

# 研究成果情報管理業務プロセスの改善事例報告と分析

林 秀和<sup>1)</sup>, 山田 一成<sup>1)</sup>, 田島 嘉則<sup>1)</sup>, 片桐 孝洋<sup>2)</sup>, 大島 聡史<sup>2)</sup>, 永井 亨<sup>2)</sup>

1) 東海国立大学機構 情報環境部 情報システム運用課

2) 名古屋大学 情報基盤センター 大規模計算支援環境研究部門

a49978a@cc.nagoya-u.ac.jp

## Improvement case report and analysis of research results information management business process

Hidekazu Hayashi<sup>1)</sup>, Kazunari Yamada<sup>1)</sup>, Yoshinori Tajima<sup>1)</sup>,  
Takahiro Katagiri<sup>2)</sup>, Satoshi Ohshima<sup>2)</sup>, Toru Nagai<sup>2)</sup>

1) Information System Operations Division, Information Technology Department,  
Tokai National Higher Education and Research System

2) High Performance Computing Division, Information Technology Center, Nagoya University

### 概要

2023年9月現在まで、名古屋大学 情報基盤センターはスーパーコンピュータ「不老」を計算資源として利用者へ提供することにより、社会貢献や成果創出を支援して来た。計算需要支援による研究力強化や、人材育成、社会連携のために、情報基盤センターはスーパーコンピュータ「不老」利用者の研究成果情報を収集・分析する必要がある。本業務のプロセスに着目すると、利用者がアンケートに回答する点、センターがアンケート結果を取りまとめて分析する点において、改善の余地が大きかった。我々は2022年度より、本業務プロセスについてアンケートツールを含め改善することを試みた。本稿では、2022年度に実施した具体的な取り組みならびにその結果を事例紹介する。また、結果分析やフィードバックにより得られた知見に基づく、2023年度の取り組みについても述べる。

## 1 はじめに

2020年7月1日より、名古屋大学 情報基盤センターはスーパーコンピュータ「不老」を運用している[1]。スーパーコンピュータ「不老」による計算需要支援による研究力強化や、人材育成、社会連携を実現するために、情報基盤センターは利用者の研究成果情報を収集・分析する必要がある。ここでいう利用者の研究成果情報とは、表1に示す4種類：論文誌、国際会議、研究会報告、著書の実績としている。

情報基盤センターは、研究成果情報を収集するために、スーパーコンピュータ「不老」の利用者を対象としたアンケートを実施する。すなわち、スーパーコンピュータ「不老」の利用者は、年度内に投稿した論文誌や、参加した国際会議等の各種情報を研究成果情報としてセンターへ、アンケート形式で回答する\*1。センターは回答内容をとりまとめ、各種情報を体系的に管理し、分析する。つまり、研究成果を俯瞰して「どの研究分野の利用者が多いか?」、「昨年度と比較して利用者の傾向はどのように変化したか?」といった分

析をおこなう。分析により得られた傾向は、中長期的なハードウェア・ソフトウェアの整備、講習会開催提案等の計算需要支援を行うための判断材料とする。

本業務のプロセスに着目すると、利用者がアンケートに回答する点、およびセンターがアンケート結果を取りまとめる点において、改善の余地が大きかった。そのため、我々は2022年度より、本業務プロセスの改善を目指した。具体的には、Excel形式の研究成果情報アンケートツールを採用し、提出されたExcelファイルの中から必要セルを抽出し、管理表であるExcelファイルへ半自動的に転記するシステムを構築した。

本稿では、業務プロセスの改善に向けた取り組み、およびその結果について事例紹介する。また、その結果分析や、フィードバックを通じて得られた知見に基づく、2023年度の取り組みについても述べる。

2章では研究成果情報管理業務のプロセスを改善するにあたり、2022年度の取り組みについて紹介する。具体的には、本業務の詳細や、改善のための具体的なアプローチ、その結果について述べる。3章では、その取り組みで得られた知見を基に、2023年度に向けアンケートツール、システムをどのように改修するのかを述べる。最後に、4章でまとめる。

