

レポート

客観的根拠（エビデンス）に基づく政策のための データ・情報基盤（第二回） ～ NISTEP 大学・公的機関名辞書～

科学技術・学術基盤調査研究室 客員研究官 小野寺 夏生、室長 伊神 正貴
第2研究グループ 総括主任研究官 富澤 宏之

概要

「客観的根拠（エビデンス）に基づく政策のためのデータ・情報基盤」シリーズの第二回として、本レポートでは、「大学・公的機関における研究開発に関するデータ整備」の中核となる「NISTEP 大学・公的機関名辞書」を取り上げる。機関レベルで研究活動の状況把握を行うための基本的情報を継続的に整備する目的を果たすため、辞書に収録する情報とその特徴、絶えず変遷するそれらの情報の追跡・収集方法、科学技術・学術政策研究所（NISTEP）のホームページにおいて公開しているデータの内容について述べる。最後に、その活用について、具体例を含め解説する。

キーワード：政策のための科学，大学，公的機関，機関名辞書，名寄せ

1. 大学・公的機関における研究開発に関するデータの整備

「客観的根拠（エビデンス）に基づく政策のためのデータ・情報基盤」シリーズでは、「NISTEP データ・情報基盤」で構築・公開されている主要なコンテンツを紹介している。第一回の NISTEP 企業名辞書¹⁾に続いて、この第二回では NISTEP 大学・公的機関名辞書を取り上げる。

「NISTEP データ・情報基盤」の中の「大学・公的機関における研究開発に関するデータ整備」²⁾は、個別機関（及びその組織）、セクター、国などの各レベルで研究開発の実態把握を行うための基礎となるデータの整備を行うものである。機関レベルで自機関と他機関の論文等の比較を行いたいと考えているリサーチ・アドミニストレーター、機関やセクターレベルで政府の研究開発投資の成果や研究開発システムに与える影響を定量的・構造的に分析したいと考えている研究者等による活用を想定している。

本稿では、その中核的役割を担う NISTEP 大学・公的機関名辞書（以下「機関名辞書」という）に収録している情報、各種データの公開や活用方法について述べる。

2. 機関名辞書とは

2-1 機関名辞書の必要性と課題

我が国における研究開発の動向を把握し、研究開発に関する方針や政策を検討するための情報源としては、様々なデータが存在する。しかし、これらの情報源を用いて機関レベルにデータを整理・分析しようとすると、機関名表記のゆらぎ、下部組織情報の不足、機関の変遷の把握の困難さ、セクター情報の不足などの問題点に直面する。

機関名辞書の構築と公開は、以上の問題点に対処するための一つの回答であり、研究開発を行う国内の主要な機関についての基本的情報を系統的・継続的に取得し、アーカイブ化することを目標としている。機関名辞書の目的や特徴については、別途詳しい報告がある^{3,4)}。

2-2 収録する情報とその特徴

「大学・公的機関名辞書」という名のように、大学等（短大、高専、大学共同利用機関を含む）と公的機関（国の機関及び国立研究開発法人等（独立行政法人、特殊法人を含む）を指す）は網羅的に収録するが、研究を行う地方公共団体の機関、民間企業、非営

利法人等もできるだけ含めている。

機関名辞書の特徴として、独立した機関（「代表機関」と呼ぶ）のみならず、主な下部組織も収録対象とする。32の大学については大学直属の下部組織を網羅的に収録するほか、他の大学についても、病院と共同研究・共同利用拠点となっている組織を収録対象とする。このほか、4つの大学共同利用機構の下部研究所、大規模な国立研究開発法人の下部組織、国立病院機構や地域医療機能推進機構等の各病院などを収録する。

もう一つの特徴は、統廃合や名称変更があって非現存となった機関（下部組織を含む）も保持し、継承の機関（組織）がある場合はそれと関係づけをすることである。最近は、大学等、公的機関、企業を問わず、機関や組織の統廃合や改組が頻繁に行われるので、この種の情報は、機関の活動を時系列で追跡するときの困難さを軽減すると考えられる。

現在公開している機関名辞書 ver.2018.2 に収録されている機関数は図表1の通りである。図表2に主な収録情報項目を示す。産学官よりも細かなセクター情報を付与しているので、詳細なセクター別の分析が可能となっている。

