

CGSIレポート

日本学術振興会グローバル学術情報センター
JSPS Center for Global Science Information

第 4 号（平成 28 年 3 月 28 日）

本レポートについて

平成 25 年 4 月の発足以来、独立行政法人日本学術振興会グローバル学術情報センター（以下、「情報センター」と略）では、本会諸事業の情報の収集・蓄積を進め、特に科学研究費助成事業（科研費）については、申請から実績報告に至るまでの主要な全データをデータベース化するなど、本格的なデータ分析への足場を整備するとともに、分析課題の検討や試行的な調査・分析を進めてきました。

一方、こうした事案のほとんどが非公開の申請・審査データに関係するものである上、調査・分析の成果は事業改善に向けて本会内部で使用することを想定しており、外部からは情報センターの活動が見えにくい状況にあります。これまで刊行してきた CGSI レポートは、創刊時に述べたとおり内容が公開できるものに限定されていることもあり、必ずしも情報センターの活動を的確に反映するものではありません。

そこで今回、CGSI レポートの場を借りて、これまでの情報センターの活動方針や主な分析課題を紹介すると共に、本年度の活動内容の概略を紹介することにしました。調査・分析の結果自体は公表できないものが多いため、多少わかりにくい部分はあるかと存じますが、本レポートが情報センターに対する皆様のご理解の深化につながれば幸いです。

日本学術振興会グローバル学術情報センター所長 前島 信

平成 27 年度の調査・分析活動報告

1. 調査・分析の方針とその背景

情報センターの主な業務の一つは、学振の諸事業に関する情報の収集・蓄積と、これらを用いた調査・分析である。しかし、本会で扱っている事業は大小併せて 30 を超え、各担当が保有しているデータも各事業を進めるために最適化されており、情報センターの限られたリソースで統一的に収集・蓄積（そのためのデータ加工を含む）するには非常な困難が伴う。このため、本格的な活動を開始して以降これまで一貫してとってきた方針は、本会の事業の中で最も規模が大きく、研究者にとっての興味・関心が高く、与える影響も大きい科研費について優先的にデータ整備を進め、併せてこれに対する分析課題を検討することである。

データの収集については、分散管理されていた科研費データのうち、重複を取り除いたものを情報センター内に集約し、独自のデータベースとして整備することで、電子申請が確立された以降の科研費データに関しては、各研究課題について申請から審査、交付、研究実績報告¹といった全過程を追いかけられる環境を整えた。

このデータベースを用いた分析課題の検討については、科研費の公募・審査等の検討や審査員の選定等を行っている本会学術システム研究センター（以下、「システムセンター」と略）の議論を踏まえてそのニーズを洗い出しつつも²、基本的には自由な発想で行った。検討にあたっては情報センターの構成員で原則週に一度の会議を開催し、そこでの議論を経て分析課題を決定する方法をとっている。

2. 具体的な調査・分析課題

前述した情報センター会議で話題に上るなどして実施した調査・分析課題について、それぞれの概要と関連する活動内容を以下に列挙する（それぞれの活動内容詳細は、4 ページ以降を参照されたい）。

1. 科研費審査データ、特に1段審査（書面審査）の評点分析

科研費の審査は多くの種目において二段審査制（複数の審査員による書面審査（1段審査）と、それとは別の審査員による合議審査（2段審査）の組み合わせ）を採用しており、1段審査で付された評点は当然審査結果に大きく影響する。1段審査は各申請に対して各審査員が各評定要素についての絶対評価を行った上で5段階の総合評点（相対評価）を付し、総合評点についてはこれをTスコア化³して合議審査で使用している。こうした評点情報を審査セット（多くは種目別かつ細目別）毎に詳細に分析した事例はなく、まずこれを行うことで審査の実情がどのようなものであるかを把握することは重要と思われ、将来その結果を審査員の割り振り等の参考とすることも考えられる。こうした動機で行われたのが「A. 主成分分析による科研費審査結果データの分析」である⁴。