以降では、スーパーコンピュータ「不老」を「不老」、

\*1 実際には、表1に定める回答項目に加え、参考情報として、論文のDOI、各種成果物の査読有無、謝辞、インパクトファクター等を回答項目として用意している。

表 1: 研究成果情報の定義

No	種類	内容	回答項目
1	論文誌	和文雑誌、英文雑誌等	著者名、論文名、雑誌名、 巻数、号数、ページ番号、発表年
2	国際会議	国際会議プロシーディング等	著者名、論文名、プロシーディング名、 編者名、版元、ページ番号、発表年
3	研究会報告	研究会報告、ポスター発表等	著者名、論文名、研究会名、巻数、号数、 ページ番号、発表年
4	著書	単行本	著者名、タイトル、版元、発表年

名古屋大学情報基盤センターを「センター」、「不老」利用者を「利用者」と表記する。

## 2 2022 年度における業務プロセス改善に向けた取り組みの報告

### 2.1 2021 年度までの業務プロセス

2021 年度まで、利用者が回答する研究成果情報アンケートのツールとして、Office365 が提供する「Microsoft Forms」を用いていた。

利用者は、アンケート期間内に Forms へアンケート項目を入力することで回答し、センターへ提出する必要がある。センターは、Forms 管理画面よりアンケート結果を Excel 形式のファイルでダウンロードした後、人的作業を含む手順を経て管理表を作成する。

この業務プロセスに着目すると、利用者は研究成果情報を Forms へ登録することは負担が大きいことがわかった。例えば、利用者が研究成果情報の 1 つとして論文誌を入力することとする。利用者は、表 1 に定める回答項目に従い、著者名、論文名、雑誌名、巻数、号数、ページ番号、発表年等をフォームへ入力する必要がある。Forms の仕様上、画面レイアウトは縦方向に長いために、項目をセルに繰り返し入力することは負担がかかることに加え、登録内容を俯瞰しにくい。また、2021 年度当時は Forms は回答が完了したアンケートを確認し、修正することができなかった。そのため、利用者は研究成果情報を追加登録したい場合は最初から登録し直す必要があった\*2。

センターがアンケートを取りまとめ管理表を作成する段階に着目すると、人的作業が発生していた。具体的には、アンケートの重複回答を目視で確認する作業や、必要に応じてその重複行を削除する必要があった。

これらのことから、業務プロセスを改善するにあたり、利用者目線では「回答しやすさ」、「確認しやすさ」、「修正しやすさ」といった要素が強く求められている。

その一方でセンター目線では、効率的 (より簡単に、より早く、より確実に) な集計を可能とする仕組みづくりが求められている。

### 2.2 2022 年度に実施した業務プロセスの改善

研究成果情報アンケートのツールとして、我々は Excel を選定した。その理由は、Excel を始めとする Office はどの現場においても導入しやすく、多くの利用者にとってシート操作、セル入力操作は馴染みがあるためである。また、利用者グループ内で Excel ファイルを共有することで、他のメンバーもアンケートに回答しやすいことも挙げられる。これらの特徴はセンターに対しても同様のメリットがあり、システム導入コスト、運用コスト (テンプレートファイルの修正しやすさ等) を低減できる。

本システムの全体像を図 1 に示す。本システムは大きく分けると、下記に示す 3 つの要素から構成される。

#### 利用者用入力 Excel ファイル

本ファイルは研究成果情報アンケートの役割がある。利用者は本ファイルのセルを埋めることでアンケートに回答し、センターへ提出する。

#### 管理者用統計 Excel ファイル

本ファイルは利用者の研究成果情報を管理し、分析する管理表の役割がある。本ファイルの各セルは、利用者用入力 Excel ファイルにおける入力セルを抽出したものである。

#### 統計自動化スクリプト

本スクリプトは、利用者用入力 Excel ファイルから管理者用統計 Excel ファイルへ自動的に転記する役割がある。提出された利用者用入力 Excel ファイル群に対して、各ファイルのセルをコピーし、管理者用統計 Excel ファイルの指定セルへペーストする。本スクリプトでは、Python をベースに実装しており、Excel ファイルを扱うライブラリ「OpenPyXL」を使用している [2]。このライブラリは Excel ファイルを開くことや、シート内の各セルに対する Read/Write を可能とする。