機関や組織の変遷に対応するため、きめ細かいデータの更新が必要であり、いろいろな方法でデータの調

査・収集を行っている。大学等と公的機関については、毎年時期を決めて網羅的調査を実施する。代表機関の存続については官公庁のサイト等にあるリストで確認し、下部組織の新設、改廃、名称変更については各代表機関のサイトで調査する。その際英語名の確認も行う。また、論文データベース中の機関名寄せ（3-2参照）において、機関同定ができなかったデータ等を点検し、必要な場合、辞書への機関や組織の追加、英語表記名の追加を行う。民間企業については、NISTEP 企業名辞書との接続（3-2参照）により変遷情報を入力する。地方公共団体の機関や非営利法人については、それぞれ数年に一度網羅的な調査を実施している。

図表1 機関名辞書への収録機関数

セクター	代表機関			下部組織		
	現存	非現存	計	現存	非現存	計
国立大学	86	15	101	1,461	396	1,857
公立大学	89	17	106	76	13	89
私立大学	607	60	667	512	76	588
短期大学	328	335	663			
高等専門学校	57	12	69			
大学共同利用機関	4	3	7	25	1	26
学校法人	661	18	679			
国の機関	54	69	123	65	15	80
国立研究開発法人等	80	87	167	370	154	524
地方公共団体の機関	746	227	973	320	71	391
会社	4,111	743	4,854	15	3	18
非営利団体・その他	3,531	3,659	7,190	85	47	132
計	10,354	5,245	15,599	2,929	776	3,705

注) 短期大学は国立短大、公立短大、私立短大の合計、高等専門学校は国立高専、公立高専、私立高専の合計、非営利団体・その他は非営利団体とその他の合計である。

図表2 機関名辞書の主要な収録情報

収録情報	内容
(1) 機関ID	○NISTEPが独自に与える機関識別ID。
(2) 機関の名称	○日本語の正式名称、英語の正式名称に加え、英語の通称、略称もできるだけ収録。 ○正式名称(Formal)、別名(Alias)、揺らぎ名(Variant)等の名称種別フラグを与える。
(3) セクター	○産・学・官よりかなり細かく、17のセクターに分類。この分類は、科学技術研究調査で用いられているものに近い。 ○これとは別に、病院の機関にはそのことを示すフラグを与える。
(4) 機関の下部組織	○主要な研究機関について一部の下部組織も収録。特に、32の大学の学部・研究科・附置研究所等、4つの大学共同利用機関である機構に属する各研究所、一部の国立研究開発法人等に属する組織は網羅的に収録。 ○代表機関と下部組織の間に関係づけを行う。
(5) 機関の変遷情報	○この15年ほどの間に統廃合、改組、名称変更等があって現存しない機関・組織についても、できるだけ収録。 ○変更のあった日付、継承機関(存在する場合)等の情報も収録。
(6) 外部データとの接続情報	○NISTEP企業名辞書ID、科研費機関番号等を収録。

3. 各種データの公開

3-1 機関名辞書の公開

機関名辞書は、参考文献2) に示すサイトで公開されている。2012年12月に最初の版を公開して以降適宜更新しており、現在はver.2018.2である。また、機関の名称データは、別途「大学・公的機関名英語表記ゆれテーブル」（以下「表記ゆれテーブル」という）として同じサイトから公開しており、公開辞書には和英の正式名のみを含む。詳細については、それぞれの利用マニュアル^{5,6)}を参照されたい。公開して

いる機関名辞書のイメージを、図表3に示す（スペースの都合により一部の項目を省いている）。

3-2 機関名辞書と外部情報源との接続情報の公開

機関名辞書の利用の便宜を図るために外部情報源とのリンク付けを行い、その結果についても公開している。図表4はこの接続のイメージを示したものである。

（論文データベースとの接続テーブル）

機関名辞書を用いて、代表的な引用索引データベースであるWeb of Science Core Collection

(WoSCC) 及び Scopus の著者所属機関データの機関同定 (名寄せ) を継続的に行っている。それぞれのデータベースの提供元の了解を得て、その結果を「機関名辞書 -WoSCC 対応テーブル (ver.2017.1)」、 「機関名辞書 -Scopus 対応テーブル (ver.2018.1)」として、参考文献 2) に示すサイトで公開している。前者は WoSCC に収録された 1998-2015 年の論文データ、後者は Scopus に収録された 1998-2016 年の論文データとの対応をとっている。詳細はそれぞれの説明書^{7,8)}を参照されたい。これらの名寄せの結果は、科学技術・学術政策研究所 (NISTEP) の種々の調査研究に利用されている^{9~11)}。

