

## J-STAGE を活用した日本の学術論文データの整備

久保 琢也<sup>1</sup>・伊藤 広幸<sup>2</sup>

**概要**：研究活動に関する IR や評価業務において、日本の学術論文のデータを整備することは特に国立大学における課題の1つとなっている。このような背景において、信州大学では、2020年より科学技術振興機構（JST）と契約を結び、J-STAGE を活用した日本の学術論文データの整備を行なっている。本発表では、J-STAGE を活用するに至った経緯や方法、注意点等について報告する。

**キーワード**：J-STAGE、Web スクレイピング、研究評価、成果を中心とする実績状況に基づく配分

### 1. はじめに

IR や評価業務において研究活動を扱う場合、その中心的な調査対象としては学術論文が挙げられる。例えば、研究 IR では各研究機関の置かれた状況や目標、戦略によって様々であろうが、毎年発表された学術論文の数やその経年変化の把握、研究機関間の比較等が行われる。また、評価業務においては、近年、国立大学では運営費交付金の算定に学術論文数が影響するため、そのためのデータ収集や体制整備が課題となっている。

一般的に、学術論文に関する調査には Clarivate Analytics 社の Web of Science や Elsevier 社の Scopus といった国際学術文献データベースが用いられることが多い。これらのデータベースは採録する媒体（雑誌、学会紀要等）の質を維持しつつも網羅性に優れており、世界の学術論文の書誌情報や被引用数等が参照可能となっている。加えて、これらのデータベースにはそれぞれ対応する分析ツールが用意されており（InCites、SciVal）、各雑誌や論文について、様々な計量書誌学的指標を用いた分析が可能となっている。

その一方で、上記のデータベースには日本の学術雑誌の多くが採録されておらず、当然ながら、これらの学術雑誌に掲載される学術論文の動向について把握することはできない。このことは、特に、日本の学術論文において研究成果を公表することが多い人文社会系の評価において課題視されており、上記のデータベースのみに依拠することの限界も指摘されてきた（林, 2018）。

それでは、どのように日本の学術雑誌に掲載される学術論文のデータを取得・整備すれば良いのだろうか。本稿では解決策の1つとして J-STAGE を活用した事例を報告する。J-STAGE とは国立研究開発法人科学技術振興機構（JST）が運営する電子ジャーナルプラットフォームであり、学協会や研究機関等における科学技術刊行物を収録・公開している（<https://www.jstage.jst.go.jp>）。以下では本取り組みの背景である国立大学運営費交付金

<sup>1</sup> 横浜国立大学 研究推進機構 特任教員（講師） メール：kubo-takuya-xv@ynu.ac.jp（投稿時は信州大学 学術研究・産学官連携推進機構に所属）

<sup>2</sup> 信州大学 学術研究・産学官連携推進機構 助教 メール：hito@shinshu-u.ac.jp

「成果を中心とする実績状況に基づく配分」による評価を取り上げ、J-STAGE を利用するに至った経緯、利用するためのプロセス、及び実際のデータ取得方法、並びに、その注意点について報告する<sup>3</sup>。

## 2. 取り組みの背景

### 2.1. 国立大学運営費交付金「成果を中心とする実績状況に基づく配分」について

現在、日本の国立大学は毎年度、または定期的に評価が行われ、それが運営費交付金の算定に影響する仕組みとなっている。例えば、国立大学法人評価は年度評価と6年毎の中期目標期間の評価からなり、評価結果は次期中期目標・計画の内容や運営費交付金の算定等に用いられている。また、平成28年度からは第3期中期目標期間の開始とともに運営費交付金に「3つの重点支援の枠組み」が導入され、毎年度、評価結果によって運営費交付金の一部が傾斜配分されるようになった。そして、令和元年度からは、新たな枠組みとして「成果を中心とする実績状況に基づく配分」（以下、「実績状況に基づく配分」）が導入され、さらに傾斜配分される運営費交付金の予算額が大幅に増大した。

「実績状況に基づく配分」の主旨は、「多額の公費により支えられている国立大学等に対して、厳格な評価とそれに基づく資源配分が求められていることから、国立大学等の法人化のメリットを活かした各大学の成果や実績を相対的に評価することで、一層の経営改革を推進する」こととされている（文部科学省, 2019）。予算規模については、令和元年度は運営費交付金の中の約700億円だったのに対して、令和2年度は850億円となっている（文部科学省, 2020）。令和2年度の運営費交付金総額が1兆971億円であるため（財務省, 2020）、「実績状況に基づく配分」は運営費交付金の1割弱を占めていることになる。また、傾斜配分率も令和元年度は基準額±10%であったのに対して、令和2年度には基準額±15%となっており、大学経営に与える影響はますます大きくなっている（文部科学省, 2020）。

