%)	150 (0%)	270 (0%)	822	10
2007	86,677 (91%)	44,773 (75%)	38,818 (77%)	47,440 (75%)	31,398 (59%)	65,646 (37%)	86,863 (25%)	141,441 (22%)	543,056	6,224
総数	95,541 (100%)	59,602 (100%)	50,645 (100%)	63,471 (100%)	53,304 (100%)	176,588 (100%)	344,605 (100%)	647,743 (100%)	1,491,499	13,131

学術論文の利用実態を明らかにするために、登録年度と利用年度をクロス集計した。縦軸は共時オブソレッセンスを、横軸は通時オブソレッセンスを示している。学位論文について調査をした結果として明確な利用の減衰は得られなかった。

³³ *はメタデータの付与ミスであると考えられる。

3.2.9 学位論文の利用実態に関する考察

学位論文は、学術論文を要件にしているため、学術論文と同様の利用傾向を示すことを予測していた。だが、実際の利用状況を見ると、学術論文と紀要の中間ともいえる結果を示した。具体的には、大学の特徴的な分野におけるコンテンツの利用回数の増加や、コンテンツが登録されてから利用されるまでにかかる時間などがこの特徴に当てはまる。

学位論文の公開義務化で利用動向などに変化が生じると考えたため、博士論文に対する分析を実施しようと考えていたが、実際の利用回数の上位のコンテンツに、情報学と図書館情報学の修士論文が含まれる結果となった。

博士論文に関する分析を実施するためには、修士論文と博士論文を区別し、要旨と本文を区別できるようなメタデータの記述が求められる。

3.2.10 利用のなかったコンテンツの調査

本研究の対象期間は、2010年3月から2017年12月までである。この約7年間に利用が観測されなかったコンテンツは、1,193件であった。機関リポジトリの導入の初期には、リポジトリとデポジトリの違いに言及し、デポジトリで終わらずにリポジトリとして利用がなされることが目標として掲げられていた。現在、OAI-PMHによって得ることの可能な43,476件のコンテンツのうちのおよそ3%が利用されていない状況である。機関リポジトリに登録されたコンテンツの97%以上は、最低でも一度は利用されたこととなる。また、2010年以前に登録されたコンテンツは、利用されなかったうちの254件であった。対象に期間中に利用のなかったコンテンツの詳細を以下に示す。

本研究の対象期間中に利用回数がゼロであったコンテンツの実態を調査した。機関リポジトリに登録されているコンテンツ全体と利用回数がゼロのコンテンツを比較した結果、2007年と2017年に登録されたコンテンツを合わせると利用回数ゼロの80%となる。利用回数がゼロのコンテンツの言語比率は、以下の図.38でも示したが、日本語:英語の比率が約4:1となる。コンテンツ全体におけるコンテンツの日本語:英語の比率は約4.3:1であることから、利用の有無に登録言語はあまり関係がないと言える。また、ここにおいて特徴的な結果となったのが、Research Paperの比率である。利用回数がゼロのコンテンツはType別に内訳を確認すると他のコンテンツが全体に対して、1%程度に留まっている。そのような結果にも、関わらず、Research Paperに関しては、Research Paper全体のうちの18%を利用回数ゼロのコンテンツが示していた。原因を明らかにするため、Research Paperについて詳細な分析を行った。

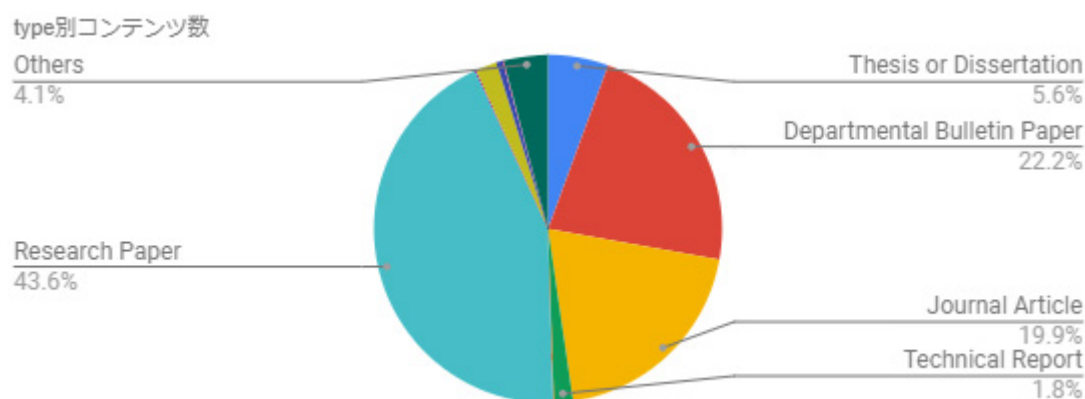


図 36 利用回数ゼロのコンテンツの type 別割合

機関リポジトリは、リポジトリのためコンテンツの利用を前提としている。そこで、利用のなかったコンテンツの属性を明らかにした。結果として、コンテンツ数は多くない Research Paper の割合が大きいことが判明した。

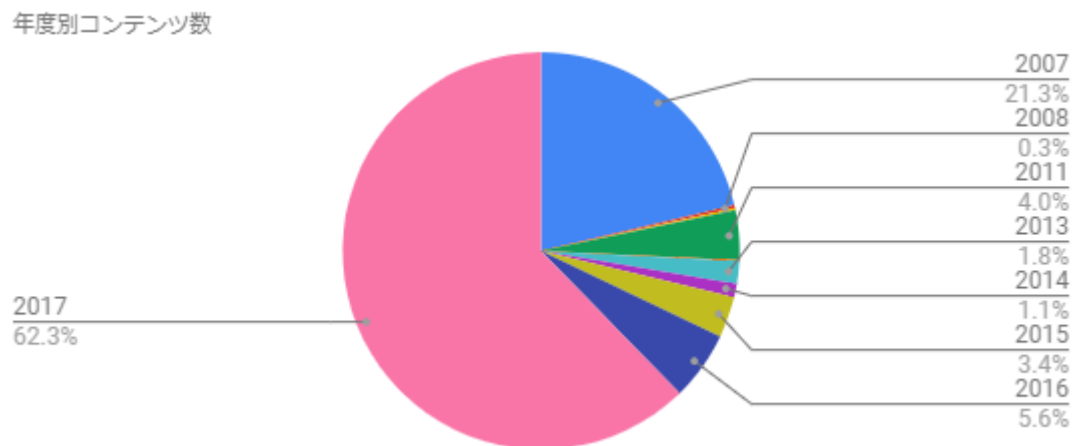


図 37 利用回数ゼロのコンテンツの登録年度

結果としては、2017 年度に登録されたコンテンツの利用が 62.3%でなされていなかった。本研究では、単位を年度にしているため、登録されてから間もないコンテンツが含まれていたことが原因であると考えられる。

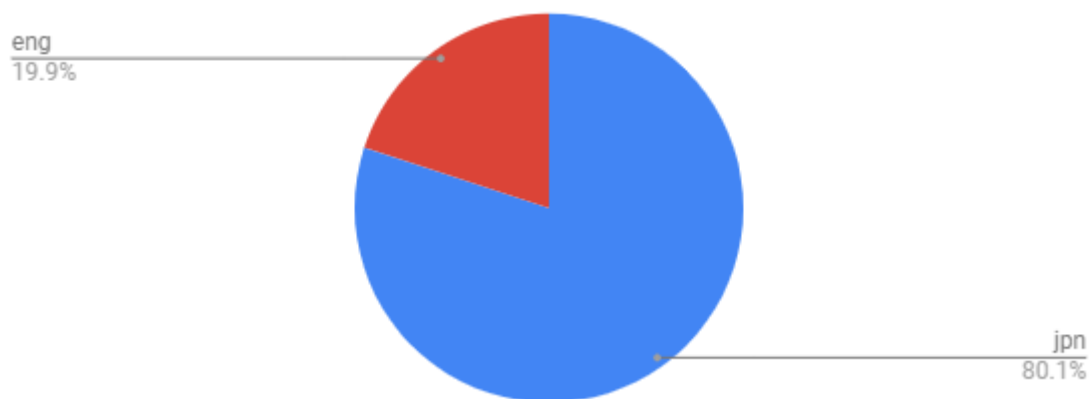


図 38 利用回数ゼロのコンテンツの言語比率

利用がなかったコンテンツの言語比率は、コンテンツ全体の比率と変わらない結果となった。

表 63 利用回数ゼロのコンテンツの実態(N=1,193)

type	利用回数ゼロ コンテンツ数	コンテンツ 全件数	全件に占める 利用ゼロの割合	利用回数ゼロ内 コンテンツ割合	全件 コンテンツ割合
Research Paper	520	2,818	18.45%	43.59%	6.09%
Departmental BulletinPaper	265	17,950	1.48%	22.21%	40.60%
Others	49	954	5.14%	4.11%	2.41%
Journal Article	237	6,993	3.39%	19.87%	16.69%
Thesis or Dissertation	67	13,131	0.51%	5.62%	30.21%
Presentation	1	137	0.73%	0.08%	0.31%
Conference Paper	7	365	1.92%	0.59%	0.86%
Preprint	0	53	0.00%	0.00%	0.11%
Learning Material	1	96	1.04%	0.08%	0.26%
Book	1	103	0.97%	0.08%	0.26%
Article	24	422	5.69%	2.01%	1.18%
Technical Report	21	475	4.42%	1.76%	1.01%

利用のなかったコンテンツの中でも、**Research Paper** に関する分析を行った。**Research Paper** に着目した理由は、アクセスログから判断することのできる利用回数がゼロのコンテンツが多いためであった。本研究の対象となる期間とそれ以前に区分して分析を行った。結果として、2014 年 5 月からコンテンツへのアクセスがなされたコンテンツが増加したことが判明した。この期間は、機関リポジトリのシステムが **DSpace** から **JAIR O Cloud** に移行した期間である。そのため、**JAIR O Cloud** にシステムが移行したことにより、それまでアクセスが少なかった登録から時間の経過したコンテンツへのアクセスが増加した可能性が高い。

表 64 Research Paper の年度別公開数(N=2,818)

公開年	公開件数
2007	370
2008	5
2009	16
2010	254
2011	181
2012	258
2013	307
2014	338
2015	349
2016	307
2017	433

Research Paper の登録年度とコンテンツ数を調査した。

表 65 Research Paper の年度別公開数(公開日 2010 年以前)

year	2010	2011	2012	2013	2014	2014	2015	2016	2017
	2010ssl	2011ssl	2012ssl	2013ssl	2014ssl				
利用可能件数	391	391	391	391	391	391	391	391	391
利用 コンテンツ数	30	18	29	31	15				
利用 コンテンツ数	32	32	31	21	10	170	191	219	217
利用コンテンツ (重複なし)	32	32	32	31	20	170	191	219	217
捕捉率	8.18%	8.18%	8.18%	7.93%	5.12%	43.48%	48.85%	56.01%	55.50%