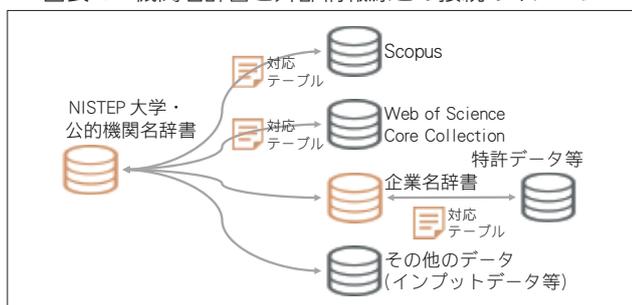
(NISTEP 企業名辞書との連携)

本シリーズの前回記事¹⁾にも述べられているように、機関名辞書は NISTEP 企業名辞書とのリンクをとっている。「NISTEP 大学・公的機関名辞書との接続テーブル (ver.2018.1)」において機関名辞書の機関 ID と企業名辞書 ID の対応を示すほか、それぞれの辞書の中にお互いの ID を収録している。NISTEP 企業名辞書に含まれている情報や、特許データ等との対応テーブルを活用することで、分析の幅を広げることが可能となる。

図表 3 公開している NISTEP 大学・公的機関名辞書 (一部)

機関ID	日本語正式名	英語正式名	セクター分類	機関階層	下部組織種別	現状
NID201200199399167	お茶の水女子大学	Ochanomizu University	国立大学	1		
NID201200434617623	お茶の水女子大学生活環境研究センター	Institute of Environmental Science for Human Life, Ochanomizu University	国立大学	2	全学組織	
NID201200072764185	愛知教育大学	Aichi University of Education	国立大学	1		
NID201200808829421	愛媛大学	Ehime University	国立大学	1		
NID201200488738503	愛媛大学医学部附属病院	Ehime University Hospital, Ehime University	国立大学	2	病院	
NID201200016556519	愛媛大学沿岸環境科学研究センター	Center for Marine Environmental Studies, Ehime University	国立大学	2	全学組織 (拠点)	
NID201200139365669	愛媛大学地球深部ダイナミクス研究センター	Geodynamics Research Center, Ehime University	国立大学	2	全学組織 (拠点)	
NID201200540848841	旭川医科大学	Asahikawa Medical University	国立大学	1		
NID201200296397387	旭川医科大学病院	Asahikawa Medical University Hospital, Asahikawa Medical University	国立大学	2	病院	
NID201200248230458	一橋大学	Hitotsubashi University	国立大学	1		
NID201200829790881	一橋大学経済研究所	Institute of Economic Research, Hitotsubashi University	国立大学	2	研究所 (拠点)	
NID201200341506245	茨城大学	Ibaraki University	国立大学	1		
NID201200894539974	宇都宮大学	Utsunomiya University	国立大学	1		
NID201200942671133	横浜国立大学	Yokohama National University	国立大学	1		
NID201200290658000	横浜国立大学環境科学研究センター	Institute of Environmental Science and Technology, Yokohama National University	国立大学	2	全学組織	No
NID201200463933413	横浜国立大学環境情報学府	Graduate School of Environment and Information Sciences, Yokohama National University	国立大学	2	大学院	

図表 4 機関名辞書と外部情報源との接続のイメージ



4. 機関名辞書の活用事例

機関名辞書は、我が国の主要な研究機関についての基本的な情報を収録した辞書として活用できる。また、3. に述べた接続情報により、外部データと組み合わせた種々の利用も考えられる。ここではその例をいくつか挙げてみる。

4-1 機関名辞書単独の利用

(1) 機関の典拠として

種々の研究データを機関別に分類するときの典拠辞書としての利用がある。英語名を含むので外国の文献にも利用でき、実際にいくつかの機関で活用されている。例えば、筑波大学の池田らが提案している雑誌評価指標 IMD (index for Measuring Diversity) の算出の基になっている著者所属機関の分類に、機関名辞書が利用されている¹²⁾。また、政策研究大学院大学、NISTEP、科学技術振興機構 (JST) が共同開発している SPIAS (研究資金投入と成果のデータを接続してイノベーション・プロセスを分析・可視化するシステム) での機関名寄せにも利用されている。