「実績状況に基づく配分」における評価項目は令和元年度に6項目であったが、令和2年度からは教育、研究に関する項目を拡充した13項目からなる評価が行われるようになっている。本稿と関係するのはこれらの中でも「常勤教員当たり研究業績数」であり、査読つき論文や学術図書、作品等の発表数が指標となっている。集計には国立大学法人評価の中で、大学改革支援・学位授与機構が実施する教育研究に関する評価の「現況分析」における評価単位である「学系」が採用されている。この学系は、対象となる教員が在籍する組織によって決まっており、各教員は「人文科学系」、「社会科学系」、「理学系」、「工学系」、「農学系」、「保健系」、「教育系」、「総合文系」、「総合理系」、「総合融合系」、「その他」に分類される。集計方法としては、学系ごとに論文数がカウントされるが、所属機関の中で学系の異なる複数の著者が含まれる論文の場合、それぞれの学系において上限を「1」とカウントされる。つまり、1つの論文に同じ学系の著者が複数いたとしても、その学系では「1」としてカウントされる。上記の集計方法は単純であるが、学術論文についてこ

<sup>3</sup> 本稿は令和3年3月22日に開催された大学評価コンソーシアム「継続的改善のためのIR/IEセミナー2021 R1：IR実務担当者セッション」において発表した内容を論文化したものである。

のような集計を行うためには、それぞれの学術論文について、各著者の所属機関情報が必要となる。しかしながら、以下で述べる通り、必ずしもこれは容易なことではない。

なお、「実績状況に基づく配分」の政策意図や導入の経緯、評価方法の問題点等については竹内(2019)、中村(2020)で詳しく述べられているため本稿では扱わない。本稿は現行の制度の中、IR・評価担当の実務者として実施した取り組みを報告するものである。

## 2.2. 信州大学における研究IR体制と課題

筆者らが行なった方法を説明する前に、信州大学における研究IR体制について簡単に述べておきたい。信州大学では、IR業務や評価業務のためにインスティテューショナル・リサーチ室(以下、IR室)を設置している。主な業務は学内外のデータの収集、分析及び調査研究を行うことによって法人としての意思決定を支援しているほか、国立大学法人評価や運営費交付金に関係する評価のための資料作成等も行っている。

信州大学のIR室はその役割によって「教務」チーム、「点検・評価」チーム、「研究、産学官・社会連携」チームから構成される。研究IRを担当するのは「研究、産学官・社会連携チーム」であるが、同チームの主な業務は信州大学の研究者による学術論文の出版状況や外部資金の獲得状況に関するデータ収集や調査、各種の世界大学ランキング(THES, QS, ARWU, etc.)の動向調査である。また、同チームでは文部科学省等による評価における研究活動に関するデータ整理等も行っており、「実績状況に基づく配分」の中でも、研究業績に関係する資料作成は「研究、産学官・社会連携」チームが担当している。

筆者らは令和2年度から「実績状況に基づく配分」に携わってきたが、そこでの課題としては、限られたマンパワーや時間の中で全研究者の研究業績を網羅的に収集、整理しなければならなかったことが挙げられる。筆者らの本来の所属は学術研究・産学官連携推進機構であり、IR室の業務の他にリサーチ・アドミニストレーション業務にも従事している。そのため、筆者らが研究IRや評価業務に費やすことができるエフォートは限られており、当該業務に限らず業務の効率化が求められている。

また、信州大学では独自の研究業績データベースを構築しているものの、これだけでは「実績状況に基づく配分」への活用には限界があったため、データの取得から検討しなければならない状況にあった。信州大学では所属する研究者の最新の学術情報環境を整備すると共に、研究成果・研究活動を広く国内外に発信するために信州大学学術情報オンラインシステム(Shinshu University Online System of General Academic Resources, 以下SOAR)を構築・運用している。SOARではCiNii ArticlesやWeb of Scienceといった外部のデータベースから研究業績情報のインポートが可能であるものの、基本的にデータの更新は研究者に依存しているため、データの網羅性という点で疑問が残る。また、研究者自身による手入力も行われていることから、表記揺れ等によるデータのクリーニングに多大な時間と労力が必要となることが予見された。例えば、学内の複数の研究者が著者である論文であっても、雑誌名や論文のタイトル等の表記が研究者によって異なるケースが散見されたためである。この場合には、何らかの形で同一の論文の共著者であることが判別できないと「実績状況に基づく配分」で正確な集計ができないため、特に注意が必要である。