2010 年までに登録されたコンテンツの利用が少なかったことが、利用のなかったコンテンツの割合を増加させた原因と考えて、年度ごとの利用を測定した。結果として、DSpace の利用期間に関しては、1 割弱の同一コンテンツのみが利用されており、残りのコンテンツの利用がないことが判明した。JAIRO Cloud に移行してからは、コンテンツの利用が増えたが、紀要のように 90%以上のコンテンツが利用されることはなかった。

表 66 Research Paper の年度別公開数(公開日 2010 年以降)

year	2010	2011	2012	2013	2014	2014	2015	2016	2017
	2010ssl	2011ssl	2012ssl	2013ssl	2014ssl				
全件数	645	826	1,084	1,391	1,729	1,729	2,078	2,385	2,818
利用 コンテンツ数	241	256	670	927	400				
利用 コンテンツ数	251	406	573	269	239	1,090	1,146	1,982	2,075
利用コンテンツ (重複なし)	253	415	683	950	550	1,090	1,146	1,982	2,075
捕捉率	39.22%	50.24%	63.01%	68.30%	31.81%	63.04%	55.15%	83.10%	73.63%

2010 年以降に登録されたコンテンツについては、80%近い利用があった。Research Paper は、最新のものは利用されるが、古いものについては、利用されない傾向があることが判明した。

3.2.11 コンテンツ利用に関する仮説の検証

本研究における仮説は、2つ設定した。1つめは、機関リポジトリに登録されたコンテンツにも、論文の被引用半減期のように、登録されてからある程度の時間が経過することで利用がされなくなる。2つめは、機関リポジトリにコンテンツが登録されてから利用されるまでには、1年程度のタイムラグが発生するという仮説である。上記の2つの仮説についてコンテンツ利用の割合の大きかった3つの分野(紀要、学術論文、学位論文)を対象として検証を行った。

結果 1

仮説：機関リポジトリに登録されたコンテンツは、時間が経過するにつれて利用数が減少する。

紀要、学術論文、学位論文に属するコンテンツを対象にオブソレッセンスを求めた。結果として、共時オブソレッセンスは、コンテンツ数に大きな差があり、コンテンツの量の影響が多いと考えられる。だが、どのコンテンツも登録した年度による影響は観測されなかった。

結果 2

仮説：機関リポジトリにコンテンツが登録されてから利用されるまでには、1年程度のタイムラグが発生する。

反応平均時間の算出による検証

反応平均時間とは、引用のタイムラグを測定する指標である。Schubert ら(1986)が、5 年の期間で物理学雑誌を対象とした調査をした指標を利用した。本研究では、引用を利用、反応時間の 5 年を 8 年として算出した。本研究で用いる単位は年度である。元の計算式は、 $MRT = -\ln(f_0 + f_1e^{-1} + f_2e^{-2} + f_3e^{-3} + f_4e^{-4})$ であった。本研究における期間を考慮した結果、 $MRT = -\ln(f_0 + f_1e^{-1} + f_2e^{-2} + f_3e^{-3} + f_4e^{-4} + f_5e^{-5} + f_6e^{-6} + f_7e^{-7})$ という計算式を用いた。ここにおける MRT が平均反応時間、ln は自然対数、f0 は登録された年度に利用された割合である。(0 の部分には年度を表す自然数が入る)

本研究では、リポジトリ全体、紀要、学术论文、学位論文における反応平均時間を算出した。

表 67 平均反応時間と平均値まとめ

	リポジトリ全体	紀要	学术论文	学位論文
利用までの平均年度	0.791	1.429	0.319	0.799
平均反応時間	0.342	0.606	0.188	0.302

利用のなかったコンテンツでは、コンテンツの利用までにタイムラグが発生するという仮説を立てることができたが、実際に利用されているコンテンツで 検証を行った結果、上記の様な数値が得られた。単純集計に平均年度と、平均反応速度算出のための計算式を用いた結果どちらも反応速度は、1 を下回る結果となった。つまり、利用の多いコンテンツに関しては、登録されてから初回利用まで 1 年かからない結果となった。

3.3 検索語句に関する調査

機関リポジトリの利用実態を明らかにするために、検索語句による分析を実施した。調査内容は、検索語句数、検索語句登場総数、検索語句によるタイトルの hit 数、検索回数登場語句上位 15 件、検索語句の言語比率である。ここにおける検索語句とは、リポジトリのサーバに蓄積された検索語句から、ストップワードや検索式を機械的に除去する処理を行った後のデータである。

機関リポジトリの検索語句を明らかにすることで、利用者が探索しているコンテンツの変化を調査することを目的として、登場回数の検索語を調査した結果、上位 15 件に入っている日本語と英語(アルファベット)の比率が 2016 年以降変化していたため、2016 年以降は検索語の言語比率に変化があったと予測し、調査をおこなったが、実際には検索語の言語比率は、2010 年からほぼ一定であることが判明した。

表 68 検索語句と年度

年	2010	2011	2012	2013	2014	2014	2015	2016	2017
検索 語句数	6,972	3,814	553	636	584	12,084	30,046	49,084	63,310
検索語句 登場総数	65,160,062	8,252,447	6,734,438	4,432,595	1,778,618	31,592	116,135	292,552	616,364
hit 数	1,633	189,966	41,039	51,747	58,559	321,107	650,862	889,815	1,191,604

検索語句数とタイトルに内包されている語句数をもとめて hit 数を算出した。2010 年までのデータは、サイト内検索のログのみが別で保存されていたため、検索語句に機械によるアクセスの回数が加算され、総数が非常に大きくなっていた。

表 69 利用者が検索した回数の多い語句(上位 10 件)

rank	2010	検索 回数	2011	検索 回数	2012	検索 回数	2013	検索 回数	2014 early	検索 回数
1	seiji	2,399,751	スギエ	1,396,356	tabuchi	2,296,882	tabuchi	907,395	yanagisaw a	644,461
2	筑波大学 平家部会 論集	1,807,207	no_searc h_analy zer	1,343,728	yanagisa wa	974,365	yanagisaw a	357,590	教育学系論 集	336,860
3	substrat e	1,097,362	戦術	790,075	コウノ	387,036	抽出	308,970	貝沼	74,104
4	sensitize d	1,014,884	sciences	763,232	ヤナギサ ワ	320,278	博乙第	262,687	徳川慶喜	64,437
5	医療	872,367	高齢者	698,537	貝沼	231,972	onda	151,257	onda	56,260
6	クロール	778,326	佐々木	670,912	クノ	161,546	クノ	111,121	クニジロウ	53,556
7	家族画	755,559	tabuchi	645,200	journal	132,138	haruo	97,963	コウノ	51,250
8	cat	690,046	授業	141,777	takamori	131,392	を通して	97,486	クノ	29,450
9	歳森	613,291	électrom agnétiqu es	121,065	タカヒロ	127,113	yasumasa	77,077	yasumasa	27,584
10	hidekazu	588,380	江草	108,126	クニジロ ウ	103,126	丸浜	70,973	習熟	24,603

2010 年 3 月から 2014 年 3 月までの検索語句のログを基に、検索回数が多い上位 10 件を表にした。上位の件数は、機械によるアクセスである可能性が高いが、検索語句には日本語と英語(アルファベット)が混ざった結果となった。

表 70 利用者が検索した回数の多い語句

rank	2014late	検索回数	2015	検索回数	2016	検索回数	2017	検索回数
1	体育系紀要	500	体育系紀要	1,184	新谷真由	964	three	6,956
2	発生運動学の方 法論	300	長谷川	677	takashi	569	hiroshi	1,343
3	コソの言語表現 の構造	155	悟郎	616	akira	505	佐藤	1,310
4	豊	94	小出	589	watanabe	474	takashi	1,227
5	星野	88	篤史	584	hiroyuki	402	鈴木	1,217
6	佐野淳	88	微積分	579	satoshi	400	suzuki	1,154
7	a!"object"&&a! b)return	84*	新谷	391	tanaka	374	akira	1,014
8	日本語	76	真由	364	nakamura	372	tanaka	955
9	筑波大学心理学 研究	71	hasegawa	316	atsushi	372	伊藤	928
10	study	69	小林	314	kobayashi	348	田中	884

2014 年 5 月以降の検索語句のログを基に、検索回数が多い上位 10 件を表にした。上位の件数は、機械によるアクセスである可能性が高いが、2015 年までと、日本語と英語(アルファベット)2016 年以降では、言語の比率が逆転していた。

表 71 各年度に占める検索語句の言語

year	all word	ja	en	other
2010	6,972	4,007	2,947	18
2011	3,814	2,414	1,391	9
2012	553	321	230	2
2013	636	382	254	0
2014early	584	366	218	0
2014late	12,084	8,925	3,132	27
2015	30,046	21,060	8,892	94
2016	49,084	32,082	16,858	144
2017	63,310	41,836	21,244	230

結果として、検索語句の日本語と英語の比率はあまり変化しておらず、検索語句の多様化が見られた。

4. 考察

4.1 アクセス元とデータの利用に関する考察

DSpace を利用して機関リポジトリを運用していた 2010 年 3 月から 2014 年 3 月までの機関リポジトリの利用状況と JAIRO Cloud に移行してからの 2014 年 5 月から 2017 年 12 月の機関リポジトリの利用状況では、コンテンツのダウンロード回数において、大きな変化があった。

DSpace で機関リポジトリを運用していた期間は、コンテンツ利用回数が 2010 年 3 月から 2014 年 3 月までの期間において大きくは変化せず、10 万件から 20 万件の範囲に収まっていた。だが、JAIRO Cloud で機関リポジトリを運用し始めてからは、利用回数が指数関数的に上昇しており、2017 年の年間の数値は、250 万件近い利用数になっていた。

アクセス元の Host の数を考えると、2010 年 3 月から 2014 年 3 月までの期間は、2011 年を除いて毎年 20 万件以上あったアクセス元が、2014 年 5 月から 2017 年 12 月までの期間では、1 万件から 4 万件程度になった。これは、コンテンツへのアクセスをした IP アドレスではなく、WEKO のトップページにアクセスした IP のみを計測しているためだと考えられる。

DSpace は、コンテンツにアクセスした IP やファイル閲覧のみした IP、トップページに訪れた IP などをアクセスログに残っている url から指定して抽出することが可能であるが、JAIRO Cloud の出力ではそれに現状では対応していない。そのため、機関リポジトリにアクセスしたアクセス元の一部を知ることではできるが、実際にコンテンツをダウンロードするアクセス元と、トップページでデータの収集のみをするアクセス元などを区別して分析することができない。また、実際のコンテンツ利用回数とアクセス元の多様性を考えても、JAIRO Cloud のアクセス元の Host が利用動向を正しく示す指標になっているとは考えられない。

筑波大学の機関リポジトリを例に考えると、トップページでのみ得ることのできる情報は、インデックスの一覧のみである。また、先行研究において示されたアクセス元として、google からコンテンツページへの遷移が利用の大半を占めていたことを踏まえて考えると、トップページへのアクセスを見ることは、機関リポジトリの利用の実態を示すデータではないといえる。だが、その点に関する指摘は公的にはなされていない。現在までに、公的に指摘がない事実を考えると、機関リポジトリに登録されたコンテンツのアクセス元は意識されていなかった可能性がある。