2. 科研費申請データのテキスト分析による応募動向の分析

前の項目で取り上げた審査データに加え、情報センターでは科研費の申請書である研究計画調書の本文も収集・データ化を進めている。科研費では毎年度約10万件の申請がなされており、しかもそれらは全て研究者の自由な発想に基づくものであるため、その集合体は日本において現在行われている（あるいは行われようとしている）学術研究のかなりの部分を網羅していると言っても過言ではない。システムセンターでは毎年度各分野の学術

¹ 研究実績報告に関しては、国立情報学研究所から科学研究費助成事業データベース（KAKEN）のデータを提供いただいている。

² 情報センター所長はシステムセンター副所長を兼任しており、システムセンターの議論を遅滞なく情報センターに還元している。

³ Tスコアについては、科研費の「第2段審査の手引」を参照されたい。

http://www.jsps.go.jp/j-grantsinaid/01_seido/03_shinsa/index.html

⁴ 同じような「審査セット毎のTスコア」による審査は本会の特別研究員事業でも行っており、科研費のデータが揃うまではその審査データを用いて分析の試行を行っていた。

動向調査を行っているが、まさに学術動向そのものと言えるこうした研究提案の膨大なテキストデータを統計的に分析することによって、より効率の良い調査が可能となるかもしれない。

関連して、こうした研究提案は現行制度では基本的にいずれかの細目で審査されているが、限られた審査員が限られた時間の中で審査しなければならない事情に鑑み、個々の細目は学術上の適切な審査可能単位というよりは、それを意識しつつも審査件数に応じてある程度現実的に設定した審査区分である。また、応募する細目を選ぶのは申請者自身であり、特に新しい分野の研究や境界分野の研究などは、細目を設定した側や審査員側の意図に関わらず、どの細目に申請されるかはわからない。このため、適切な審査に近づけるためには実際の応募内容を考慮した不断の細目見直しとそこに適切な審査員を配置することが極めて重要であり、研究提案のテキスト解析はそのための基本的情報の一つといえる。システムセンターでは平成 30 年度応募に向けた細目等の見直しを進めており⁵、その中でも当然のことながらこれまでの各細目への応募内容が考慮されているが、これを助ける定量的な分析として、ある細目の申請全体に含まれる単語から細目同士の距離を測るクラスター分析を昨年度は試行的に実施し、今年度はこれに引き続き、「B. 科研費申請内容の潜在意味解析と可視化」といった分析も行っている。

3. その他、科研費審査業務の効率化・改善にかかる案件

これまでに述べたような科研費審査に関する分析課題を検討する上で、副次的に出てきたアイデアもいくつか存在する。そのうちの 하나가延べで 6000 人を超える審査員の選定の効率化である。これは毎年度システムセンターで行っているが、用いる審査員データベース⁶に搭載されている審査員候補者（ほとんどが科研費の受給経験者）の情報のうち、研究内容にかかる部分は本人の申告制であることもあって必ずしも充実しておらず、ある審査候補者が本当にある細目をカバーできるか否かは、個別に研究業績や研究分野を調べたりすることにより、ゼロから人の目で判断するしかないのが現状である。一方、例えばその候補者の過去の科研費の実績報告や申請時の研究計画調書のテキストと、前項で述べた細目単位のテキストデータとの類似性を見ることで、自動的に各細目への審査員割付案を作成することが原理的には可能である。このための基礎研究が「C. 審査員の選定および査読割り当ての自動化に関する調査」である。

この他にも、1 段審査において各審査員の付す評点はその審査対象について正規分布するという T スコアによる審査の前提について、より自然な仮定として、各申請の質が申請グループ全体で見ると正規分布しているとした場合に、どのような評点の付し方と処理が可能かについて基本的な考察を行うなど（「D. 項目反応理論(IRT)に基づく得点の標準化に関する検討」）、科研費審査の効率化や改善は常に念頭に置いている。