\*2 2023 年 9 月 26 日時点の仕様では、Forms アンケートを確認し、提出後に修正することは可能である。しかし、上記の画面デザインの理由より、詳細情報を確認して修正する操作は負担がかかるといえる。

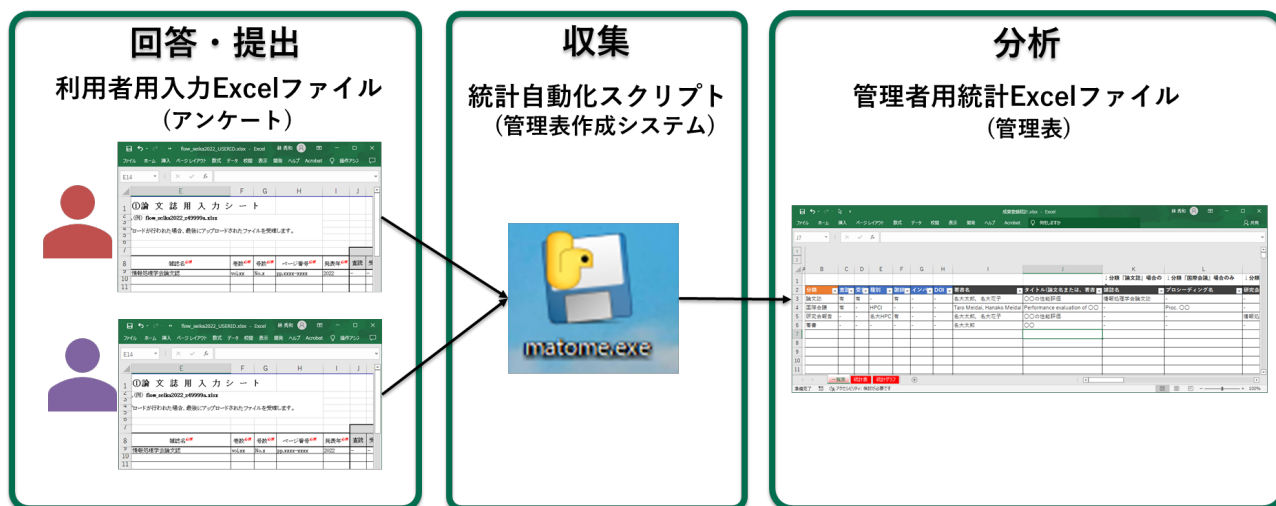


図 1: システムの全体像

なお、Excel ファイルを主軸とするために、利用者、センターが用いる端末 OS は Windows であることを前提としている。本システムをベースとした業務プロセスを下記に示す。

- (1) 利用者は、利用者用入力 Excel ファイルを回答する。
- (2) 利用者は、NUSS(Nagoya University Storage Service) 教育研究ファイルサービス (以下、NUSS) へ Excel ファイルをアップロードする。
- (3) センターは、NUSS より提出されたファイル群をダウンロードする。
- (4) センターは、統計自動化スクリプトを実行することで、管理者用統計 Excel ファイルを更新する。
- (5) センターは、必要に応じて管理者用統計 Excel ファイルを確認する。

システムの設計方針、動作フロー、運用ルール等の詳細情報については、大学 ICT 推進協議会 (AXIES) 2022 年度年次大会にて報告している [3]。

### 2.3 実業務プロセスの振り返り

**■利用者がアンケートに回答するプロセス** センターが「不老」を導入した 2020 年度から 2022 年度に至るまでの、研究成果情報の登録件数を示したグラフを図 2 に示す。2022 年度は合計 235 件の登録が行われており、この数値は前年度以前と比較して約 3 倍である。2022 年度の各種類ごとの登録件数も全て増加しており、特に研究会報告 (査読無し) は、前年度と比較して、6.5 倍増加した。

2022 年度はアンケートツールの変更に伴い登録件数が増加していることから、利用者はアンケートへ回答しやすくなったといえる。増加した理由は、2.2 節で述べた通り、Excel ファイルは、どの利用者に対しても操作に馴染み深く、研究メンバー内で共有可能と