(2) 下部組織の情報源として

機関名辞書は多数の下部組織も収録している (図表 1 参照)。特に、32 の大学では、収録する下部組織を

次の10種に分類し、階層関係も付けているので、系統的な整理ができる：学部；大学院；学部・大学院統合；専攻科・別科；教員組織；研究所；研究所（拠点）；全学組織；全学組織（拠点）；病院。

(3) 機関・組織の変遷の情報源として

機関や組織の変遷情報（変遷のあった日付、変遷の種別、継承機関）により、複雑な変遷関係をたどることができる。図表5は、機関名辞書に収録される32

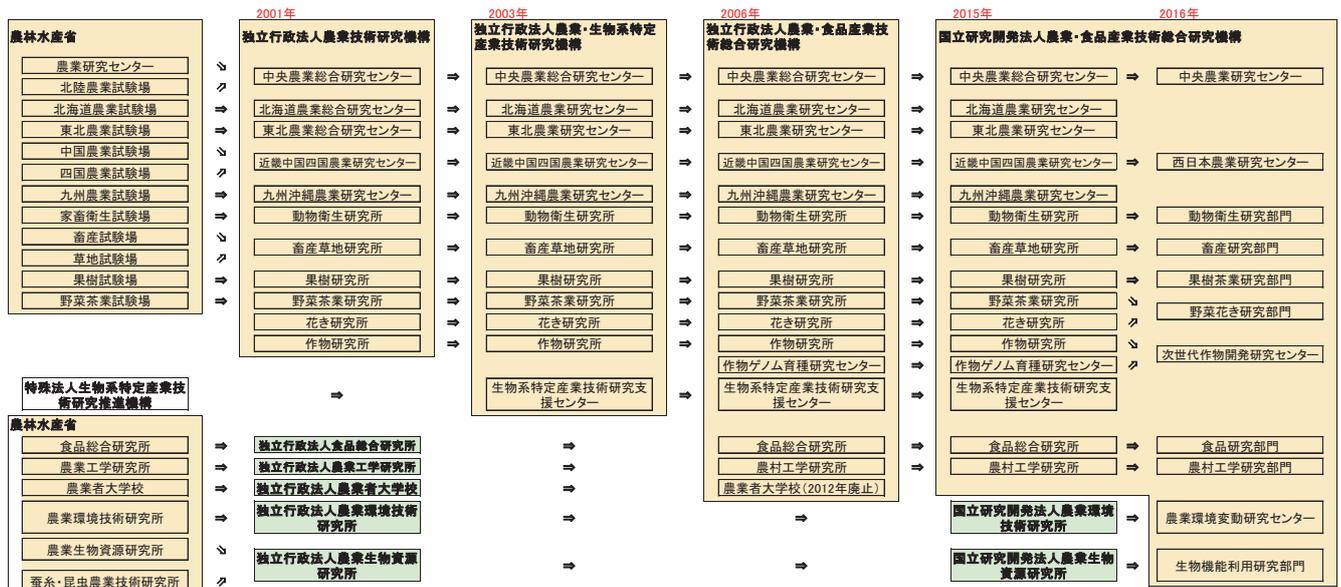
大学の下部組織数を、2013年（下部組織の網羅的収録を開始した年）と2018年の間で比較したものである。増加した組織には、この間に新設されたものだけでなく、それ以前から存在していたものを追加した場合もあるが、それを考慮しても頻繁な組織変更が行われていることが分かる。また、図表6は、国立研究開発法人農業・食品産業技術総合研究機構（NARO）の各研究部門の、農林水産省の研究所であった時代以来の変遷を図示したものである。