### 2.3. 国内学術論文の抽出における課題

「はじめに」で述べたとおり、国際学術雑誌に掲載された論文の書誌情報は、Web of Science や Scopus といったデータベースによって比較的容易に取得可能である。例えば、Web of Science からダウンロード可能なデータでは、若干のデータ操作が必要であるものの、著者名と所属機関情報が紐付いていることから、複数の著者を有する論文であっても、誰が信州大学に所属している研究者であるのかを判別することができる。また、論文の著者情報と学内の研究者情報とを結びつけることにより、学生の著者を取り除いた上で部局単位での集計も可能である。その一方で、日本国内の学術雑誌に掲載された論文については公的なデータベースの整備が充実しているものの、以下の通り研究 IR や評価業務には活用しづらいのが現状である。

CiNii Articles (<https://ci.nii.ac.jp/ja>) は国立情報学研究所が運営する学術論文に関するデータベースであり、日本の学協会誌や研究紀要等のおよそ 2063 万件を超える膨大な論文データを収録している。また、CiNii Articles では Web サービスとして検索した論文集合の書誌情報をダウンロードすることができるため、ある程度、研究 IR のためのデータ分析が可能にはなっている。その一方で、CiNii Articles からダウンロードできる情報には著者の所属情報が含まれておらず、複数の著者がいる場合には、誰が所属機関の研究者であるのか判別がつかない。なお、令和 3 年 4 月より、文献だけでなく機関リポジトリ等の研究データ等も横断検索できる CiNii Research が本公開されているが (<https://cir.nii.ac.jp/ja>)、本稿執筆時は検索した論文集合をダウンロードする機能は提供されていない。

一方、これらのデータベースでは API (Application Programming Interface) も提供されていることから、API を通じたデータの取得にも触れておきたい。まず、CiNii Articles の論文検索の OpenSearch では、著者名や著者の所属機関、論文タイトルや雑誌名等を用いて、該当する論文の書誌情報を一度に取得可能である。しかしながら、この方法では、Web ブラウザを用いてダウンロードする場合と同様に、書誌情報に著者の所属機関が含まれておらず、誰がどの研究機関の研究者であるのか判別することはできない。他にも、CiNii Articles と CiNii Research の API では、各論文のメタデータを RDF/XML 形式で取得できるようになっている。この場合は、各論文の ID を用いて一報一報データを取得していくことになるが、著者の所属機関情報も含んでいる点で期待が持てそうである。しかしながら、そもそもこれらのデータベースでは、目次や巻頭言といった論文以外のデータが含まれていたり、査読の有無の手がかりが得られなかったりする点で「実績状況に基づく配分」で利用するにしてもデータ取得後の確認作業に多くの労力を要することが予見される。

他にも、「はじめに」で述べたように、日本の学術論文に関する公的なデータベースとしては J-STAGE が挙げられる。J-STAGE は科学技術振興機構 (以下、JST) が運営する電子ジャーナルプラットフォームであり、日本の学協会や会議の刊行物を収録している。J-STAGE は Web ブラウザ上で論文が閲覧できる他、雑誌の査読の有無や文献種別、分野を選択した上で文献検索が行える点で検索機能に優れている。その一方で、J-STAGE では検索した論文の書誌情報をダウンロードする機能を提供していない。また、J-STAGE も API を提供しているものの、取得できる情報には CiNii Articles と同様に著者の所属機

関の情報は含まれていない。更に、J-STAGEのWebAPI利用規約では、APIで取得したデータをそのままの状態でも24時間以上保持することが禁止されていることから、実質的に研究IRや評価業務での利用は難しい。

このように、既存の公開サービスや学内のリソースだけでは国内の学術雑誌に掲載される論文データの収集・整備には課題が多いのが現状であった。そのため、研究IRや各種の評価業務のためにも、より効率的、かつ網羅的に論文データを収集できる方法の確立が求められていた。