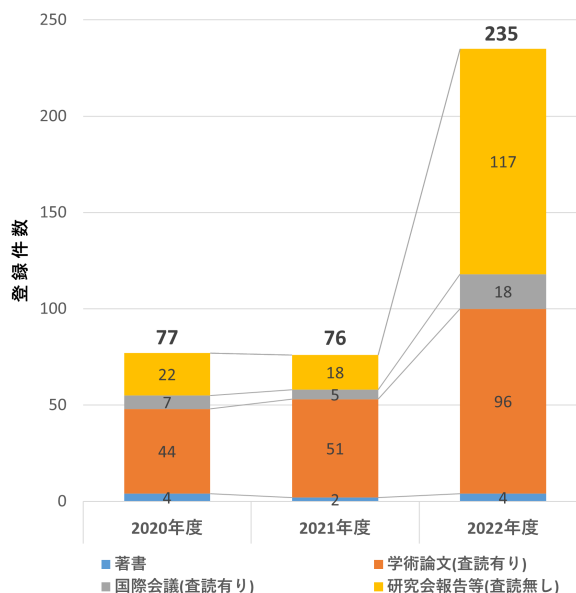


図 2: 年度ごとの研究成果情報の登録件数推移

いった特徴を持ち回答しやすかったことも要因の一つとして考えられる。

**■センターがアンケート管理表を作成するプロセス** 本システムを導入することにより、手作業工程を減らし、作業ミスを減らすことができた。これは、2.1 節で述べた管理表作成時の手動調整工程が減ったためである。既存プロセスにおいて利用者が Forms で複数回登録した際、登録内容が重複するケース、重複しないケースの両者が発生する。前者の例は、利用者が Forms で全項目を最初から登録し直したケースである。後者の例は、利用者が Forms で変更分のみを登録するケースである。このことから既存プロセスにおいて、センターは目視で登録内容が重複しているかを判断し、手作業で削除する必要があった。2021 年度は重

複登録が特に多く発生したために、提出後のアンケートを集計して管理表を作成するまでの作業工程は、1回あたり1時間削減した。

#### 2.4 大学 ICT 推進協議会 (AXIES)2022 年度年次大会でのフィードバック

2.2 節で述べた通り、本業務プロセスの改善方針について、大学 ICT 推進協議会 (AXIES) の 2022 年度年次大会で報告した [3]。

質疑応答の中で、聴講者より「統計自動化スクリプトが Python で構築されているならば、デジタルオブジェクト識別子 (Digital Object Identifier: DOI) をもとに、それに紐づく詳細情報を参照する設計とすると、より高信頼なデータを取得可能と思われる。」という助言があった。DOI とはインターネット上で公開されている資料、記事、研究データ等のコンテンツに付与される永続的識別子であり、ISO によって標準化されている国際的規格である。”<https://doi.org/>”もしくは”<https://dx.doi.org/>”の後に DOI を付与した URL は、そのコンテンツへの恒久的なアクセスが保証される。

2022 年度にセンターへ提出された研究成果情報アンケートファイルの回答項目を確認すると、誤字や空欄が含まれるケースが散見された。すなわち、利用者が登録する研究成果が増えるほど手入力によるミスも起こりやすくなる。このことから、利用者の入力項目を DOI のみに限定する方針は、利用者に対しては入力時間の削減につながる一方、センターに対しては高信頼なデータを得られる点で双方にメリットがある。

### 3 2023 年度の取り組みについて

#### 3.1 基本設計

2023 年度の本業務プロセスで用いる利用者用入力 Excel ファイルを、図 3 のように修正する。

図 3 からわかるとおり、研究成果に対して DOI が割り当てられている場合、利用者はその DOI を Excel へ入力することで登録を完了することとした。一方で、DOI が割り当てられていない場合、または不明な場合には、従来通り表 1 に定める各種項目を入力する仕組みとした。

ただし、利用者が DOI を入力ミスする可能性があり、意図しない成果を登録される可能性もある。これを防ぐために、利用者が入力した DOI は意図した値なのかを確認させるためのセルを用意した。具体的には、DOI をセルに入力した際、隣接するセルには DOI に紐づいた参照用の URL が自動入力される。また、入力した DOI が意図した値なのかを確認するよう、本 Excel ファイルの注釈文に追記した。