2016 年以降にアクセス元の比率が変化した原因の一部は、クローラリストの更新の停止によるクローラからのアクセスの除外が一部なされていないことにあると考えられる。クローラを作成し、情報を収集している元が jp ドメイン以外であることを考えると、情報としての需要に応じて、コンテンツを提供する環境を整えることも必要であると考えられる。2016 年からは、OpenAIRE へのデータ提供も開始したため、アクセス数の増加とともに、クローラからのアクセスも増加するため、特に影響が大きかったと考えられる。こ

のように、データを処理する環境が変化することは、今後も考えられる。データリポジトリとしての機能を持ち、コミュニティ機能を持つリポジトリの実装となれば他のデータリポジトリとの連携も考えうる。また、著作権法の改正後は、データ利用が許容される範囲が広がるため、他の機関からのアクセスも増加する可能性がある。このように機関リポジトリへのアクセス元に関する環境は日々変化が予測されている。

このことに関する問題点として、アクセスログのフィルタリングやスクリーニングは、独自構築の機関リポジトリであれば、手元にあるデータに対して処理を増やすことでクローラなどの影響に対応することが可能である。だが、本研究でも利用した **WEKO** の出力データは、データを業務統計用に切り出した形で提供されている。これは、業務担当者の手間という観点では簡便で良い仕組みであるかもしれないが、今回の分析のように、大学における利用実態を調査する際のフィルタリングに利用することができる情報が少ないともいえる。**JAIR Cloud** がフィルタリングする IP アドレスの更新をやめると、システムを利用している大学のデータすべてに影響がある。出力データから個別の大学がアクセスログに適切な処理を追加することができない。大学別に発行している業務統計や大学の競争を考えた際の運営戦略に使用することのできる信頼性の高い個別のデータがないこととなる。国立大学の法人化などで各大学に個性が求められる中、各大学が個別に学術的生産物としての需要を正しく把握できないことは、デメリットであるといえる。

Fei(2010)の研究でも、電子資源に関する統計は、意思決定に有用であることが示されたが、同時に図書館員の手間が大きいことも示されていた。本研究では、図書館員の手間を軽減するシステムを利用した結果、意思決定に必要な情報を得られない可能性を示唆したため、異なる観点から同一の結果を得られたといえる[121]。

機関リポジトリや **JAIR Cloud** に今後期待される役割として、オープンサイエンスの文脈でのデータ基盤構築がある。大学や研究所で生産された学術論文のエビデンスデータを保存や、データの 2 次利用に関する需要が高まっているためである。現在、機関リポジトリでは、研究者によるコンテンツの登録があまりなされておらず、図書館員による登録が主となっていることが、実践報告として挙げられており、先導的な取り組みとしても紹介されていた[212]。今後、求められているこれらの機能では、学術論文の生産と公開、査読に伴い研究者による自主的なデータの登録、利用が求められている。現在、**NII** と **NIMS** によって開発が進められている次世代リポジトリ[213]は、**COAR** の要件に従って開発が進められている[214][215]。次世代リポジトリは、研究データ保管のプラットフォームやコミュニティ機能の実装が望まれている。研究者によるコンテンツ管理・保管のためのプラットフォームとしての機能である。本研究で示した、データの利用可能性と分析に用いられるデータの管理の手間が **JAIR Cloud** では解決されていた。だが、本研究における批判の様に、データの利用の幅が狭まっている側面もある。今後、データ基盤構築が進められると、研究者による自主的なデータ公開のための支援が図書館員に必須の役割

となることは確実であろう。そのためにも、現在も育成が期待されているデータライブラリアンの需要が増加することが予測される[216]。

4.2 対象機関におけるメタデータの現状

アクセスログ分析を通して、メタデータに関する問題点も明らかとなった。筑波大学の機関リポジトリのコンテンツ数は、ID で考えると 48226 件、実数 46947 件であり 1,200 件ほどのメタデータが欠損していた。著作権や著者の希望でコンテンツを削除した可能性もある。だが、コンテンツの利用がログに残っており、現在も利用が可能な 8 件のコンテンツのメタデータも存在しなかった。博士論文公開義務化のあとに junii2 では、`<degreeName>`というメタデータが付与された[211]。それ以前は、`<description>`に記載する内容であった。メタデータをハーベスティングし、リクエストした形式に変換して xml を提供する API を利用しているため、`oai_dc`形式でリクエストをした場合は、学位が`<description>`に記載されるが、junii2 でリクエストした場合は、`<degreeName>`と`<description>`に分かれて表示された。学位名をパースしたところ、5,712 件から 10,339 件まで自動的に付与されたメタデータ数が変化した。`oai_dc`に対応している状態のみで良いのかを検討する必要がある。例えば、`oai_dc`を利用することを考えた python のライブラリなどでは日本語に対応しておらず、文字コードの変換に必要以上の手間が生じる。日本語のメタデータの扱いやすさは、英語のメタデータと比較し、利用がしにくい。メタデータフォーマットや記述言語を管理者と利用者を意識した形で登録することは重要である。OAI-PMH で利用可能なデータとしてメタデータを公開しているということは、図書館が内部向けに資料管理用のメタデータを付与しているのとは大きく異なる現状の理解が求められる。

著作権の一部を改正する法律が 2019 年 1 月に施行され、「デジタル・ネットワーク化の進展に対応した柔軟な権利制限規定の整備」で、論文等のデータが利用可能になる[80]。今後いっそう、OA コンテンツの利用が促進されることとなると考えられる。メタデータが不十分だと正しい利用がなされない可能性もある。そのため、著作権法の施行の前に機関リポジトリのメタデータの確認は必須事項とも言える。

4.3 今後求められるデータ管理と図書館員の役割

4.2 で対象機関におけるメタデータの現状を確認したが、項目の記述が適切でない事例が散見された。4.2 では、博士論文におけるメタデータ項目の`<degreeName>`が`<description>`から変更されておらず、現状に迫いついていない事例を紹介したが、`<NIItype>`と`<NDC>`など様々な項目においてメタデータが記述されていない現状がある。今後、図書館員に求められる役割は、コンテンツの登録だけでなく、データやコンテンツを登録する研究者の支援や、データの整備、管理運用も含まれることになる。全文検索可能な資料は、メタデータが適切に記述されていなくも本文のテキストデータを利用すること

で、検索することが可能である。著作権法の改正により、全文検索の際に、著作者の許諾なくデータを利用できるようになるので、テキストデータの存在する学术论文等の資料に関しては、この傾向が強化されると予測できる[80]。今後の図書館員に求められるデータリテラシーとして、適切なメタデータ項目の記述や、利用可能な形でメタデータの運用に関する知識がある。そのためにも、現状のメタデータ項目の整理や、メタデータの管理運用ができる人材の育成が必要である。また、図書館は”象牙の塔”における資料の貸し出しや、リポジトリへの資料の提供だけでなく、研究者とのコミュニケーションを意識する必要がある。現在、機関リポジトリで行われている研究者による生産物である資料を代理で登録する仕組みと違い、研究におけるデータ管理や、データの公開・利用、データの検索、メタデータの適切な記述などは、研究段階から求められている。現在開発の進められている次世代リポジトリでは、**JIARO Cloud** を利用したクラウドベースのデータ基盤が想定される。現在、機関リポジトリを運営している図書館員は、データの利活用にプラットフォームを利用する研究担当者や研究者をデータ基盤構築に携わる立場として支援する必要がある。本研究で指摘したように、データの分析やデータの適切な記述やメンテナンスは今後予測されるオープンデータの時代において必須であると考えられるため、図書館にはデータ利用に対応できる基盤としての役割を担う存在として変化することを期待したい。

4.4 コンテンツ利用に関する考察

DSpace を利用して機関リポジトリを運用していた 2010 年 3 月から 2014 年 3 月までの機関リポジトリの利用状況と **JAIR O Cloud** に移行してからの 2014 年 5 月から 2017 年 12 月の機関リポジトリの利用状況では、コンテンツのダウンロード回数において、大きな変化があった。このリポジトリに登録されているコンテンツ利用の増加は、機関リポジトリに登録されたデータ全体の傾向であるとともに、個別に分析を実施した、紀要、学术论文、学位論文、**Research Paper** すべてに共通した。

4.4.1 紀要に関する考察

紀要は、日本の機関リポジトリの特徴として諸外国と比較しても非常に大きな割合を占めるコンテンツであった。日本全体で見ると、機関リポジトリに登録されているコンテンツの 5 割以上が紀要であることから日本の機関リポジトリが紀要リポジトリと呼ばれることもわかる[194]。紀要は、学术论文という位置づけの雑誌もあれば、大学の記録としての灰色文献としての位置づけの雑誌も存在する。また、雑誌内でこれらが混合していることもあるので、文献単位での適切なメタデータの付与が必要である。

本研究で得た紀要の分析結果は、コンテンツ数の多い紀要の利用が多いことは当然ながら、コンテンツ数の多い雑誌よりも利用数が多い雑誌も上位には複数見られた。コンテンツ数の一番多い紀要は、『**Tsukuba Journal of Mathematics**』であり、上位 15 件で唯一の

STEM 分野であった。だが、利用数では 8 番目であり、コンテンツ数で 1/3 程度の紀要よりも利用が少なかった。対象の機関リポジトリの傾向として、コンテンツ数と利用数は同程度の順位を示す結果が得られていたので、この利用数は少ないといえる。代わりに、『筑波大学心理学研究』『体育科学系紀要』『心身障害学研究』『人文科教育研究』などの人文・社会科学系の分野から学術分野の紀要が上位を占めていた。また、これらの分野は筑波大学の中でも有名な学問分野である。機関リポジトリに登録された紀要が実際に利用されており、一部においては STEM 分野の紀要と比較しても利用が多いという事実は、研究成果発表の場として、国際誌ではなく紀要などを選択する機会の多いとされる人文・社会科学分野の研究者のメリットとなる。また、JAIRO Cloud 移行前後で比較をすると、利用されているコンテンツの割合が急激に増加した。2010 年 3 月から 2014 年 3 月までに登録されたコンテンツが同期間内に利用された割合が 2%弱であったのに対して、2014 年 5 月から 2017 年 12 月までに登録されたコンテンツが利用された割合は 96%であった。また、全期間を通して利用されたコンテンツの割合が、97%以上になった。前半の期間では、利用されるコンテンツが固定化されており、毎年同じコンテンツが利用されていたが、後半の期間では広範なコンテンツが利用されることとなった。これは、機関リポジトリが紀要デポジトリとも呼べる状態であったことが改善されたことを示す結果となった。

また、紀要の利用上位 15 件のうち 12 件は異なる雑誌に掲載されたコンテンツであることから、利用されているコンテンツが多岐にわたることをうかがえる。

4.4.2 学術論文に関する考察

学術論文は、Green OA の本来の目的であるポストプリントのコンテンツである。対象の機関リポジトリは、日本の機関リポジトリの中でもコンテンツに占める学術論文の多さが特徴である。この背景には、対象の大学が電子図書館プロジェクトに長く携わってきたことがあると考えられる。OA として公開されるようになった学術論文については、引用数の増加や利用の分散などの議論があるが、結論は出ていない。本研究では、単一の大学の機関リポジトリのコンテンツ利用を調査したため、これらの議論に対する回答は持ち得ていない。