### 3. J-STAGEによる論文データの取得

上記の背景を踏まえ、筆者らは改めてJ-STAGEに着目した。図1はJ-STAGEの書誌画面の例であるが、ここでは論文のタイトルや著者情報、キーワードやDOI、また、文献の種別(例:事例報告)等の情報が整備されていることがわかる。これらの情報を収集すれば、「実績状況に基づく配分」のための論文データを整備できそうであるが、手作業で収集しては、多大な労力を要することは想像に難くない。

一方、図2は図1のソースコードの一部であるが、ここでは、論文の書誌情報や著者情報が規則的に記述されていることが分かる。そこで、筆者らはWebスクレイピングによって機械的にこれらの情報を取得することを検討した。Webスクレイピングとは、Webサイトから特定の情報を抽出するための技術であり、PythonやRといったプログラミング言語やオンラインのサービス等により実行可能である。

Webスクレイピングを行うことのメリットとしては、機械的、効率的な情報収集が可能であるため、例えば対象とする論文数が多かったとしても短時間での情報の取得が期待できる。また、手入力の際に生じる見間違いや入力時のミスといったエラーを防ぐことができるというのもWebスクレイピングの大きなメリットである。

The image shows a screenshot of a J-STAGE article page. The header is dark blue with the text '情報の科学と技術'. Below the header is a navigation bar with links: '資料トップ', '巻号一覧', 'おすすめ記事', and 'この資料について'. The main content area has a breadcrumb trail: 'J-STAGEトップ / 情報の科学と技術 / 71巻(2021)2号 / 書誌'. The article title is '事例報告 研究力分析の効率化・高度化に関するCode for Research Administrationの取り組み: URAによる機関を越えた連携'. The author information is '平井 克之, 岡崎 麻紀子, 奥津 佐恵子, 久保 琢也, 矢吹 命大, 渡邊 優香'. There is a '+ 著者情報' button. The keywords are 'リサーチ・アドミニストレーター (URA), 研究力分析, プログラミング, コーディング, シビックテック'. There are 'ジャーナル' and 'フリー' buttons. The page information is '2021年71巻2号 p. 80-86'. The DOI is 'https://doi.org/10.18919/jkg.71.2\_80'.

図1. J-STAGEにおける各論文のページ例

```

<meta name="journal_title" content="情報の科学と技術">
<meta name="journal_abbrev" content="情報の科学と技術">
<meta name="publisher" content="一般社団法人 情報科学技術協会">
<meta name="authors" content="平井 克之">
<meta name="authors_institutions" content="新潟大学研究企画室">
<meta name="authors_orcids" content="0000-0003-1657-0035">
<meta name="authors" content="岡崎 麻紀子">
<meta name="authors_institutions" content="京都大学学術研究支援室">
<meta name="authors_orcids" content="0000-0002-0507-1372">
<meta name="authors" content="奥津 佐恵子">
<meta name="authors_institutions" content="名古屋市立大学産学官共創イノベーションセンター">
<meta name="authors_orcids" content="0000-0002-1475-1898">
<meta name="authors" content="久保 琢也">
<meta name="authors_institutions" content="信州大学学術研究・産学官連携推進機構">
<meta name="authors_orcids" content="0000-0002-6219-3835">
<meta name="authors" content="矢吹 命大">
<meta name="authors_institutions" content="横浜国立大学大学戦略情報分析室">
<meta name="authors_orcids" content="0000-0001-9008-7100">
<meta name="authors" content="渡邊 優香">
<meta name="authors_institutions" content="九州大学学術研究・産学官連携本部">
<meta name="authors_orcids" content="0000-0002-3207-0041">
<meta name="title" content="研究力分析の効率化・高度化に関するCode for Research Administrationの取組み：URAによる機関を越えた連携">
<meta name="publication_date" content="2021/02/01">
<meta name="online_date" content="2021/02/01">
<meta name="volume" content="71">

```

図 2. 図 1 のソースコードの一部

しかしながら、J-STAGE の利用規約上、機械的な大量のデータ取得は禁止されており、サーバに負荷がかかる規模での Web スクレイピングを行うことはできない点に注意が必要である。このことから、筆者らは 2020 年の 6 月より科学技術振興機構と協議を開始し、データ利用契約書を交わすことにより、J-STAGE 上の論文の書誌情報や著者情報を Web スクレイピングによって取得する許可を得た。