#### 3.2 詳細設計

DOI を参照することで、詳細な研究成果情報を自動的に取得する設計を述べる。

2022 年度に提出された研究成果数は、図 2 に示すとおり合計 235 件であった。この内、参考情報として利用者が DOI を記載した件数は計 88 件であった。本年度は、この 88 個の DOI をサンプルとし、スクリプトを設計することとする。これらの DOI を分析すると、おおよそ 2 種類に分類できた：(1)Crossref で参照可能な研究情報、(2)J-STAGE で参照可能な研究情報。(1)については 86 件の登録があり、(2)については 2 件の登録があった。これらのことから、我々は DOI を Crossref、J-STAGE より参照するようスクリプトを改修することとする。

##### 3.2.1 Crossref

Crossref は、学術コンテンツを対象とする DOI 登録機関 (RA) であり、学術コミュニケーションの改善を目的として世界中の学術出版社や学協会が参加する非営利組織である [4]。Crossref に DOI が登録されている場合には、同組織が管理している REST API サービス `crossrefapi` を用いることで、メタデータを検索可能である [5]。具体的には、検索対象 DOI を付与した URL を、`crossrefapi` に対して GET リクエストした場合、その DOI に紐づくメタデータを json 形式のレスポンスで得られる。実際に得られるレスポンスの例を表 2 に示す。

##### 3.2.2 J-STAGE

J-STAGE とは、文部科学省所管の国立研究開発法人科学技術振興機構が運営する学術ジャーナル発信オンラインプラットフォームである [6]。J-STAGE に公開された研究成果は DOI が付与されるが、`crossrefapi` を用いても検索できない。また、J-STAGE が運営・管理する WebAPI サービス J-STAGE WebAPI があるが、これは DOI をリクエストパラメータとして指定できない [7]。

以上のことから、J-STAGE で公開された研究成果情報を取得するために、Web スクレイピングを行うこととする。Web スクレイピングとは、Web サイトより情報を抽出する技術を指す。本ケースでは、J-STAGE の HTML ドキュメントを解析し、タグ要素を取得することとする。具体例として、図 4 に示す。図 4(a) では J-STAGE に公開された記事であり、図 4(b) はこの記事の HTML ドキュメントである。すなわち、タグ要素を機械的に確認することで、研究成果情報を抽出することができる。

なお、J-STAGE の利用規約 第 3 条 (禁止事項) において、機械的及びそれに準じた手段を用いて登載データを大量にダウンロードすることを禁止している [9]。先行事例を挙げると、久保らは J-STAGE の約 1700