機関リポジトリに学術論文を登録することで、コンテンツの利用が長期間にわたり、広範なコンテンツであったことを明らかにした。紀要と比較すると各雑誌のコンテンツ数はあまり多くない。だが、総利用回数が多い雑誌が紀要よりも多いことが明らかになった。つまり、紀要よりも学術雑誌の方がコンテンツ利用が多くなされているということである。その傾向は、2010 年 3 月から 2014 年 3 月までのコンテンツ利用割合に顕著に表れている。紀要が 2%であったのに対し、学術論文は 75%程度の利用があった。2014 年 5 月から 2017 年 12 月の期間でも 93%、全期間では、98%のコンテンツが利用されている。JAIRO Cloud 移行前後において、利用の割合が大きく変化しているが、紀要ほどの影響はなかった。ここから、学術論文はコンテンツとしての価値があるため、プラットフォーム

への依存が紀要ほど大きくないという予測を立てることができる。また、興味深い結果として、メガ OA 誌として知られている PLOS ONE と Scientific Reports に関しては、後者に属するコンテンツが前者のコンテンツ数よりも 20 件ほど多かったにも関わらず、雑誌としての利用数では、前者が後者のほぼ 2 倍の利用であることを明らかにした。この 2 誌の関係は、OA の推進のために作られた学術情報出版社(PLOS)と、伝統的な学術情報出版社(Nature)の関係であり、コンテンツ数で、後者が OA 誌で前者を抜いて首位になったことが知られている。OA ジャーナルの機関リポジトリでの OA は、相互作用をもたらすのか、利用が分散するのか、今後調査をする必要がある分野であると考えられる。

4.4.3 学位論文に関する考察

学位論文は、2013 年 4 月 1 日に博士論文の公開義務化が施行されてからコンテンツ利用の実態に注目が集まっているコンテンツであるといえる。本来ならば、博士論文に焦点を絞った分析をするべきであったが本研究では、博士論文と修士論文を区別していない。理由としては、メタデータが< Thesis or Dissertation >であり、学位もメタデータの適切な記述がなされているか不明な分野だったため、区別を行わなかったためである。実際に、コンテンツ利用上位 15 件のうち、『イメージ操作は脳の中でどのように行われるのか：心的回転の回転方向に関する実験的検討』は情報学、『子どもの読書の意義と効果に関する研究』は図書館情報学の修士論文であることが確認できた。学位別にみると、工学、医学、Science の学位論文数が 1,000 件を超えるなど分野としての論文数や利用回数は多かったものの、上位 15 件には、工学以外の 2 分野は登場しなかった。代わりに、コンテンツ別の上位 15 件には、体育科学が 3 件、工学が 2 件、その他が 1 件ずつと広範な分野の学位論文が利用されていることが判明した。

学位論文などの学術論文と異なり、引用による計量分析が困難とされている論文の分析の指標として期待されているオルトメトリクスを Ceek.jp にて調査した結果、工学、体育科学は 5 位、6 位であり、上位には、図書館情報学、教育学、学術、文学が現れた。その下位には医学、情報学、国際政治経済学、言語学などが続いた。情報学、国際政治経済学などのコンテンツ数は、上位 15 件に含まれておらず、逆に含まれている農学、理学のオルトメトリクスは多くなかった。今回の分析の結果からは、コンテンツ数や、学位別の利用回数とオルトメトリクスには、あまり関係がないといえる。

4.4.4 Research Paper に関する考察

Research Paper に関する分析は、利用のなかったコンテンツの分析において、割合が顕著に高かった type 種別であることに着目して実施した。結果として、利用のなかったコンテンツの登録年が 2010 年以前と 2017 年に多いことが判明した。後者に関しては、利用されるまでの時間があるため、あまり不思議ではないが、他の分野の分析でオブソレッセンスがあまり見られず、2007 年のコンテンツが利用されていたコンテンツの 50%に至

る結果も見られた中で、過去のコンテンツがあまり利用されていない結果が明らかになった。紀要は、2%程度の利用だったコンテンツが 96%利用されるようになったため、JAIRO Cloud に移行したことによりコンテンツの広範な利用が実現していたが、Research Paper は全体で 73%程度と低い結果になった。DSpace の際も 50%近く利用されていたため、あまり変化の影響を受けていない。

4.4.5 コンテンツ利用に関するまとめ

DSpace から JAIRO Cloud への移行によって、最も影響を受けたのは紀要である。また、学術論文や Research Paper は、利用されるコンテンツの範囲が広がり、利用回数が上昇するなどの変化はあったが、紀要ほど顕著な変化は生じなかった。学位論文に関しては、影響が学術論文と紀要の中間程度であった。だが、学位論文に関しては、システムの移行期間と博士論文の公開義務化という 2 つの大きな変化を与える要因が被っているため、システムを継続して利用し続けた大学の利用動向を見て、変化への影響を判断するのが妥当であると考えられる。

筑波大学のコンテンツ利用実態には、大学としての特徴を示すようなデータが示された。筑波大学において特徴的な分野である体育科学や物理、言語学などの分野のコンテンツ利用数が、コンテンツ数のランキングとは異なり上位に来たためである。オブソレッセンスを測定した際にコンテンツの利用数は、コンテンツ数に依存しており、コンテンツの登録年による減衰が見られなかった。このことからコンテンツ数は影響の強い変数であると考えられる。だが、紀要や学術雑誌、授与学位を単位とした利用数のランクでは、コンテンツ登録数を直接反映する結果とはならなかった。その際に、上記で示した分野が利用の上位となっていたため、大学の特徴と一致するようなコンテンツの利用があると判断した。

この結果から、機関リポジトリのコンテンツ利用実態は、大学の特徴と一致するという仮説を構築した。本研究の限界として、筑波大学の機関リポジトリのみを対象としたため、仮説の検証には至らなかった。仮説が成立した場合、大学の特色となる情報を発信、利用可能なメディアとしての位置づけが可能になる。仮説としてまとめると、「機関リポジトリの利用実態として、コンテンツ数は大きな影響を持つが、大学の特色ともいえる分野に関しては、量的に少なくとも利用が多い雑誌が存在する。このことから、機関リポジトリの利用実態は、機関によって特徴のある学問分野を表すメディアとなっている。」である。その場合、JAIRO Cloud の利用がデメリットになりうるので、今後検証する必要性がある。

DSpace と JAIRO Cloud の利用実態を調査して明らかになったメリットは、リポジトリに死蔵されていたコンテンツが利用されるようになった点である。紀要や Research Paper は、2010 年以前に登録されたが、2010 年 3 月から 2014 年 3 月までの間で、10%程度の資料しか利用されていなかった。だが、2014 年 3 月から 2017 年 12 月までに間に

それらの資料は 50%以上が利用される結果となった。紀要に関しては、99%近いコンテンツが利用される様になった。

DSpace においては紀要や Research Paper といった type 種別の利用されるコンテンツが固定化されており、全体としてもコンテンツ利用数が安定していた。だが、JAIRO Cloud 導入後は、利用数や多様なコンテンツが利用されるようになった。サイト内で利用された検索語句を見ても、同様の傾向が見られた。

機関リポジトリの登録コンテンツとして、代表的な紀要、学位論文、学術論文を比較した。利用を変数とした平均反応時間では、学術論文が一番平均反応時間が早く、紀要が一番遅い結果となった。今回の分析では、それぞれにオブソレッセンスを調査したが利用の減衰は見られず、どのコンテンツもコンテンツ数に依存した結果となったが、初回の利用までにかかる期間は、異なることが明らかとなった。だが、どれも年度を単位として、1 年度を超えることはなく、OA コンテンツが登録されてから利用されるまでの時間は、非常に短くなっていた。

4.5 考察のまとめ

DSpace から JAIRO Cloud への移行によって、コンテンツの利用では、利用されるコンテンツが固定化されていた機関リポジトリのコンテンツ利用が広範になる変化が見られた。また、全体のとしての利用数や各コンテンツの利用数が毎年約 2 倍になるペースで増えている。だが、国立情報学研究所によるクローラリストの更新停止など、システム面のサポートに関する不備で、各大学のコンテンツ利用を示すデータにノイズが入っている可能性を否定できない。本研究では、コンテンツ利用に関するデータの分析、アクセス元に関する分析やメタデータの調査を通して、機関リポジトリを運用するシステムの変化の影響を調査した。この調査を通して、機関リポジトリの利用促進という面において、JAIRO Cloud への移行後の良い側面が明らかになった。だが、一方で大規模大学である筑波大学のように、大学でデータライブラリアンを育成可能な組織である場合は、JAIRO Cloud の利用でなく、独自構築の機関リポジトリを利用したほうが、適切なデータによる利用状況の調査が可能である可能性も示唆した。今後、次世代リポジトリの運用が開始された場合、独自構築の機関リポジトリを運用している大学も、コミュニティやデータリポジトリとして JAIRO Cloud を利用する可能性がある。その場合は、図書館による意思決定だけでなく、データ分析が大学の学術情報生産傾向を示す材料になることを考慮したうえで、システムを選択する必要があるだろう。また、各大学でのデータライブラリアンの育成、メタデータの適切な記述や研究者への指導など、データを扱う図書館員としての準備が必要になる。本研究の対象以外のリポジトリの調査が OA コンテンツや博士論文に注目した調査には必要であると考えられるため、図書館員によるデータ分析やデータ利用の事例を増やして、経験を蓄積していく必要があるだろう。

5. 結論

2010年3月から2017年12月までの筑波大学の機関リポジトリを通したオープンアクセスコンテンツの利用実態を明らかにした。結果として、2010年3月から2014年3月まで運用されていたDSpaceと2014年5月から2017年12月までと現在も運用されているJAIR Cloudでは、利用されるコンテンツの多様性と利用されるコンテンツ数に大きな変化が生じたことが示された。コンテンツ利用の多様化や、コンテンツ利用数の変化の割合の大きさなどは、長期的なアクセスログ分析でのみわかる結果である。インフラストラクチャとしての機関リポジトリは、2014年段階でも安定して運用されていたが、資料の一部が利用されず死蔵されている大学図書館と同様の傾向を示していた。だが、クラウドベースのシステムの導入後は、コンテンツ利用の多様化が進んだ。

コンテンツ利用の実態を明らかにするために、DSpaceとJAIR Cloudのアクセスログデータを用いたことによって、データの性質の違いが明らかになった。アクセスログデータの利用において、独自構築されたりリポジトリとクラウドベースで管理されたりリポジトリのメリットとデメリットと考えられる特徴があった。アクセスログデータを分析する際の効率化と各機関で利用するために加工する際の自由度が例として挙げられる。