4. 諸外国の学術振興機関の事業の実施状況に関する情報の収集・蓄積

1 ページで述べたような方針に準じ、諸外国のファンディングエージェンシーが実施する

⁵ http://www.mext.go.jp/a_menu/shinkou/hojyo/1362786.htm

⁶ http://www.jsps.go.jp/j-grantsinaid/14_kouho/db.html

科研費のような公的研究費に関する情報や、主要国の学術研究活動に関連した政策動向等の情報の収集を進めている（「F. 諸外国の学術研究動向等の調査分析」）。

5. 事業等の定量的な成果分析

学振で行っている事業はほぼ全てが学術研究の推進とそれを担う人材の育成（いずれも国際交流を含む）に関するものであり、個々の採択課題も当然のことながら学術的に判断する必要がある。このため、有識者による評価（ピア・レビュー）が本会の事業評価の原則である。

一方、こうした評価を何らかの形で定量化できないかという点は情報センター発足当初から分析課題として検討されてきたが、現在のところ極めて難しいと言わざるを得ない。個別の評価対象というより、その集合体としての事業単位については、定量的な評価法としていわゆる計量書誌学的方法があり、情報センターでも試行してきた（「G. 文献データベースを利用した学振の事業による支援の効果の分析」）。この手法は学術研究の一つの側面を表すものではあるが、そのみで全体を評価するものとして使うことはできない。

このように事業等の定量分析は困難ではあるが、今後もこの点については継続して検討していきたい。

3. 今年度の調査・分析活動報告

以下、3名の分析研究員により、今年度の調査・分析活動の詳細を報告する。なお、本レポートで紹介するには至らなかった調査・分析課題や、現在進行中の課題もあり、以下に述べられる内容が情報センターの活動の全貌ではないことに注意されたい。

A. 主成分分析による科研費審査結果データの分析（三浦分析研究員）

主成分分析は、多数の変量が様々な値をとる中で、そこに存在する相互依存性を相関係数により計測し、そこから各変量の共通変動成分を取り出すものである。取り出された共通変動成分は主成分と呼ばれ互いに独立（相関がゼロ）であり、その数は元の変数の数と同じである。元の変数全体の総変動を最も強く説明する主成分を第1主成分、次に強いものを第2主成分、さらに順に第3主成分…という。説明する力の強さは寄与率として表される。これは相関係数行列から計算される各固有値の固有値総和に対する比率である。例えば科研費のある種目では審査員が4名であり、元の変数を基準化して計算を行った場合変数数は4であり、したがって総変動は4である。各主成分の寄与率の和はこの総変動4に等しく、各主成分の寄与率は各主成分に対応する固有値をこの4で割ったものである。

第1、第2の主成分の寄与率がかなり大きければ（例えば主成分あたりの平均である4分の1より大きければ）、その大きさに応じて、例えば第1主成分だけで、あるいは第1と第2の主成分により各審査員の各申請に対する審査評価点をかなり説明できることを意味する。解釈としては、例えば、第1主成分の寄与率が格段に大きい場合、4名の審査員の審査基準が相互にかなり近いと考えられる。また、第1主成分の寄与率がある程度大きいが、しかし十分ではなく、第2主成分の寄与率あるいは説明力を加えることによってのみ各申請に対する評価点を十分に説明できる

場合もある。この場合は二つの異なる審査基準を持っていると考えられる。これは当該研究分野の方法論などの差異を反映するのかもしれない。

一例として、ある年度・種目の2つの系に対して上記の分析を行った結果を図1に示す。上図と下図はそれぞれ異なる系についての結果である。散布図の各点は、一つの審査セット（基本的に細目）を表している。審査セットによっては第1主成分の寄与率が0.7を超えるものがあり、他方で0.4以下のものもある。ここでは系別の傾向の違いを見てみたが、こうした分析をより細かい分野別に行ったり、時系列で行ったりすることも可能である。