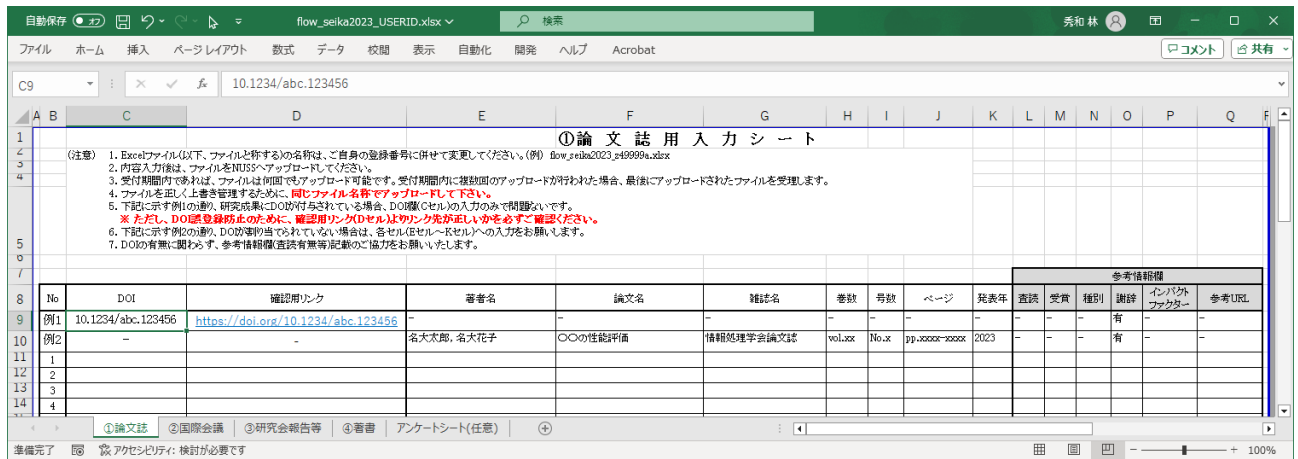


図 3: 利用者入力 Excel ファイル

表 2: crossrefapi の json レスポンスの値参照情報

No	key 名	value 値	value 値の具体例
1	type	研究成果分類	journal-article, proceedings-article, posted_content, book_chapter
2	author	筆者の姓名、及び所属	{'family': 'Meidai', 'given': 'Taro'} <sup>a</sup>
3	title	研究成果のタイトル	[Performance evaluation of OO] <sup>b</sup>
4	container-title	ジャーナル、プロシーディング名等	Proc. OO
5	page	ページ番号	1-6
6	volume	巻数	50
7	issue	号数	1
8	publisher	版元	ACM
9	published	発表年	{date-parts: [2023,9,1]} <sup>c</sup>

<sup>a</sup> value 値の型は dict である。dict 内の value 値を結合させ、1 人の著者として集計する。上記の例では、'Meidai Taro' となる。

<sup>b</sup> value 値の型は list である。list 内の要素を結合させ、タイトルとして集計することとする。

<sup>c</sup> value 値の型は dict である。dict 内の value 値の内、先頭要素を発表年として集計する。上記の例では、'2023' となる。

報分の記事を Web スクレイピングするために、科学技術振興機構と協議を行っており、データ利用契約書を交わしている [8]。一方、本ケースは 3.1 節で述べた通り、研究成果情報登録が行われた総数 235 件に対して J-STAGE の DOI 記載があった件数は 2 件であり、先行研究と比較して大量といえない。そのため、J-STAGE が運用する Web サーバに対する負荷を十分に配慮した上で、Web スクレイピングを行うものとする。具体的には、J-STAGE に対して Web スクレイピングを 1 回行った場合には、次の Web スクレイピングを行うまでに 10 秒以上の時間的インターバルを設けることとする。

HTML ドキュメントを解析する Python パッケージとして、我々は BeautifulSoup を用いた [10]。

### 3.2.3 統計自動化スクリプトの実行フロー

スクリプトのフローチャートを図 5 に示す。従来より使用していた Step2 最新版ファイルチェック機構、Step3 ファイル名称チェック機構は流用するものとする [3]。Step4 にて、利用者の DOI に入力の有無を判断する機構を設けた。DOI の入力がある場合には、この crossrefapi を用い DOI を検索条件としたリクエストを行う。リクエストが成功したならば、そのレスポンスである json の value 値を参照して管理表へ記載する。一方で、リクエストが失敗したならば、次に J-STAGE に対して Web スクレイピングを行い、その結果を管理表へ記載する。一方で、DOI の入力がない場合は、従来通り各セルを参照し転記する。

## 応用数理

資料トップ 巻号一覧 この資料について

J-STAGEトップ / 応用数理 / 20巻 (2010) 3号 / 書誌

「数値計算のための自動チューニング」特集号に寄せて(〈特集〉数値  
めの自動チューニング)