作業の手順としては、まず、J-STAGE が提供する API を通じて、対象となる期間における信州大学の研究者を著者に含む論文の書誌画面のリンクを取得した。次に、J-STAGE では査読付き雑誌とそうでない雑誌のカテゴリ分けがなされているため、査読つき雑誌に掲載された論文の書誌画面のリンクのみを抽出した。そして、これらのリンクを用いて Web スクレイピングを行い、各論文の著者名、所属機関名、文献種別等を取得した。これらの一連の作業は R 言語と主に httr パッケージ、rvest パッケージを用いて行った。httr パッケージは API を通じたデータ取得に、rvest パッケージは Web スクレイピングのために用いた。

具体的な Web スクレイピングの方法を説明する前に、簡単に HTML について触れておきたい。HTML とは Web ページを制作するために開発された言語であり、現在公開されている Web ページのほとんどは主に HTML で記述されている。HTML では、文書内の各部分がどのような役割を持っているのかタグ「<>」によって示される。例えば、最上位の見出しを表す h1 タグ (<h1>)、一つの段落を表す p タグ (<p>)、また、文書のメタデータを表す meta タグ (<meta>) などである。タグには特定の機能を追加するために属性を付加することがあり、「<タグ名 属性名="属性値">」のように記述される。

さて、話を戻すと、J-STAGE からの Web スクレイピングでは主に head タグ内にある meta タグの属性に着目し、情報の抽出を行なった。著者名は name 属性の値が「authors」である meta タグの content 属性の値であり（例：<meta name="authors" content="信大 太郎">）、所属機関は name 属性の値が「authors\_institutions」である meta タグの content 属性の値である（例：<meta name="authors\_institutions" content="信州大

学”>)。著者情報は著者ごとに氏名、所属機関という連続する meta タグで記述されているが、著者によっては複数の所属機関情報を有する場合がある点に注意が必要である。この場合、それぞれの所属機関情報は異なる meta タグで表されるため、name 属性が「authors」と「authors\_institutions」の meta タグを別々に抽出してもそれぞれの個数が一致せず、著者名と所属機関情報の照合は難しい。

そのため、少々手間ではあるが、筆者らは以下のような方法で著者情報の収集と加工を行なった。基本的な考え方としては、name 属性の値が「authors」または「authors\_institutions」の meta タグを抽出し、著者ごとに ID を付与することで、著者名と所属機関の照合を行なった。具体的には、まず、(1) 各書誌画面の HTML から name 属性の値が「authors」または「authors\_institutions」の meta タグを抽出し、name 属性の値と content 属性の値を列とするデータフレームを作成する。データフレームとは R における行と列からなる 2 次元配置のオブジェクトであり、Excel の表のような形式である。次に、(2) name 属性が「authors」の場合には「1」、「authors\_institutions」の場合には「0」となる著者名フラグ列を作成する。そして、(3) 著者名フラグ列の累積和である ID 列を加えることで著者情報に ID を付与することができる。

(1)		(2)	(3)
name 属性	content 属性	著者名フラグ	ID (著者名フラグの累積和)
authors	信大 一郎	1	1
authors_institutions	信州大学 社会学部	0	1
authors	信大 花子	1	2
authors_institutions	信州大学 社会学部	0	2
authors_institutions	社会科学研究所	0	2

次に、(4) 上記のデータフレームを name 属性の値 (authors/authors\_institutions) によって 2 つのデータフレームに分割する。この際には name 属性の列や著者名フラグ列は必要ないため削除しても構わない。

authors のデータフレーム		authors_institutions のデータフレーム	
content 属性	ID	content 属性	ID
信大 一郎	1	信州大学 社会学部	1
信大 花子	2	信州大学 社会学部	2
		社会科学研究所	2

最後に、(5) ID 列をキーとして上記の 2 つのデータフレームを結合すれば、著者名と所属機関情報の照合が可能となる。

ID	氏名	所属機関
1	信大 一郎	信州大学 社会学部
2	信大 花子	信州大学 社会学部
2	信大 花子	信州大学 社会科学研究所

なお、上記の方法では著者の所属機関情報が複数あった場合に複数の行に別れているが、1 つの行に集約することも可能である。しかしながら、1 つの行に集約してもしなくても