図表5 機関名辞書に収録された32大学の下部組織数

大学	2013年版	2018年版	増加率	大学	2013年版	2018年版	増加率	大学	2013年版	2018年版	増加率
北海道大学	69	109	58%	金沢大学	18	94	422%	長崎大学	23	60	161%
東北大学	71	112	58%	信州大学	22	52	136%	熊本大学	24	55	129%
筑波大学	89	115	29%	岐阜大学	19	42	121%	大阪府立大学	21	32	52%
群馬大学	15	37	147%	名古屋大学	51	76	49%	大阪市立大学	19	38	100%
千葉大学	48	71	48%	京都大学	73	90	23%	慶應義塾大学	39	63	62%
東京大学	52	108	108%	大阪大学	69	96	39%	早稲田大学	65	109	68%
東京医科歯科大学	14	39	179%	神戸大学	47	86	83%	東海大学	54	81	50%
東京工業大学	71	101	42%	岡山大学	25	68	172%	東京理科大学	34	62	82%
東京農工大学	13	29	123%	広島大学	52	58	12%	日本大学	70	84	20%
新潟大学	24	48	100%	徳島大学	30	58	93%	近畿大学	51	61	20%
富山大学	26	45	73%	九州大学	111	165	49%	計	1409	2344	66%

出典：NISTEP 大学・公的機関名辞書を用いて筆者が集計（2013年は内部保存版、2018年はver.2018.2）。

図表6 国立研究開発法人農業・食品産業技術総合研究機構（NARO）の研究部門の変遷



出典：NISTEP 大学・公的機関名辞書（ver.2018.2）を用いて筆者が作成。

4-2 対応テーブルの利用、及び外部データと接続しての利用

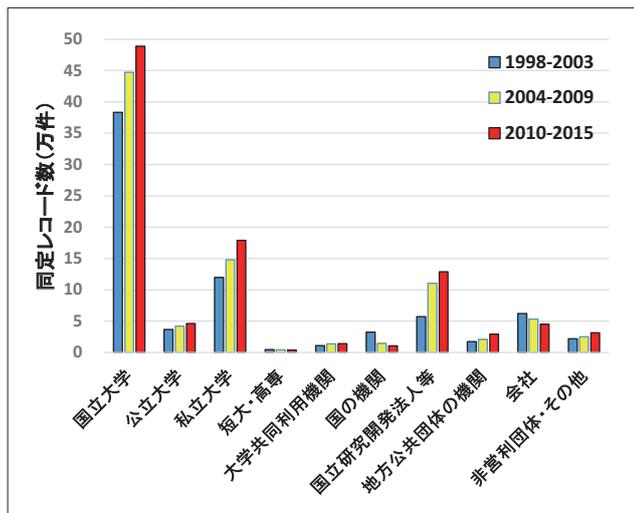
(1) WoSCC、Scopus との対応テーブルの利用（高精度の機関検索と機関成果分析）

3-2 に述べた「機関名辞書 -WoSCC 対応テーブル」又は「機関名辞書 -Scopus 対応テーブル」は、これら単独でも、機関や組織の研究アウトプットの分析がある程度できる。図表7と図表8に、機関名辞書 -WoSCC 対応テーブル（ver.2017.1）を利用した例を示す。図表7は1998～2015年の期間に同定され

た機関をセクター別に集計したものである。対応テーブルのレコード数の単純集計結果によるので、1論文内で同じセクターの異なる機関や組織が複数同定された場合、その数が重複してカウントされている。図表8はよりミクロレベルに大学の組織別同定レコード数を求めた結果で、京都大学の例である。これらはいずれも、6年ずつに区切った各期間の累積レコード数を表す。

更に、これらのテーブルによって、WoSCCやScopusをより精度よく利用することができる注。機

図表7 機関同定レコード数のセクター別集計の推移 (整数カウント)



出典: WoSCC-NISTEP 大学・公的機関名辞書対応テーブル (ver.2017.1) を用いて筆者が集計。図表1と同様にセクターをまとめている。

関(あるいは組織)による検索を行うとき、機関名を入力しなくても、対応テーブルからその機関のIDに紐付いたデータベース記事IDを取り出し、データベースと照合すればよい。この方法により、データベース中の機関名表記のゆれに起因する検索漏れを大部分防ぐことができる。また、検索してダウンロードした結果を機関・組織レベルで分析する際にも、この対応テーブルは有効である。

(2) 企業名辞書との対応テーブルの利用 (特許と論文を結びつけた企業分析)