片桐 孝洋

著者情報

片桐 孝洋  
東京大学情報基盤センター

ジャーナル フリー

2010年 20巻 3号 p. 188-190

DOI <https://doi.org/10.11540/bjsiam.20.3.188>

詳細

(a) 閲覧者用の Web ページ

```

<ul class="accordion_body_ul">
  <li>
    <span style="font-size:1.1em">片桐 孝洋</span>
    <p class="accordion_affilinfo"></p>
  </li>
</ul>
</div>
<div class="corr-auth-sapopup-wrap" id="saPopUp" style="display:none"></div>
</div>
<div class="global-para"></div>
<div class="global-tags">
  <span class="tags-wrap original-tag-style" title="ジャーナル">ジャーナル</span>
  <span class="tags-wrap freeaccess-tag-style" title="フリー">フリー</span>
  <br>
</div>
<p class="global-para"> 2010年 20巻 3号 p. 188-190 </p>
<div class="articleoverview-doi-wrap" id="fig-doi">
  <span class="doi-icn">DOI</span>
  <a href="https://doi.org/10.11540/bjsiam.20.3.188" class="blueink-style" style="wo
rd-break:break-all">https://doi.org/10.11540/bjsiam.20.3.188</a>
</div>
<div class="articleoverview-doi-wrap"></div>
<div class="top-margin-2x bottom-margin-2x"></div>
<div class="accordion_container"></div>
</div>

```

(b) Web ページのソースコード

図 4: J-STAGE の資料・記事情報 Web ページの分析

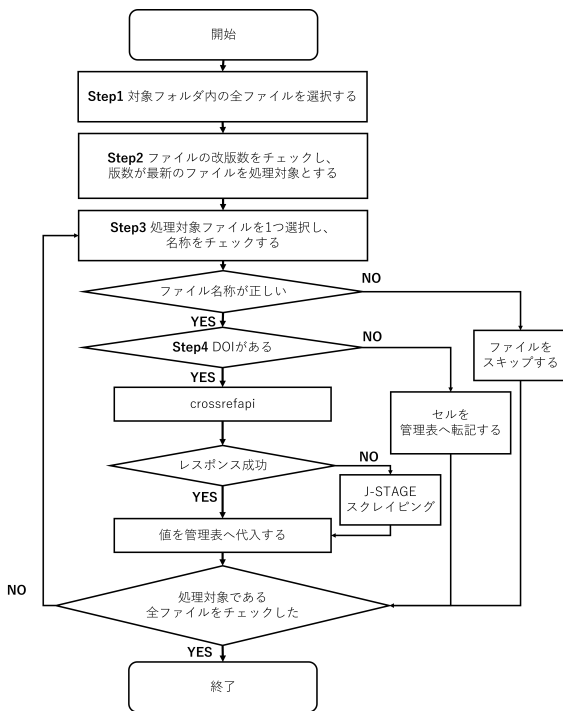


図 5: スクリプトの実行フローチャート

## 参考文献

- [1] 名古屋大学 情報連携推進本部 “スーパーコンピュータシステム” <https://icts.nagoya-u.ac.jp/ja/sc/> (accessed 2023-09-15).
- [2] Pypi “openpyxl” <https://pypi.org/project/openpyxl/> (accessed 2023-09-15).
- [3] 林 秀和, 山田 一成, 田島 嘉則, 片桐 孝洋, 大島 聡史, 永井 亨, 『研究成果情報管理業務のシステム化による業務プロセスの改善』, 大学 ICT 推進協議会 2022 年度全国大会, 2022.
- [4] Crossref <https://www.crossref.org/> (accessed 2023-09-15).
- [5] Pypi “crossrefapi” <https://pypi.org/project/crossrefapi/> (accessed 2023-09-15).
- [6] J-STAGE <https://www.jstage.jst.go.jp/browse/-char/ja/> (accessed 2023-09-15).
- [7] J-STAGE WebAPI について <https://www.jstage.jst.go.jp/static/pages/JstageServices/TAB3/-char/ja> (accessed 2023-09-15)
- [8] 久保 琢也, 伊藤 広幸, 『J-STAGE を活用した日本の学術論文データの整備』, 情報誌「大学評価と IR」第 12 号, 2021.
- [9] J-STAGE 利用規約・ポリシー <https://www.jstage.jst.go.jp/static/pages/TermsAndPolicies/ForIndividuals/-char/ja> (accessed 2023-09-15).
- [10] Pypi “beautifulsoup4” <https://pypi.org/project/beautifulsoup4/> (accessed 2023-09-15).

## 4 まとめ

本稿では、研究成果情報管理業務のプロセスに着目し、改善に向けた取り組みを紹介した。2022 年度は、利用者の登録件数を増加させつつ、センターの業務負担を削減できた。また、AXIES2022 年度年次大会で得られた知見を基に、利用者の負担を下げつつ、より高信頼なデータを取得するようシステムを改修した。