本研究では一大学の機関リポジトリの長期的なアクセスログに対する分析を行ったことで、インフラストラクチャとしての機関リポジトリの利用実態や、機関リポジトリ運用の際のメリット、デメリットとなる点を明らかにすることができた。だが、1つの大学のみで調査を行ったため、「機関リポジトリの利用実態として、コンテンツ数は大きな影響を持つが、大学の特色ともいえる分野に関しては、量的に少なくとも利用が多い雑誌が存在する。このことから、機関リポジトリの利用実態は、機関によって特徴のある学問分野を表すメディアとなっている。」という仮説部分についての検証を行うことができなかった。機関リポジトリの更なる利用実態を明らかにするためには、複数の大学を対象とし、長期的なアクセスログを用いた検証の必要性があると考えられ、今後の課題となった。

謝辞

本研究を進めるにあたり、ご指導いただいた逸村裕教授、歳森敦教授に深く感謝申し上げます。データの提供にご協力いただきました筑波大学附属図書館の皆様、データの確認や入手に関するアドバイスをいただいたデジタルライブラリ担当の嶋田様、松野様、リポジトリ担当の佐藤様にも感謝申し上げます。また、大学院在学中にデータの取り扱いやデータ分析に関する指導をしてくださった **Clarivate Analytics Japan** の **Life Science Solution** 部門の皆様にも感謝いたします。また、分析の際に必要なプログラミングに必要な思考法のコーチングと実践の場をいただいた岡本様にも感謝いたします。大学院在籍にあたり、ご支援いただいた肥田野信東京女子医大名誉教授に感謝いたします。本研究を進めるにあたり、計算機資源を貸して下さり、助言、校正をいただいた逸村研究室の皆様やアドバイスをいただいた逸村研究室の卒業生の皆様に感謝を申し上げます。

文献リスト

- [1] “Read the Budapest Open Access Initiative”. Budapest Open Access Initiative. 2002, <http://www.budapestopenaccessinitiative.org>, (accessed 2018-07-18).
- [2] Jean-Claude Guéron. Budapest Open Access Initiative. “Open Access: Toward the Internet of the Mind”. 2017, <http://www.budapestopenaccessinitiative.org/boai15/Untitleddocument.docx>, (accessed 2018-07-18).
- [3] 栗山正光. デジタル資料保存リポジトリの動向. Current Awareness Portal. 2005, <http://current.ndl.go.jp/ca1561>, (accessed 2019-01-08).
- [4] "Uniform Resource Identifier (URI): Generic Syntax". The Internet Society. 2005, <https://tools.ietf.org/html/rfc3986>, (accessed 2018-12-25).
- [5] 倉田敬子. オープンアクセスとは何か. 情報の科学と技術. vol. 60, no. 4, 2010, p. 132-137.
- [6] "ArXive.org". ArXive.org. <https://arxiv.org>, (accessed 2018-07-18).
- [7] "Los Alamos National Laboratory". Los Alamos National Laboratory. <https://www.lanl.gov>, (accessed 2018-07-18).
- [8] "cornell university library". cornell university library. <https://www.library.cornell.edu>, (accessed 2018-07-18).
- [9] "Cornell Computing and Information Science". Cornell Computing and Information Science. <https://cis.cornell.edu>, (accessed 2018-07-18).
- [10] "PMC". PMC US National Library of Medicine National Institutes of Health. <https://www.ncbi.nlm.nih.gov/pmc/>, (accessed 2018-07-18).
- [11] "U.S. National Library of Medicine". U.S. National Library of Medicine. <https://www.nlm.nih.gov>, (accessed 2018-07-18).
- [12] "Welcome to NCBI". National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov>, (accessed 2018-07-18).
- [13] 佐藤義則. シリアルズ・クライシスと学術情報流通の現在: 総括と課題. 情報管理. 2011, vol. 53, no. 12, p. 680-683.
- [14] 三根慎二. オープンアクセスジャーナルの現状. 大学図書館研究. no. 80, 2007, p. 54-64.
- [15] “Ten years on from the Budapest Open Access Initiative: setting the default to open BOAI 10 Prologue: The Budapest Open Access Initiative after 10 years”. Budapest Open Access Initiative. 2012, <https://www.budapestopenaccessinitiative.org/boai-10-recommendations>, (accessed 2018-07-18).
- [16] 文部科学省. “学位規則の一部を改正する省令(平成 25 年文部科学省令第 5 号)”. 学位規則の一部を改正する省令の施行について. 2013,

- http://www.mext.go.jp/a_menu/koutou/daigakuin/detail/1331790.htm, (accessed 2018-07-18).
- [17] 日本学術振興会. “科学研究費助成事業 オープンアクセス”. 2010, https://www.jsps.go.jp/j-grantsinaid/01_seido/08_openaccess/index.html, (accessed 2018-07-20).
- [18] 文部科学省. “機関リポジトリの活用による情報発信機能の強化について”. 研究環境基盤部会学術情報基盤作業部会. 2012, http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/002-1/siryo/attach/1323931.htm, (accessed 2018-07-20).
- [19] 高池宣彦. 大学認証評価における大学図書館の評価：認証評価機関の評価基準と評価結果を中心に. 三田図書館・情報学会, 2016, no. 75, p. 1-36.
- [20] "The Clarivate Analytics Impact Factor". Clarivate Analytics. <https://clarivate.com/essays/impact-factor/>, (accessed 2018-07-18).
- [21] The PLoS Medicine Editors. "The Impact Factor Game". PLoS Med. vol. 3, no. 6, <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0030291>, (accessed 2018-07-18).
- [22] "San Francisco Declaration on Research Assessment". DORA, <https://sfedora.org/read/>, (accessed 2018-07-18).
- [23] 逸村裕, 池内有為. インパクトファクターの功罪: 科学者に与えた影響とそこから生まれた歪み. 化学, 2013, vol. 68, no. 12, p. 32-36.
- [24] 栗山正光. ハゲタカオープンアクセス出版社への警戒. 情報管理. vol. 58, no. 2, 2015, p. 92-99.
- [25] Beall, Jeffrey. "Scholarly Open Access Critical Analysis of Jeffrey Beall's Blog - Open Access Publishing". <http://scholarlyoa.net>, (accessed 2018-07-18).
- [26] Alexander M. Clark ; David R. Thompson. Five (bad) reasons to publish your research in predatory journals. 2016, vol. 73, no. 11, p. 2499-2501.
- [27] 林和弘. オープンアクセスを踏まえた研究論文の受発信コストを議論する体制作りに向けて. 科学技術動向. 2014, no. 145, p. 19-25.
- [28] 佐藤翔. 査読の抱える問題とその対応策. 情報の科学と技術. 2016, vol. 66, no. 3, p. 115-121.
- [29] 山地一禎. "オープンサイエンスと学術リポジトリとの発展的連携とその期待". 国立情報学研究所 TITEC 平成 29 年度第 1 回情報活用 IR 研究会. 2017, <http://rcos.nii.ac.jp/item/2017/0904/20170904yamaji.pdf>, (accessed 2018-12-25).
- [30] 国立大学図書館協会 学術情報委員会 学術情報の利用促進と保存プロジェクトチーム. 学術情報の利用促進と保存プロジェクトチーム報告. 2014, p. 1-16. <https://www.janul.jp/j/projects/si/gkjhoukoku201406b.pdf>, (accessed 2018-07-18).

- [31] 森石みどり. コレクション構築の現在: シェアード・プリント: 米国の大学図書館における冊子体資料の共有と保存. 情報の科学と技術. 2015, vol. 65, no. 9, p. 386-391.
- [32] 長谷川豊祐. 図書館経営: 変化への適応: 大学図書館における業務分析と業務管理手法. 2011, vol. 61, no. 8, p. 311-316.
- [33] 熊谷俊夫, 重里信一. リソース・シェアリング: 資料利用のための協力: 学術資源の全国的保存システムと共同保存図書館. 情報の科学と技術, 1993, vol. 43, no. 11, p. 996-1005.
- [34] “ランキング 統計期間 2017-12-25 -2018-12-25”. つくばリポジトリ.
<https://tsukuba.repo.nii.ac.jp>, (accessed 2018-12-25).
- [35] 佐藤翔, 逸村裕. 機関リポジトリとオープンアクセス雑誌: オープンアクセスの理念は実現しているか?. 情報の科学と技術. 2010, vol. 60, no. 4, p. 144-150.
- [36] Peter Suber. "Timeline of the Open Access Movement". Timeline of the Open Access Movement. 2009, <http://legacy.earlham.edu/~peters/fos/timeline.htm>, (accessed 2018-12-26).
- [37] 岡部晋典, 佐藤翔, 逸村裕. Budapest Open Access Initiative の思想的背景とその受容. 情報知識学会. 2011, vol. 21, no.3, p. 333-349.
- [38] 図書館情報学用語辞典編集委員会. "フォーマルコミュニケーション". 図書館情報学用語辞典. 第4版, 2013, 284p.
- [39] 科学技術・学術政策研究所. "第4章 研究開発のアウトプット". 科学技術指標 2016. 2016, http://www.nistep.go.jp/sti_indicator/2016/RM251_41.html, (accessed 2018-12-26).
- [40] 倉田敬子. "学術コミュニケーションの「場」はどう変わってきたのか-直接対話からプラットフォームへ-". NIRA 研究報告書 孤立する日本の研究プラットフォーム-放置すれば日本の科学そのものが衰退する-. 公益財団法人 総合研究開発機構. 2015, 111p, <http://www.nira.or.jp/pdf/1502report.pdf>, (accessed 2018-12-26).
- [41] 横井慶子. 学術雑誌におけるオープンアクセスジャーナル. 慶應義塾大学, 2014, 博士論文.
- [42] Clarivate Analytics. "Journal Citation Report".
<https://clarivate.jp/products/journal-citation-reports/>, (accessed 2018-12-26).
- [43] 山崎茂明. インパクトファクターをめぐる議論: 正しい理解と研究への活かし方. 情報管理. 1998, vol. 41, no. 3, p. 173-182.
- [44] 呑海沙織. 学術情報流通と大学図書館: 学術情報基盤, 電子ジャーナル, オープンアクセス. 図書館界. 2010, vol. 61, no. 5, p. 528-541.
- [45] 上田修一. 学術情報の電子化: 学術情報の電子化は何をもたらしたのか. 情報の科学と技術. 2015, vol. 65, no. 6, p. 238-243.