前号の記事¹⁾に述べられているように、NISTEP 企業名辞書は、企業データ、特許データ等外部の多くの情報源との接続テーブルを持つ。したがって、「NISTEP 大学・公的機関名辞書との接続テーブル」(3-2 参照)を介して、機関名辞書と企業名辞書自体のほか、それらの外部情報源とも接続することができる。図表9は、参考文献1)の図表8に示された新日鐵住金(株)関係の特許出願データに、「機関名辞書-WoSCC 対応テーブル」から得られた論文生産データを組み合わせたものである。特許と論文はおおむね似た推移を示しているが、論文は研究遂行時から遅れて発表されることもあるため、2012年の合併後も旧社名での発表が見られる。

(3) その他の外部データとの接続 (名寄せ・分類等)

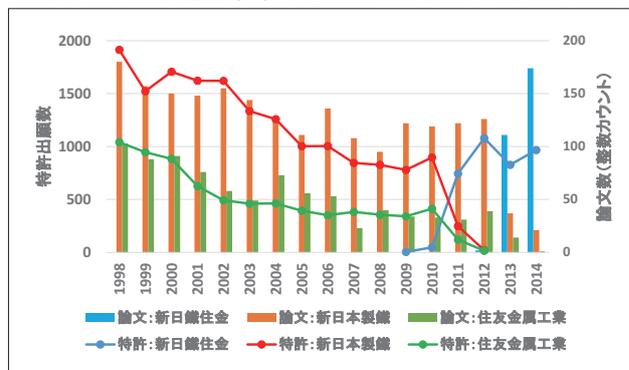
図表4に示したように、機関名辞書を研究インプット情報と結びつければ、より多様な応用が考えられる。

図表8 大学の組織別同定レコード数: 京都大学の例 (整数カウント)

組	組織	1998-2003	2004-2009	2010-2015
学部・大学院	理学部+理学研究科+生命科学研究所	5,023	6,239	6,480
	工学部+工学研究科+情報学研究科	6,087	7,350	7,203
	医学部+医学研究科+附属病院	6,287	7,487	8,883
	薬学部+薬学研究科	1,007	1,345	1,184
	農学部+農学研究科	2,163	2,710	2,525
	総合人間学部+人間・環境学研究科+エネルギー科学研究科	1,581	1,469	1,503
	文系学部+研究科	99	325	432
附置研究所	人文科学研究所+経済研究所+東南アジア地域研究研究所	50	92	132
	数理解析研究所	261	370	457
	基礎物理学研究所	613	839	1,084
	化学研究所	1,609	1,758	1,595
	霊長類研究所	342	409	617
	エネルギー理工学研究所	415	645	561
	複合原子力科学研究所	665	770	608
	生存圏研究所	16	575	663
	防災研究所	300	527	563
	ウイルス・再生医学研究所+iPS細胞研究所	1,306	1,612	1,738
その他	共同利用・共同研究拠点	438	520	702
	全学組織	209	1,034	3,137
	下部組織区分け不能	2,737	2,249	2,354
総計	31,208	38,325	42,421	

出典: WoSCC-NISTEP 大学・公的機関名辞書対応テーブル (ver.2017.1) 及び NISTEP 大学・公的機関名辞書 (ver.2018.2) を用いて筆者が集計。非現存組織は継承組織に統合している。

図表9 新日鐵住金(株)の特許出願数と論文発表数の推移



出典: WoSCC-NISTEP 大学・公的機関名辞書対応テーブル (ver.2017.1) 及び企業名辞書からの提供データを用いて筆者が集計。

例えば、国立情報学研究所が提供する科学研究費助成事業データベース (KAKEN) とリンクさせれば、機関の研究資金取得データを分析することができる。

機関名辞書及び表記ゆれテーブルは、より簡便な方法で、いろいろな情報源に現れる機関の識別、名寄せ、分類等に用いることが可能である。天野らは、生物系実験材料名及びその提供機関名の辞書の作成や、当該辞書を利用しての生物系実験材料の利用影

注 WoSCC の利用にはクラリベイト・アナリティクス社との、Scopus の利用にはエルゼビア社との契約が必要である。

響度の測定に、大学・公的機関名辞書を利用している¹³⁾。また、山下らは、特許における論文引用の所属機関分類に、WoSCC- 機関名辞書対応テーブルを利用した¹⁴⁾。

5. おわりに

NISTEP では、今後 2～3 年間の機関名辞書の目標として、次の 2 つを考えている。一つは、収録する情報（特に下部組織の変遷情報）の一層の充実のため、収録対象機関との連携を進めることである。NISTEP だけの力では情報の収集・確認に限界がある。下部組織情報を網羅的に収録している大学を現在の 32