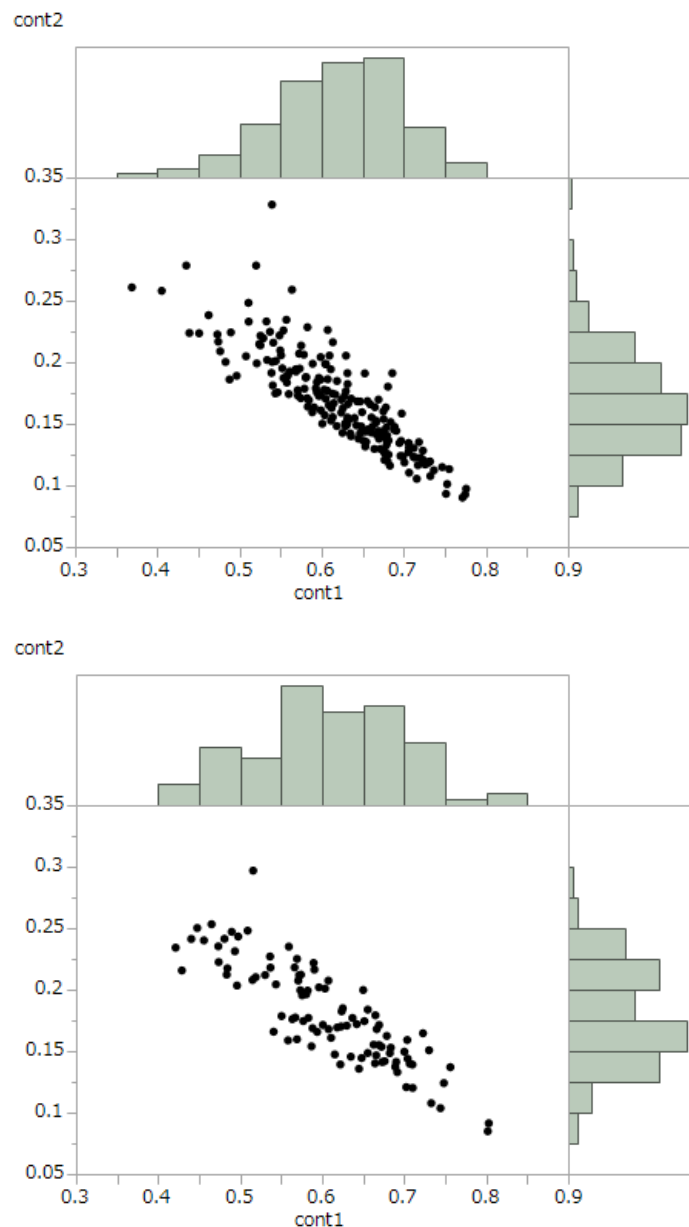


図 1. 科研費申請のある年度・種目・系について、各審査セットにおいて付された評点の第1・第2主成分の寄与率の散布図。対象となった審査セット数は上図の系で200、下図の系で99である。各図の上及び右にあるヒストグラムはそれぞれ各点の横軸及び縦軸の値のヒストグラムである。

B. 科研費申請内容の潜在意味解析と可視化（持橋分析研究員）

科研費申請書の科学的内容が全体としてどう分布しているかを把握することは、基礎的で重要な知見となり、また、細目の設定を考える上でも重要な情報である。このため、ある分野の各申請書の「研究目的」のテキストを単語に分割し、その頻度ベクトルを確率的潜在意味解析法の一つである潜在ディリクレ配分法(LDA)によって統計的に解析し、t-SNE という非線形可視化法によって図2のように可視化する予備調査を行った。この結果は、色で表されている細目の散らばりが直観と合致し、さらに、前年度までに行っていた細目を単位としたクラスター分析（言語データに必ずしも適さない多変量解析によって得られた細目間の相関係数）より広い情報を視覚的に把握することができる。

結果がトピック数に依存するためそれを自動推定することや、本分析を申請全体に適用すること、階層的な分野構造を表現するより精密な統計モデルを適用すること、などが来年度の課題である。