- [46] "Darwin Correspondence Project". University of Cambridge.
<https://www.darwinproject.ac.uk/letters/darwins-life-letters>, (accessed 2018-12-26).
- [47] 大学図書館コンソーシアム連合事務局. "海外学術雑誌価格の推移". 大学図書館コンソーシアム連合. 2018, <https://www.nii.ac.jp/content/justice/documents>, (accessed 2018-12-25).
- [48] "Open DOAR Statistics". OpenDOAR.
http://v2.sherpa.ac.uk/view/repository_visualisations/1.html, (accessed 2018-12-26).
- [49] "Content sources: By country". BASE. https://www.base-search.net/about/en/about_countries_doc_dn.php?#tab, (accessed 2018-12-26).
- [50] "IRDB コンテンツ分析". 国立情報学研究所.
<https://irdb.nii.ac.jp/analysis/index.html>, (accessed 2018-12-18).
- [51] 国立情報学研究所. "機関リポジトリ公開数とコンテンツ数の推移". 学術機関リポジトリ構築連携支援事業. 2018, <https://www.nii.ac.jp/irp/archive/statistic/>, (accessed 2018-07-18).
- [52] 竹内比呂也. 灰色文献: 大学紀要というメディア: 限りなく透明に近いグレイ?. 情報の科学と技術. 2012, vol. 62, no. 2, p. 72-77.
- [53] 林和弘, 餌取直子, 武内八重子, 中原由美子, 伊勢幸恵, 新岡美咲, 直江千寿子. "日本の機関リポジトリから発信される学術情報の分析: 「三大学図書館連携機関リポジトリ研修」の成果より". 情報プロフェッショナルシンポジウム予稿集. 東京, 2014-12-04/05, 一般社団法人 情報科学技術協会, 2014, p. 155-158.
- [54] "「学術情報リポジトリ」による研究成果の発信". 奈良文化財研究所. 奈文研ニュース, 2015, no.56, p. 7.
- [55] 日本学術会議. 人文・社会系の分野における研究業績評価のあり方について. 2005, <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-19-t1026.pdf>, (accessed 2018-12-18).
- [56] 富本壽子. 大学図書館の機関リポジトリにおける Article コンテンツの検証. 筑波大学, 2009, 修士論文.
- [57] 科学技術・学術審議会・研究計画・評価分科会・情報科学技術委員会・デジタル研究情報基盤ワーキング・グループ. "学術情報の流通基盤の充実について(審議まとめ)". 文部科学省. http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu2/toushin/020401.htm, (accessed 2018-12-18).
- [58] 文部科学省研究振興局情報課. "学術情報発信に向けた大学図書館機能の改善について(報告書)". <https://www.janul.jp/j/documents/mext/kaizen.pdf>, (accessed 2018-12-18).
- [59] 科学技術・学術審議会 学術分科会 研究環境基盤部会・学術情報基盤作業部会. "学術情報基盤の今後のあり方について(報告)". 文部科学省. 2006, http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/toushin/1213896.htm, (accessed 2018-12-18).

- [60] 国立大学図書館協会. "オープンアクセスに関する声明". 文部科学省. 2009,
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/toushin/attach/1283016.htm,
(accessed 2018-12-25).
- [61] 情報科学技術委員会. "第 4 期科学技術基本計画における情報通信分野の重点事項について". 文部科学省.
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu13/siryu/attach/1290981.htm,
(accessed 2018-12-25).
- [62] "第 4 次科学技術基本計画 (平成 23~27 年度)". 内閣府.
<https://www8.cao.go.jp/cstp/kihonkeikaku/index4.html>, (accessed 2018-12-18).
- [63] "平成 21 年度文部科学白書". 文部科学省.
http://www.mext.go.jp/b_menu/hakusho/html/hpab200901/detail/1295628.htm,
(accessed 2018-12-25).
- [64] 科学技術・学術審議会 学術分科会 研究環境基盤部会 学術情報基盤作業部会. "大学図書館の整備について(審議まとめ)-変革する大学にあって求められる大学図書館像-". 文部科学省. http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/toushin/1301602,
(accessed 2018-12-25).
- [65] ジャーナル問題に関する検討会. "大学等におけるジャーナル環境の整備と我が国のジャーナル発信力強化の在り方について". 文部科学省.
http://www.mext.go.jp/b_menu/shingi/chousa/shinkou/034/gaiyou/1351118.htm,
(accessed 2018-12-25).
- [66] "学術機関リポジトリ構築ソフトウェア実装実験プロジェクト". 国立情報学研究所.
<https://www.nii.ac.jp/metadata/irp/>, (accessed 2018-12-18).
- [67] 国立情報学研究所. 学術機関リポジトリ構築ソフトウェア実装実験プロジェクト報告書. 2005, 180p, <https://www.nii.ac.jp/irp/archive/basic/pdf/NII-IRPreport.pdf>, (accessed 2018-12-18).
- [68] "Genii". 国立情報学研究所. <https://ge.nii.ac.jp/genii/jsp/index.jsp>, (accessed 2018-12-18).
- [69] 国立情報学研究所. "【お知らせ】学術コンテンツ・ポータル (GeNii ジーニイ) の終了等サービスの一部見直しについて". 学術コンテンツサービス サポート.
<https://support.nii.ac.jp/ja/news/cinii/20140203>, (accessed 2018-12-18).
- [70] "NII の JAIRO Cloud、スタンフォード大学図書館による「研究図書館イノベーション賞」の功労賞を受賞". Current Awareness Portal. 2014,
<http://current.ndl.go.jp/node/25744>, (accessed 2019-01-08).
- [71] 平成 21 年度学術ポータル担当者研修. "機関リポジトリと著作権. 概論". 2009,
<https://www.nii.ac.jp/hrd/ja/portal/h21/txt9.pdf>, (accessed 2018-12-18).

- [72] 黒澤節男. “機関リポジトリと著作権 Q&A”. 広島大学図書館. 2013,
<https://ir.lib.hiroshima-u.ac.jp/files/public/2/23065/20170823151125629505/Repository-Copyright.pdf>, (accessed 2018-12-18).
- [73] “学術情報リポジトリと著作権”. 静岡大学附属図書館.
http://www.lib.shizuoka.ac.jp/denshi/sure/repo_tyosakuken.pdf, (accessed 2018-12-18).
- [74] 阪口幸治. “機関リポジトリと著作権実務 国内紙を中心に”. 一橋大学附属図書館.
http://www.tulips.tsukuba.ac.jp/pub/SCPJ_WS/H24SCPJ_Copyright_practice.pdf,
(accessed 2018-12-18).
- [75] 引原隆士. “博士論文のオープンアクセス化と研究・教育”.
https://www.nii.ac.jp/irp/event/2013/OA_summit/docs/2_2.pdf, (accessed 2018-12-18).
- [76] Jonathan P. Tennant; François Waldner; Damien C. Jacques; Paola Masuzzo; Lauren B. Collister; Chris. H. J. Hartgerink. The academic, economic and societal impacts of Open Access: an evidence-based review. F1000Research. 2016, vol. 5, no. 632, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4837983>, (accessed 2018-12-18).
- [77] “学位規則の改定概要”. 文部科学省.
http://www.mext.go.jp/component/a_menu/education/detail/__icsFiles/afieldfile/2013/03/13/1331809_03.pdf, (accessed 2018-12-25).
- [78] “著作権法の一部を改正する法律の概要”. 文部科学省.
http://www.mext.go.jp/b_menu/houan/kakutei/detail/__icsFiles/afieldfile/2018/05/21/1405213_01.pdf, (accessed 2018-12-25).
- [79] “学位規則の一部を改正する省令(平成 25 年文部科学省令第 5 号)”. 文部科学省. 2013,
http://www.mext.go.jp/component/a_menu/education/detail/__icsFiles/afieldfile/2013/03/13/1331809_01.pdf, (accessed 2018-07-18).
- [80] “著作権法の一部を改正する法律”. 文部科学省. 2018,
http://www.mext.go.jp/b_menu/houan/kakutei/detail/1405213.htm, (accessed 2018-07-18).
- [81] “著作権法の一部を改正する法律(平成 30 年法律第 30 号)について”. 文化庁.
http://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h30_hokaisei/, (accessed 2018-07-18).
- [82] “Hathi Trust Digital Library”. Hathi Trust Digital Library.
<https://www.hathitrust.org>, (accessed 2018-12-18).
- [83] 逸村裕. “日本における機関リポジトリの展開: 学術情報流通と蓄積の変容”. Current Awareness Portal. 2007, <http://current.ndl.go.jp/ca1626>, (accessed 2018-12-26).
- [84] 国立情報学研究所. “次世代学術コンテンツ基盤共同構築事業”.
<https://www.nii.ac.jp/content/>, (accessed 2018-12-26).

- [85] 栗山正光. “デジタルリポジトリ連合(DRF)の活動を振り返って”. *Current Awareness Portal*. 2017, <http://current.ndl.go.jp/e1908>, (accessed 2018-12-26).
- [86] “DRF(デジタルリポジトリ連合)”. 北海道大学. <http://drf.lib.hokudai.ac.jp/drf/>, (accessed 2018-12-18).
- [87] 前田朗, 加藤寛士, 高橋菜奈子, 山地一禎. システム基盤としての JAIRO Cloud. *大学図書館研究*. 2016, vol. 103, p. 9-15.
- [88] 国立情報学研究所学術コンテンツ課. “筑波大学 JAIRO Cloud 移行実験レポート”. JAIRO Cloud, https://community.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=196&item_no=1&page_id=24&block_id=100, (accessed 2018-07-18).
- [89] 文部科学省. “平成 29 年度学術情報基盤実態調査”. 2018, <https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00400601&tstat=000001015878>, (accessed 2018-07-18).
- [90] 南山泰之. 研究データ管理における機関リポジトリの可能性. *大学図書館研究*. 2016, vol. 103, p. 16-23.
- [91] 大園隼彦, 田口忠祐, 片岡朋子, 林豊, 高橋菜奈子, 南山泰之. JPCOAR スキーマの策定 日本の学術成果の円滑な国際的流通を目指して. *情報管理*. 2018, vol. 60, no. 10, p. 719-729.
- [92] 大場高志. “SPARC Japac: その創設期の頃”. *SPARC Japan Newa Letter*. 2012, no. 12, <https://www.nii.ac.jp/sparc/publications/newsletter/html/12/fa1.html>, (2019-01-07).
- [93] 外国雑誌センター館会議承認. 外国語雑誌センター館のこれまでの活動と今後の課題. 2012, <https://www.janul.jp/ncop/docs/2012sokatsu.pdf>, (2019-01-07).
- [94] 上田修一. 学術情報の電子化: 学術情報の電子化は何をもたらしたのか. *情報の科学と技術*. 2015, no. 65, vol. 6, p. 238-243.
- [95] 守屋文葉. 大学図書館コンソーシアム連合(JUSTICE)とは何か. *看護と情報*. 2013, vol. 20, p. 39-44.
- [96] 田中久徳. 科学技術リテラシーの向上をめぐって-公共政策の社会的合意形成の観点から-. *リファレンス*. 2006, p. 57-83.
- [97] UNESCO. “科学と科学的知識の利用に関する世界宣言”. 文部科学省. 2010, http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/siryo/attach/1298594.htm, (2019-01-08).
- [98] Lynch, Clifford A. *Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age*. *ARL Bimonthly Report* 226, 2003, <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>, (accessed 2018-12-25).