より広げるには、大学との連携が不可欠と考えられる。今後種々の機関の御協力をお願いしたいと考えている。

もう一つの目標は、データベースとの対応テーブルの作成に用いている名寄せプログラムの公開である。公開することにより、利用機関でそれぞれの目的に応じたカスタマイズをすることも可能となる。この目標に向けて、プログラムのモジュール化等、分かりやすさとユーザビリティの向上のための検討を開始したところである。

この記事をお読みになった方が、機関名辞書の活用に関心を持たれば幸いである。必要な場合は、data-infra@nistep.go.jp に御相談いただきたい。

参考文献

- 1) 中山保夫, 富澤宏之. 客観的根拠(エビデンス)に基づく政策のためのデータ・情報基盤(第一回)～ NISTEP 企業名辞書～. *STI Horizon*, 2018, 4 (2), 47-53. <http://doi.org/10.15108/stih.00134>
- 2) “大学・公的機関における研究開発に関するデータ.” 科学技術・学術政策研究所. <http://www.nistep.go.jp/research/scisip/randd-on-university>
- 3) 小野寺夏生. 大学・公的機関における研究開発に関するデータの整備－マイクロデータ分析への貢献－. NISTEP NOTE No.11. 科学技術・学術政策研究所, 2014年5月. <http://hdl.handle.net/11035/2926>
- 4) 小野寺夏生, 伊神正貫, 阪彩香. NISTEP 大学・公的機関名辞書の整備とその活用－大学下部組織レベルの研究データ分析に向けて－. NISTEP NOTE No.15. 科学技術・学術政策研究所, 2015年10月. <http://hdl.handle.net/11035/3085>
- 5) “NISTEP 大学・公的機関名辞書 (ver.2018.2) 利用マニュアル.” 科学技術・学術政策研究所. 2018年8月. http://www.nistep.go.jp/wp/wp-content/uploads/manual_organization_dictionary_ver_2018.2.pdf
- 6) “大学・公的機関名英語表記ゆれテーブル (Ver.2018.1) 利用の手引き.” 科学技術・学術政策研究所. 2018年3月. http://www.nistep.go.jp/wp/wp-content/uploads/guide_name_variations_ver_2018.1.pdf
- 7) “WoSCC-NISTEP 大学・公的機関名辞書対応テーブル 説明書.” 科学技術・学術政策研究所. 2018年4月. http://www.nistep.go.jp/wp/wp-content/uploads/guide_WoS_NID_corres_table_ver2017_1.pdf
- 8) “Scopus-NISTEP 大学・公的機関名辞書対応テーブル 説明書.” 科学技術・学術政策研究所. 2018年8月. http://www.nistep.go.jp/wp/wp-content/uploads/Guide_Scopus_Table_of_the_NISTEP_Affiliation_Dictionary_ver.2018.1.pdf
- 9) 村上昭義, 伊神正貫. 日本の大学システムのアウトプット構造：論文数シェアに基づく大学グループ別の論文産出の詳細分析. 調査資料 271. 科学技術・学術政策研究所, 2018年3月. <http://doi.org/10.15108/rm271>
- 10) 村上昭義, 伊神正貫. 科学研究のベンチマーキング 2017－論文分析でみる世界の研究活動の変化と日本の状況－. 調査資料 262. 科学技術・学術政策研究所, 2017年8月. <http://doi.org/10.15108/rm262>
- 11) 村上昭義, 伊神正貫, 阪彩香. 論文データベース分析から見た大学内部組織レベルの研究活動の構造把握. 調査資料 258. 科学技術・学術政策研究所, 2017年3月. <http://doi.org/10.15108/rm258>
- 12) Ikeda, J. and Morimoto, Y. Research Metrics for the Rest of Us. *Inter Faculty*, 2016, 95-98.
- 13) 天野晃ほか. 論文マイニングのための生物系実験材料名辞書及び提供機関名辞書の作成. 情報メディア学会第 18 回研究会発表資料, 2-5, 2016-11-26.
- 14) 山下泰弘, 治部眞里. 特許からの論文引用に関する定量的評価指標の検討－日本の研究機関についての試行的分析－. *JAPIO Year Book*, 2017, 144-155. http://www.japio.or.jp/00yearbook/files/2017book/17_2_04.pdf