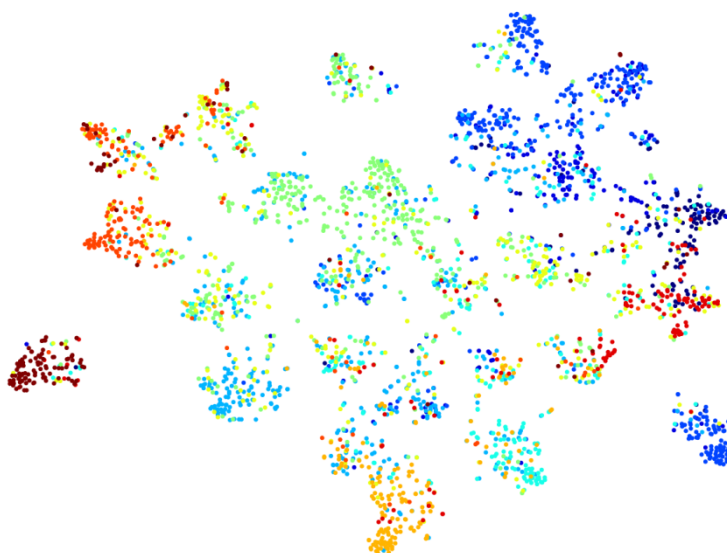


図 2. 科研費申請の確率的潜在意味解析による、ある分野の申請の可視化結果の例。各点が申請を、色は細目を表している。距離が近いほど申請内容が類似していることを意味する。

C. 審査員の選定および査読割り当ての自動化に関する調査（持橋分析研究員）

科研費の審査員の選定、および論文割り当ては、現在のところシステムセンターの研究員による手作業で行われており、きわめて負荷が高いものになっている。計算機科学の一部の国際会議では投稿された論文の査読割り当てが「B. 科研費申請内容の潜在意味解析と可視化」で述べた潜在ディリクレ配分法などの統計モデルにより自動化され、高精度な割り当てが実現されており、コミュニティにも受け入れられていることから、これらの内容について調査し、今後学振において同様の支援システムが導入できないか検討した。

D. 項目反応理論(IRT)に基づく評点の標準化に関する検討 (持橋分析研究員)

科研費の審査において、1 段審査員によって付される総合評点は、原則として 5 段階がほぼ正規分布となるように付すよう求められており、その結果が T スコアとよばれる値に変換されて用いられている。しかしこの方法は、離散的な観測値をそのまま表面的に変換するものであり、各審査員の評点が厳密に正規分布に従うわけではないこと、審査員による評価軸の違いを考慮に入れていないこと等の問題が指摘されてきた。

これに対して、計量心理学の分野で提案され、TOEIC をはじめ各種テストにおいて広く使われ始めている項目反応理論 (IRT) に基づく標準化について検討した。具体的には、科研費の審査の場合には評点は 1 から 5 のように段階で示されることから、Samejima(1969)の段階反応モデルを基にして、潜在的な「真のスコア」 θ を計算する方法を提案した。これは、 θ に対して潜在的に一種のロジスティック回帰モデルがあり、その結果観測値である評点が得られたとする統計モデルである。ロジスティック回帰モデルの係数は審査員によって異なる潜在変数であり、これも評点全体から推定される。

IRT によるこの方法は、(1) 各審査員が評点を必ずしも正規分布させてつける必要がない、(2) 審査員の「評価軸」や「信頼性」が統計モデルから得られる、(3) 離散的な評点に対する精密な統計モデルであり、評点を表面的に変換するものではない、といった複数の利点を同時に持っている。