「実績状況に基づく配分」における集計の際には特に大きな影響はない。というのも、「実績状況に基づく配分」においては、仮に1つの論文に対して同一の著者で2重に集計したとしても、1つの論文に対して各学系の上限は「1」にしかならないためである。

ここまで、著者情報を取得する方法を紹介したが、その他にも DOI (Digital Object Identifier) や論文タイトル、雑誌タイトル、出版年などの情報も meta タグには含まれており、必要に応じてこのデータフレームに追加していくことが可能である。また、論文によっては文献の種別がタイトルの上部に付されているため (図1では「事例報告」)、この情報も併せて取得しておくことを特にお勧めする。J-STAGE に掲載される記事は学術論文だけでなく、巻頭言や編集後記といった研究成果の発表とは関係のないものも含まれているため、研究 IR ではこれらの文献を除外する作業が必要になる。その際には、文献の種別が分かっているならば逐一論文データを確認する手間を省くことができるため、非常に有益な情報となる。詳しい説明は省くが文献の種別については head タグではなく body タグ内に格納されているので、各自で確認してもらいたい。

なお、上記の説明はあくまで1つの論文に関するデータの取得方法である。本稿執筆時には2017年から2019年に発行された信州大学の研究者が著者である論文は約1700報であり、最終的にはこれらの書誌画面のリンクを用いて1件1件ループさせて上記の処理を実行していくことになる。だが、これに要する時間はループ間隔に十分な余裕を持たせたとしても1日程度で十分に可能であった。

#### 4. J-STAGE から論文データを取得・利用する際の注意点

上記の方法で J-STAGE から論文データを取得、利用する際の注意点としては以下の3点が挙げられる。

第一に、Web スクレイピングする際には J-STAGE のサーバーに対する配慮が求められる点が挙げられよう。本稿で紹介した手法は機械的に J-STAGE にアクセスし HTML よりデータを抽出するというアプローチである。そのため、対象とする1つ1つの論文にアクセスする間隔を短くすることで、短時間のうちに大量のデータの取得も可能である。しかしながら、短時間のうちに大量にアクセスすることは、サーバーへの負荷を高め、サーバーダウンに繋がる恐れがある。

第二に、Web スクレイピングにより機械的にデータを取得できたとしても最終的には目視による確認も必要である。J-STAGE では各学術雑誌が査読制度を有しているか否かという情報を有していることから、事前に査読制度を有する雑誌を対象を絞って Web スクレイピングを行うことができる。しかしながら、個々の論文が実際に査読を経たものであるか否かに関する情報は判別できないため、それぞれの論文を精査しなければならない。その際には、前節でも触れた通り文献種別を予め取得し、巻頭言や編集後記といった対象外の文献を予め除外しておくことで作業時間を短縮できるだろう。

第三の注意点としては、実施体制が挙げられる。上記の通り、機械的にデータの取得が可能であるとはいえ、人の手によってある程度の作業が必要である。特に、「実績状況に基づく配分」では、J-STAGE だけでなく Web of Science や Scopus、また、各機関が独自で有するデータベースも用いているが、これらのデータベース間でのデータの重複を排除し



ていかなければならない。信州大学では筆者らの他、大学院生を2名雇用することにより、「実績状況に基づく配分」の研究業績に関するデータの整理に当たったが、作業時間はおよそ1ヶ月程度を要している。信州大学の承継教員数は2020年5月1日時点で約800名であるが、これよりも教員数の大きい大規模大学で同様の作業を実施するならば、より長期間の作業時間を確保するか、あるいは、より多い人員を確保する必要があるだろう。

## 5. 終わりに

本稿では日本語の学術論文データを整備するためにJ-STAGEを活用するに至った背景や具体的なデータの取得方法について報告した。本稿で報告したアプローチは事前に契約手続きが必要であるほか、何らかのプログラミング言語の知識を前提とする点で、研究IRで実践するためのハードルは低いとは言えない。しかし、通常では取得することが難しいデータを取得できる点が本稿のアプローチの大きな強みと言える。実際に、筆者らの事例ではJ-STAGEから取得した論文データのうち、Web of ScienceやScopusに採録されているものはわずか1割ほどである。

ただし、先述のようにJ-STAGEにも論文以外のデータも混在することから、J-STAGEの論文データを研究IRや評価に用いる際には更にデータの精選が求められる点に注意が必要である。筆者らの事例ではJ-STAGEから取得した論文データのうち、Web of ScienceやScopusに採録されている論文や、査読なしの論文、論文以外の記事を除いた、4割ほどのデータを「実績状況に基づく配分」のために用いている。手間は生じるものの、より正確に自機関の研究成果を捕捉するために必要なプロセスであることは特に強調しておきたい。