- [99] Crow, Raym. The case for Institutional Repository: A SPARC Position Paper. Scholarly Publishing and Academic Re-sources Coalitions (SPARC), 2002, <http://www.arl.org/bm~doc/instrepo.pdf>, (accessed 2018-12-25).
- [100] Crow, Raym. 機関リポジトリ擁護論 : SPARC 声明書. 栗山正光訳 Scholarly Publishing and Academic Re-sources Coalitions (SPARC), 2002, http://mtkuri.web.fc2.com/case_for_ir_jptr.html, (accessed 2018-12-25).
- [101] "Stanford Prize for Innovation in Research Libraries (SPIRL)". Stanford Libraries. <http://library.stanford.edu/projects/stanford-prize-innovation-research-libraries-spiral>, (accessed 2018-12-25).
- [102] Dame Lynne Brindley, Elisabeth Niggemann, Ann Okerson. "National Institute of Informatics: judges' comments". Stanford Libraries. 2014, <http://library.stanford.edu/projects/stanford-prize-innovation-research-libraries-spiral/2014-prizes/national-institute>, (accessed 2018-12-25).
- [103] 内閣府. “我が国におけるオープンサイエンス推進のあり方について～サイエンスの新たな飛躍の時代の幕開け～”. 国際的動向を踏まえたオープンサイエンスに関する検討会. 2015, <http://www8.cao.go.jp/cstp/sonota/openscience/>, (accessed 2018-07-20).
- [104] 武田英明. "オープンデータ～日本版データリポジトリの可能性～". 国立情報学研究所. 2014, https://www.nii.ac.jp/csi/openforum2014/docs/b-3_takeda.pdf, (accessed 2018-12-25).
- [105] 南山泰之. 研究データにおける機関リポジトリの可能性. 大学図書館研究. 2016, vol. 103, p. 16-23.
- [106] 丸山修一. "オープンサイエンスに関する諸施策について". 文部科学省. 2017, https://japanlinkcenter.org/rduf/doc/170626_03_mext.pdf, (accessed 2018-12-25).
- [107] "Canadian Association of Research Libraries". Canadian Association of Research Libraries. <http://www.carl-abrc.ca>, (accessed 2018-12-26).
- [108] "Open Archives Initiative Protocol for Metadata Harvesting". Open Archives Initiative Protocol for Metadata Harvesting. <https://www.openarchives.org/pmh/>, (accessed 2018-12-26).
- [109] Jordan, Mark. The CARL metadata harvester and search service. Library Hi Tech. 2016, vol. 24, no. 2, p. 197-210.
- [110] "Deutsche Initiative für Netzwerkinformation(DINI)". Deutsche Initiative für Netzwerkinformation. <https://dini.de>, (accessed 2018-12-26).
- [111] "COUNTER CONSISTENT CREDIBLE COMPARABLE". COUNTER CONSISTENT CREDIBLE COMPARABLE. <https://www.projectcounter.org/about/>, (accessed 2018-12-26).

- [112] Dobratz, Susanne; Scholze, Frank. DINI institutional repository certification and beyond. *Library Hi Tech*. 2006, vol. 24, no. 4, p. 583-594.
- [113] "The Hong Kong University of Science and Technology". The Hong Kong University of Science and Technology. <https://www.ust.hk>, (accessed 2018-12-26).
- [114] Ki-Tat Lam; Diana L.H. Chan. Building an institutional repository: sharing experiences at the HKUST Library. *OCLC Systems & Services*. 2007, vol. 23, no. 3, p. 310-323.
- [114] "National Coronal Information System". National Coronal Information System. <http://www.ncis.org.au>, (accessed 2018-12-26).
- [115] "ePrint@IISc". ePrint@IISc. <http://eprints.iisc.ernet.in>, (accessed 2018-12-26).
- [116] "SUSHI". COUNTER CONSISTENT CREDIBLE COMPARABLE. <https://www.projectcounter.org/code-of-practice-sections/sushi/>, (accessed 2018-12-26).
- [117] Jayakanth, Francis; Filbert Minj; Silva, Usha; Jagirdar, Sandhya. ePrints@IISc: India's first and fastest growing institutional repository. *OCLC Systems & Services*. 2008, vol. 24, no. 1, p. 59-70.
- [118] "SCHOLARLY Stats Helping Libraries Streaming the Title Selection Process". SCHOLARLY Stats Helping Libraries Streaming the Title Selection Process. <https://www.mpsscholarlystats.com>, (accessed 2018-12-26).
- [119] "360 Counter". PROQUEST. <https://www.proquest.com/products-services/360-Counter.html>, (accessed 2018-12-26).
- [120] "Scholarly iQ". Scholarly iQ. <http://www.scholarlyiq.com>, (accessed 2018-12-26).
- [121] Xu, Fei. Implementation of an electronic resource assessment system in an academic library. *Program*. 2010, vol. 44, no. 4, p. 374-392.
- [122] "The National electronic Library of Infection". The National electronic Library of Infection. http://www.neli.org.uk/IntegratedCRD.nsf/NeLI_Home1?OpenForm, (accessed 2018-12-26).
- [123] Jordan, Mark. The CARL metadata harvester and search service. *Library Hi Tech*. 2006, vol. 24, no. 2, p. 197-210.
- [124] "Strathprints: The University of Strathclyde institutional repository". University of Strathclyde, <https://strathprints.strath.ac.uk>, (accessed 2018-12-26).
- [125] "Web of Science". Web of Science. <http://apps.webofknowledge.com/>, (accessed 2018-12-26).
- [126] "What does fractional count (FC) mean?". nature INDEX. <http://feedback.natureindex.com/knowledgebase/articles/423525-what-does-fractional-count-fc-mean>, (accessed 2018-12-26).

- [127] Bonilla-Calero, A I. Scientometric Analysis of a Sample of Physics-Related Research Output Held in the Institutional Repository Strathprints (2000-2005). *Library Review*. 2008, vol. 57, no. 9, p. 700-721.
- [128] "Journal Usage Statistics Portal (JUSP)". Jisc. <https://www.jisc.ac.uk/journal-usage-statistics-portal>, (accessed 2018-12-26).
- [129] "IRStats". iplus. <http://iplus.ukoln.ac.uk/technology/irstats.html>, (accessed 2018-12-26).
- [130] "IRStats". eprint repository software. <http://wiki.eprints.org/w/IRStats>, (accessed 2018-12-26).
- [131] "What is AWStats". AWStats official web site. <https://awstats.sourceforge.io>, (accessed 2018-12-26).
- [132] "Access Statistics for Participating RePEc Services". Access Statistics for Participating RePEc Services. <https://logec.repec.org>, (accessed 2018-12-26).
- [133] "About /robots.txt". The Web Robots Pages. <http://www.robotstxt.org/robotstxt.html>, (accessed 2018-12-26).
- [134] AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES. "Project Report of the Benchmark Statistics Service (BEST)". p. 1-7, https://openresearch-repository.anu.edu.au/bitstream/1885/46638/5/final_report.pdf, (accessed 2018-12-26).
- [135] Johan Bollen. MESUR: A Survey of Usage-based Scholarly Impact Metrics. *Against the Grain*. 2010, vol. 22, no. 4, p. 20-26.
- [136] "MESUR is now operating at Indiana University – School of Informatics and Computing". MESUR studying science from large-scale usage data. <http://mesur.informatics.indiana.edu/?p=1>, (accessed 2018-12-26).
- [137] "PIRUS2". The PIRUS2 Project. <http://www.cranfieldlibrary.cranfield.ac.uk/pirus2/tiki-index.php>, (accessed 2018-12-26).
- [138] "WELCOME TO THE IFABC". ifabc. <http://www.ifabc.org>, (accessed 2018-12-26).
- [139] "The DOI® System". doi. <https://www.doi.org>, (accessed 2018-12-26).
- [140] "Factsheet DOI® System and OpenURL". doi. https://www.doi.org/factsheets/DOI_OpenURL.html, (accessed 2018-12-26).
- [141] Merk, Christine; Scholze, Frank; Windisch, Nils. Item-level usage statistics: A review of current practices and recommendations for normalization and exchange. *Library Hi Tech*. 2009, vol. 27, no. 1, p. 151-162.
- [142] Clarivate Analytics. "This essay was originally published in the Current Contents print editions June 20, 1994, when Clarivate Analytics was known as The Institute for Scientific Information® (ISI®)". The Clarivate Analytics Impact Factor. <https://clarivate.com/essays/impact-factor/>, (accessed 2018-12-28).

- [143] Mike Thelwall. A comparison of sources of links for academic Web impact factor calculations. *Journal of Documentation*. 2002, vol. 58, no. 1, p. 66-78.
- [144] "What does Altmetric do?". Altmetrics. <https://www.altmetric.com>, (accessed 2018-12-26).
- [145] "Ceek.jp Altmetrics (α ver.)". Ceek.jp Altmetrics. <http://altmetrics.ceek.jp/>, (accessed 2018-01-04).
- [146] Janet Dagenais Brown. Citation searching for tenure and promotion: an overview of issues and tools. *Reference Services Review*. 2014, vol. 42, no. 1, p. 70-89.
- [147] "北海道大学学術成果コレクション". 北海道大学附属図書館. <https://eprints.lib.hokudai.ac.jp/dspace/index.jsp>, (accessed 2018-12-26).
- [148] "京都大学学術機関リポジトリ 紅". 京都大学. <https://repository.kulib.kyoto-u.ac.jp/dspace/>, (accessed 2018-12-26).
- [149] "Science Citation Index". Clarivate Analytics. http://mjl.clarivate.com/scope/scope_sci/, (accessed 2018-12-26).
- [150] 佐藤翔, 永井裕子, 古賀崇, 三隅健一, 逸村裕. 機関リポジトリへの登録が論文の被引用数と電子ジャーナルアクセス数に与える影響. *情報知識学会*. 2011, vol. 21, no. 3, p. 383-402.
- [151] "つくばリポジトリ". つくばリポジトリ. <https://tsukuba.repo.nii.ac.jp>, (accessed 2018-12-26).
- [152] "HERMES-IR (一橋大学機関リポジトリ) によろこそ". HERMES-IR. <http://hermes-ir.lib.hit-u.ac.jp/ir/index.html>, (accessed 2018-12-26).
- [153] "ARRIDE". IDE-JETRO. <https://ir.ide.go.jp>, (accessed 2018-12-26).
- [154] 佐藤翔. コンテンツ入手元として機関リポジトリが果たしている役割. 筑波大学. 2013. 博士論文.
- [155] "九州大学学術情報リポジトリ (QIR)". 九州大学附属図書館. <https://www.lib.kyushu-u.ac.jp/ja/collections/qir>, (accessed 2018-12-26).
- [156] 馬場謙介. リポジトリのアクセスログ解析. 九州大学附属図書館研究開発室年報. 2011, vol. 2010/2011, p. 5-8.
- [157] 木下仁, 馬場謙介, 伊東栄典, 廣川佐千男. 大学機関リポジトリのアクセスログに現れる検索語と論文題名の関係について. *情報処理学会報告研究*. 2011, p. 1-8, https://catalog.lib.kyushu-u.ac.jp/opac_download_md/19684/B-3-2-id049.pdf, (accessed 2018-12-26).
- [158] "Japan Link Center". Japan Link Center. <https://japanlinkcenter.org/top/>, (accessed 2018-12-26).
- [159] 吉川次郎. DOI リンクがウェブ上の学術情報流通に果たす役割 : Wikipedia を対象に. 筑波大学, 2015, 修士論文. [ja/news/cinii/20150611](http://news.cinii/20150611), (accessed 2018-07-18).