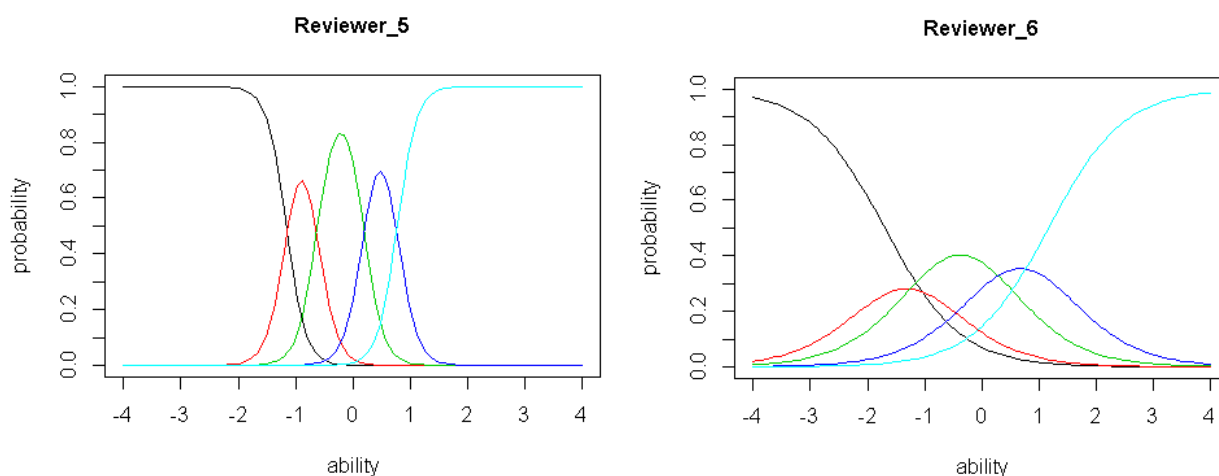


図 3. IRT (項目反応理論) によって導かれた、審査データから計算された各審査員の項目反応曲線の例。横軸が θ (ability) の値を、縦軸がその θ でそれぞれの評点となる確率を表している。

計算を安定化させること、および従来の T スコアと比較して実際にどのような違いが得られるのか等について、来年度も引き続き調査していく。

E. 諸外国の学術研究動向等の調査分析 (遠藤分析研究員)

近年、各国の学術振興機関、政府機関、大学、学術団体、シンクタンク、そして様々な国際機関等は日々膨大な情報を公表し、その多くはインターネット経由で容易に入手可能であることが

ら、これらを学振の業務に役立てる目的で収集・分析を行った。特に、主要国の学術振興機関が行うグラント等による研究資金の配分に関する情報は、学振の研究助成事業をはじめとする諸事業の実施の検討に直接参考となり得ることから重点的に取りまとめを行っており、昨年度は米国、ドイツ、英国、今年度は中国、韓国等のファンディングエージェンシーを対象として調査を行った。この成果は、学振内部向けには「ファンディングエージェンシーの審査システム」報告書として共有し、外部向けにはCGSIレポートとして公表した⁷。

また、学振が行う人材育成事業に関連した取り組みとしては、主要国のファンディングエージェンシーの研究者の育成を目的とした諸事業の情報や、各国の研究人材のキャリアパス等の情報の収集を進めている。

さらに、主要国の政府・学術関係機関や国際機関が発表する様々な情報についても幅広く収集を行っており、毎年主要文書数十編を選択し、それぞれに解説を付し「米欧の主要国における学術・科学技術に関する主要文書」という学振内部向けの資料を作成している。この中には各国の基本的な科学技術政策文書、アカデミー等による政策提言を含む報告書、大学協会等による要望書、政府や国際機関が作成した科学技術や研究人材に関する統計資料などが含まれている。

各国の学術研究の動向や学術振興機関の事業の実施状況に関する情報の収集を通して明らかとなっていることは、各国が科学技術政策やイノベーション政策を推進する中で学術研究活動がその基盤に位置付けられていることである。研究者の自由な発想に基づくボトムアップ的な研究に対する支援はいずれの国においても学術振興機関の事業の中核に据えられ、また、個々の研究に対する支援の決定は専門を同じくする研究者のピア・レビューを通して行われるということも共通である。

しかしながら、各国の状況を広い視野でみた場合、我が国との間で様々な相違点があることも事実である。例えば大学に対する政府の基盤的経費と競争的研究資金の支出の考え方については国により大きな違いがある。この違いは高等教育システムの発展の歴史や国による大学への関与の考え方の相違などが背景となっているが、そのような多様な背景の理解に基づき、我が国の学術研究システムにおいて最も相応しい支援の方策を学振が検討するための助けとなるような情報を収集したいと考えている。

F. 文献データベースを利用した学振の事業による支援の効果の分析（遠藤分析研究員）

学術誌に掲載される論文は、著者の研究の成果としての学術的な知見が記されていると同時に、論文に含まれる著者の情報や引用に関する情報により、学術研究の動向を理解する手掛かりを与えてくれる。そして、論文に含まれる著者情報や引用情報は、民間企業が提供する文献データベースを利用することで定量的な分析が可能となっている（いわゆる科学計量学と呼ばれる研究分野である）。