本稿は令和3年3月22日に開催された大学評価コンソーシアム「継続的改善のためのIR/IEセミナー2021 R1: IR実務担当者セッション」において発表した内容を論文化したものであるが、最後に、当セッションでの質疑応答の1つを取り上げて改めて考えてみたい。寄せられた質問は、「J-STAGEの運営者側でデータのダウンロード機能も提供すべきではないか」という内容であった。J-STAGEに限らず研究活動に関しては公的な機関が必要なデータを保有している場合が多く、同様の議論は他のデータベースに関しても耳にする機会も多い。そのため、この種のニーズは研究IRや評価担当者間で一定程度存在するものと思われる。

確かに、先述のような前提条件なくデータを利用できる環境が用意されているのは、研究IRや評価業務の担当者にとって理想的な状態かもしれない。しかしながら、運営者側もニーズや効果がどれ程のものか不確かな状態では、多大なコストを投じてまで実現するのは難しいのではないと思われる。とは言え、研究IRや評価業務は定期的、または随時発生するため、ただ手をこまねいている訳にもいかないだろう。このような中、研究IRや評価業務の担当者として必要なのは、運営者側とのコミュニケーションのもと、まずは現状でもできることを模索していくことではないかと考える。そして、長期的には利用者側と運営者側の相互理解のもと、研究IRや評価業務にも適した環境が実現されることを期待したい。

## 謝辞

本取り組みで J-STAGE を活用させていただくに当たり、ご協力いただいた国立研究開発法人科学技術振興機構の皆様にご心より御礼申し上げます。

## 引用文献

- 竹内健太（2019）「国立大学法人運営費交付金の行方―「評価に基づく配分」をめぐって―」『立法と調査』No.413, p.67-76. ([https://www.sangiin.go.jp/japanese/annai/chousa/rippou\\_chousa/backnumber/2019pdf/20190603067.pdf](https://www.sangiin.go.jp/japanese/annai/chousa/rippou_chousa/backnumber/2019pdf/20190603067.pdf))
- 中村真也（2020）「第6章国立大学法人運営費交付金とEBPM―評価に基づく配分について―」『EBPM(証拠に基づく政策形成)の取組と課題 総合調査報告書』p.101-118. ([https://dl.ndl.go.jp/view/download/digidepo\\_11460685\\_po\\_20190308.pdf?contentNo=1&alternativeNo=](https://dl.ndl.go.jp/view/download/digidepo_11460685_po_20190308.pdf?contentNo=1&alternativeNo=))
- 林隆之（2018）「大学評価の現場における人文・社会科学の研究評価の現状」『学術の動向』23(10), p.16-23. ([https://www.jstage.jst.go.jp/article/tits/23/10/23\\_10\\_16/pdf-character/ia](https://www.jstage.jst.go.jp/article/tits/23/10/23_10_16/pdf-character/ia))
- 文部科学省（2019）「令和元年度国立大学法人運営費交付金における新しい評価・資源配分の仕組みについて(成果を中心とした実績状況に基づく配分の仕組みの創設)」([https://www.mext.go.jp/content/1417264\\_001.pdf](https://www.mext.go.jp/content/1417264_001.pdf))
- 文部科学省（2020）「令和2年度成果を中心とする実績状況に基づく配分の仕組みについて」([https://www.mext.go.jp/content/20200722-mxt\\_hojinka-000008505\\_1.pdf](https://www.mext.go.jp/content/20200722-mxt_hojinka-000008505_1.pdf))
- 財務省（2020）「財政制度等審議会財政制度分科会歳出改革部会 文教・科学技術（参考資料）」([https://www.mof.go.jp/about\\_mof/councils/fiscal\\_system\\_council/sub-of\\_fiscal\\_system/proceedings\\_sk/material/zaiseier20201026/04.pdf](https://www.mof.go.jp/about_mof/councils/fiscal_system_council/sub-of_fiscal_system/proceedings_sk/material/zaiseier20201026/04.pdf))

\* オンライン文献の最終閲覧日は全て 2021 年 7 月 31 日である。

[受付：令和3年8月5日 受理：令和3年9月29日]