- [160] "筑波大学附属図書館". 筑波大学附属図書館. <https://www.tulips.tsukuba.ac.jp/lib/>, (accessed 2018-07-18).
- [161] 高久雅生, 小幡将司, 江草由佳. OPAC 利用ログに基づく文献検索システムの試作と評価. 知識情報学会. 2018, vol. 28, no. 2, p. 111-120.
- [162] "About DSpace". DSpace. <https://duraspace.org/dspace/about/>, (accessed 2018-12-26).
- [163] "JAIRO Cloud トップページ". JAIRO Cloud. <https://community.repo.nii.ac.jp>, (accessed 2018-12-26).
- [164] "JAIRO Cloud 移行 FAQ". JPCOAR. <https://jpcoar.repo.nii.ac.jp/>, (accessed 2018-12-26).
- [165] 国立情報学研究所. "機関リポジトリ公開数とコンテンツ数の推移". 学術機関リポジトリ構築連携支援事業. 2018, <https://www.nii.ac.jp/irp/archive/statistic/>, (accessed 2018-07-18).
- [166] 佐藤翔. 機関リポジトリ収録コンテンツにおける利用数とアクセス元、アクセス方法、コンテンツ属性の関係. 三田図書館・情報学会研究大会発表論文集. 2009, p. 9-12.
- [167] 稲永晶子. "IRDB から見る機関リポジトリの現状 および周辺の問題の整理について". https://www.nii.ac.jp/hrd/ja/jitsumu/h24/h24-2_seika.pdf, (accessed 2018-12-26).
- [168] "IRDB コンテンツ分析". 国立情報学研究所. <https://irdb.nii.ac.jp/>, (accessed 2018-07-18).
- [169] "IRDB". 国立情報学研究所. <https://dev.irdb.nii.ac.jp/>, (accessed 2019-01-04).
- [170] 真中考行. つくばリポジトリへの JAIRO Cloud への移行. Current Awareness Portal. 2014, <http://current.ndl.go.jp/e1574>, (accessed 2018-12-26).
- [171] 国立情報学研究所 学術コンテンツ課. "平成 25 年度 筑波大学 JAIRO Cloud 移行実験レポート". JAIRO Cloud. 2014, https://community.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=196&item_no=1&page_id=24&block_id=100, (accessed 2018-12-26).
- [172] 千葉大学. 機関リポジトリのアウトプット評価 プロジェクト最終報告書. 2013, https://www.ll.chiba-u.jp/roat/document/ROAT_Final_Report_J.pdf, (accessed 2018-12-26).
- [173] 総務省. "デジタルデータの長期保存・利用について". 震災関連デジタルアーカイブ構築・運用のためのガイドライン. http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/02ryutsu02_03000114.html, (accessed 2018-12-26).
- [174] "キャンパスネットワークの管理". 筑波大学情報環境機構 学術情報メディアセンター. https://www.cc.tsukuba.ac.jp/wp/service/etc_network/, (accessed 2018-12-28).

- [175] 佐藤義則, 竹内比呂也, 武内八重子, 竹内茉莉子機関リポジトリのアウトプット評価プロジェクト最終報告書. 千葉大学. 2013, 33p.
- [176] 佐藤翔. 機関リポジトリにおける利用数とアクセス元、アクセス方法、コンテンツ属性の関係. 三田・図書館情報学会研究大会発表論文集. 三田図書館情報学会, 2009, p. 9-12.
- [177] "Referrer Policy". W3C. 2017, <https://www.w3.org/TR/referrer-policy/#integration-with-fetch>, (accessed 2018-07-18).
- [178] "マスコミ報道だけでは分からない岡崎図書館事件". Librahack. <http://librahack.jp>, (accessed 2018-07-18).
- [179] 安形輝. 図書館ウェブサイトの公開性-クローラに対するアクセス制御に関する調査-. 2010, http://www.mslis.jp/am2010yoko/08_agata.pdf, (accessed 2018-07-18).
- [180] 国立国会図書館. "国立国会図書館法によるインターネット資料の収集について", 2016, http://warp.da.ndl.go.jp/bulk_info.pdf, (accessed 2018-12-28).
- [181] "筑波大学附属図書館の robots.txt". 筑波大学附属図書館. <http://www.tulips.tsukuba.ac.jp/robots.txt>, (accessed 2018-12-26).
- [182] "Repository Module on NC2 WEKO". WEKO. <http://weko.at.nii.ac.jp>, (accessed 2018-07-18).
- [183] "WEKO 管理". JAIRO Cloud. <https://community.repo.nii.ac.jp/weko/help/ja/html/setting.html>, (accessed 2018-07-18).
- [184] "WEKO メタデータ取得 API-OAI-PMH". WEKO. <https://weko.at.nii.ac.jp/demo/weko/help/ja/html/api/WK03-01.html>, (accessed 2018-12-26).
- [185] "Corporation for National Research Initiatives (CNRI)". Corporation for National Research Initiatives. <https://www.cnri.reston.va.us>, (accessed 2018-12-26).
- [186] "Factsheet DOI® System and Internet Identifier Specifications". doi. <https://www.doi.org/factsheets/DOIIdentifierSpecs.html>, (accessed 2018-12-26).
- [187] "Persistent identifiers: awareness level". ands. <https://www.ands.org.au/guides/persistent-identifiers-awareness>, (accessed 2018-12-26).
- [188] "DSpace システム説明書: 機能概要". 国立情報学研究所. https://www.nii.ac.jp/metadata/irp/dspace-docs-jp_1_3/functional.html, (accessed 2018-01-04).
- [189] 永井一樹. "永続性と可視性をカッコに入れる". <https://www.nii.ac.jp/hrd/ja/portal/h22/txt10.pdf>, (accessed 2018-12-26).
- [190] "「OAI-PMH と図書館サービス」". 国立情報学研究所. https://www.nii.ac.jp/metadata/oai-pmh/oai-pmh_chiba.html, (accessed 2018-12-26).

- [191] 国立国会図書館. "OAI-PMH の要点". 2015, http://iss.ndl.go.jp/somoku/wp-content/uploads/2013/08/h27kensyu_sanko01.pdf, (accessed 2018-12-26).
- [192] "junii2 レスポンスフォーマット一覧". WEKO.
<https://weko.at.nii.ac.jp/demo/weko/help/ja/html/api/WK03-01.html>, (accessed 2018-12-26).
- [193] 佐藤義則. 機関リポジトリの利用統計のゆくえ. Current Awareness Portal. 2008, <http://current.ndl.go.jp/ca1666>, (accessed 2018-12-26).
- [194] "Google Analytics". Google.
<https://analytics.google.com/analytics/web/provision/?authuser=0#/provision>, (accessed 2018-12-26).
- [195] "What is it? The Webalizer". The Webalizer. <http://www.webalizer.org>, (accessed 2018-12-26).
- [196] "visitors". visitors. <http://www.hping.org/visitors/>, (accessed 2018-12-26).
- [197] 岸田和明. 図書館の評価: 利用統計を用いた蔵相評価の手法. 情報の科学と技術. 1994, vol. 44, no. 6, p. 300-305.
- [198] Virgil Diodato. "平均反応時間". 軽量書誌学用語辞典. 芳鐘冬樹. 岸田和明. 小野寺夏生訳. 日本図書館協会. 2008. P. 161-162.
- [199] Oracle. "Oracle Text で提供されるストップリスト". Oracle Text リファレンス.
https://docs.oracle.com/cd/E16338_01/text.112/b61357/astopsup.htm, (accessed 2018-12-26).
- [200] "WebUIP 利用マニュアル". 国立情報学研究所. https://www.nii.ac.jp/CAT-ILL/MAN2/WebUIP/TOSHO1/furoku_a.html, (accessed 2018-12-26).
- [201] "gTLD とは". 日本ネットワークインフォメーションセンター.
<https://www.nic.ad.jp/ja/basics/terms/gtld.html>, (accessed 2019-01-04).
- [202] "ドメイン名の種類". 日本ネットワークインフォメーションセンター.
<https://www.nic.ad.jp/ja/dom/types.html#gtld>, (accessed 2019-01-04).
- [203] "科学技術指標". 科学技術・学術政策研究所. 2018,
<http://www.nistep.go.jp/research/science-and-technology-indicators-and-scientometrics/indicators>, (accessed 2019-01-04).
- [204] 国立情報学研究所. "機関リポジトリの現状(各国比較)". 文部科学省.
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/040/shiryo/__icsFiles/afieldfile/2017/06/20/1386600_003.pdf, (accessed 2019-01-07).
- [205] "HathiTrust へのデジタル化資料の登載について". 慶應義塾大学メディアセンター (図書館). <http://www.lib.keio.ac.jp/jp/hathitrust/>, (accessed 2019-01-04).

- [206] Nick Shockey; Heather Joseph; Melissa Hagemann. “BOAI 15 Survey Report”. LIS Scholarship Archive Works. 2018, <https://osf.io/preprints/lissa/znf2w/>, (accessed 2019-01-04).
- [207] “Scientific Reports 誌が出版論文数で PLOS ONE を逆転（記事紹介）”. Current Awareness Portal. 2017, <http://current.ndl.go.jp/node/33827>, (accessed 2019-01-04).
- [208] ”Who We Are Openness Inspires Innovation”. PLOS. <https://www.plos.org/who-we-are>, (accessed 2019-01-04).
- [209] “Scientific Reports”. Nature. <https://www.nature.com/srep/>, (accessed 2019-01-04).
- [210] 国立情報学研究所. “CiNii Dissertations について”. NII 学術コンテンツサービスサポート. https://support.nii.ac.jp/ja/cid/cinii_dissertations, (2019-01-04).
- [211] ”メタデータ・フォーマット junii2”. 学術機関リポジトリ構築連携支援事業. <https://www.nii.ac.jp/irp/archive/system/junii2.html>, (2019-01-07).
- [212] ”研究支援(機関リポジトリ)”. 文部科学省. http://www.mext.go.jp/component/b_menu/other/__icsFiles/afieldfile/2011/12/20/1314399_03.pdf, (accessed 2019-01-07).
- [213] “オープンサイエンス時代の次世代リポジトリソフト開発に着手”. 国立研究開発法人物質・材料機構. <https://www.nims.go.jp/news/press/2017/11/201711070.html>, (accessed 2019-01-08).
- [214] Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group. ”Next Generation Repositories”. 2017, <https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf>, (accessed 2019-01-08)
- [215] 林正治. “次世代リポジトリの機能要件および技術勧告”. Current Awareness Portal. <http://current.ndl.go.jp/e2011>, (accessed 2019-01-08)
- [216] 科学技術・学術審議会 総合政策特別委員会. “オープンサイエンスの推進について”. 文部科学省. http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu22/siryo/__icsFiles/afieldfile/2016/12/08/1380241_04.pdf, (accessed 2019-01-08)