公的機関によるこの類の分析例としてこれまでも我が国においては科学技術・学術政策研究所による Thomson Reuters 社の Web of Science 等のデータを用いた分析があるほか⁸、海外のファンディングエージェンシーにおいても文献データベースを利用して支援の成果の分析を行っている事例が見られ、例えば英国において競争的研究資金を配分しているリサーチカウンシルのひ

⁷ http://www.jsps.go.jp/j-cgsi/chousa_bunseki.html

⁸ 例えば「サイエンスマップ 2010 & 2012」<http://data.nistep.go.jp/dspace/handle/11035/2933>

とつである医学研究会議（Medical Research Council: MRC）は、支援を行った成果論文の被引用数のデータを分析し、MRCによる支援の効果について報告している⁹。情報センターでも、文献データベースを利用した事業の支援効果の分析を試みており、昨年度にはCGSI レポート第1号「Scopus 収録論文との関連付けを通じた科研費成果論文の分析」を刊行し、科学研究費助成事業による支援が我が国の学術研究活動の向上に寄与している状況について報告した。

本年度は Elsevier 社が提供する文献データベース Scopus 及び分析ツール SciVal を学振内で利用できる環境を整えた上で、研究者個人を単位とした研究成果の分析として、人材育成事業の中でも海外特別研究員を対象とした分析を試行した。具体的には、平成 18 年度採用の海外特別研究員を対象とし、採用期間中及び採用期間終了後の論文発表を通じた研究活動の変化について分析を試みた。Scopus においては、研究者一人一人を同定した著者 ID（Author ID）¹⁰の機能があることから、これを利用して個々の海外特別研究員の論文から得ることができる共著者、所属機関、被引用の情報を取りまとめた。結果として、対象者全員の平均を見ると派遣期間中及び派遣期間終了後に発表した論文に被引用数の増加が見られ、国際共著論文の割合も増加していることが確認された。ただし、対象となった研究者の数が 129 人と比較的少数であり、分野・個人により論文発表の傾向が大きく異なる等の理由から単純平均では正しい傾向が見えない恐れがあることから、海外特別研究員事業の効果を検証するためには更なる分析が必要と思われる。

4. 今後の調査・分析方針

ここまでで紹介してきた本会事業にかかる調査・分析内容のうち、科研費審査等に関する試行的なものについては、今後実用に耐えうるような段階まで分析手法の検討と吟味を重ねることが情報センターにとって現在最も重要な課題と考えている。併せて、科研費以外、特に特別研究員事業に関するデータの収集や分析（特に 2 ページで述べた審査データに関する分析）も、科研費データの調査・分析で得られたノウハウを生かして進めることができるものと期待している。

今後は、調査・分析案件の選定についてシステムセンターや各事業担当部署とこれまで以上に連携を図り、分析結果の利活用を見越して、意味のある調査・分析を進めたいと考えている。

⁹ 例えば MRC, Economic Impact Report 2011/12, p10

<https://www.mrc.ac.uk/publications/browse/economic-impact-report-2011-12/>

¹⁰ この著者 ID の機能は Elsevier 社独自のアルゴリズムにより著者の名寄せが行われたもので、個人単位の文献情報を得るには便利なものだが、同社の説明 (http://help.scopus.com/flare/ja_JP/Content/h_autsrch_intro.htm) にもあるように論文の著者情報が不十分な場合がある等の理由により完全に著者が同定されたものではない。このため、今回の調査では目視による確認を含めた作業を行った。

独立行政法人日本学術振興会グローバル学術情報センター
CGSI レポート 第4号
平成28年3月28日発行

独立行政法人日本学術振興会グローバル学術情報センター
〒102-0083 東京都千代田区麹町 5-3-1 麹町ビジネスセンター
電話：03-3263-1971 電子メール：cgsi@jsps.go.jp

