

## 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（第一回） ～ NISTEP 企業名辞書～

第2研究グループ 客員研究官 中山 保夫、総括主任研究官 富澤 宏之

## 概要

科学技術・学術政策研究所（NISTEP）では、文部科学省の「科学技術イノベーション政策における『政策のための科学』推進事業（SciREX）」の一環として、政策形成を支えるエビデンスの充実のためのデータ・情報基盤の構築を進めている。その成果として、「NISTEP データ・情報基盤」の公開を行っており、様々な研究者等による活用も進んでいるが、潜在的利用者の掘り起こしの余地は大きいと考えられる。ここでは、三回にわたり、「NISTEP データ・情報基盤」の具体的な利用方法について紹介する。第一回は、産業における研究開発・イノベーションに関するデータ基盤の中核となる「NISTEP 企業名辞書」を取り上げ、紹介する。

キーワード：科学技術イノベーション，政策のための科学，企業名辞書，名寄せ

## 1. 客観的根拠（エビデンス）に基づく政策の実現に向けて：NISTEP データ・情報基盤

文部科学省が平成 23 年度（2011 年度）より実施している「科学技術イノベーション政策における『政策のための科学』推進事業（SciREX）」は、“客観的根拠（エビデンス）に基づく政策”の実現に向けて、政策研究の推進や専門人材の育成等を一体的に行う取組である。

その一環として、科学技術・学術政策研究所（NISTEP）は、エビデンスの充実や政策研究の高度化のための「NISTEP データ・情報基盤」の構築・公開を進めている。STI Horizon 誌では、本稿を含めて三回にわたり、「NISTEP データ・情報基盤」の主要なコンテンツを取り上げ、具体的な利用方法について紹介する。第一回に当たる本稿では産業における研究開発・イノベーションに関するデータを分析する際の基礎となる「NISTEP 企業名辞書」を紹介する。今後、第二回では大学・公的機関の研究開発に関するデータを分析する際の基礎となる「NISTEP 大学・公的機関名辞書」、第三回では、科学技術に関する施策

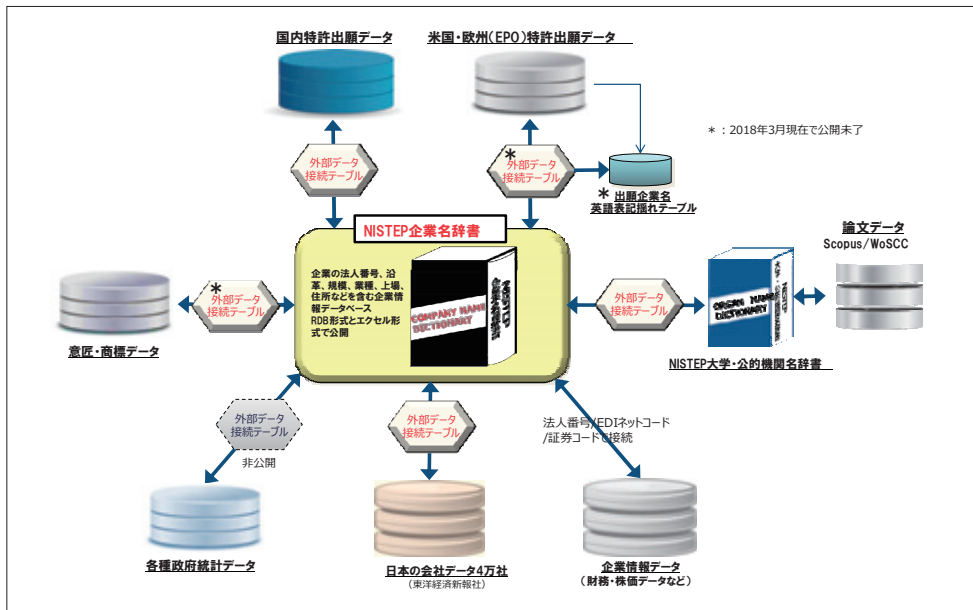
や「NISTEP デルファイ調査」、 「NISTEP 定点調査」に関する各種検索システムを取り上げて紹介する。

## 2. 産業における研究開発・イノベーションに関するデータ・情報基盤

NISTEP は、産業セクターを対象とした研究開発とイノベーションの実態、イノベーションプロセスなどの分析・研究を行うために、国内営利企業（以下、「企業」と呼ぶ）の知的財産、研究開発投資、事業展開等のデータを体系的に関係付けできる「産業における研究開発・イノベーションに関するデータ・情報基盤」の構築を行っている。

この「産業における研究開発・イノベーションに関するデータ・情報基盤」は、図表 1 に示すように、企業情報検索の索引ともいえる「NISTEP 企業名辞書」（以下、企業名辞書と略す）を核として、企業の研究開発に関するインプット、アウトプット及びアウトカムの各データをリンクし、企業レベルでイノベーションの分析・研究を可能としている。

図表 1 「産業における研究開発・イノベーションに関するデータ・情報基盤」の全体構成図



### 2-1 名寄せ問題の解決

産業セクターの研究開発やイノベーションに関するデータを対象とした分析・研究を行う場合、企業名の名寄せという問題が立ちはだかる。この問題には二つの側面がある。一つは多くの企業データには誤記も含んだ企業名表記の揺らぎが存在し、同一企業の判断をする上で障害となるという側面、もう一つは統合、合併又は事業再編等による組織形態の変化や名称の変更、場合によっては清算・倒産などの変遷を考慮して企業データを取扱いつける必要があるという側面である。

前者の例として、特許データの出願人に記載された企業名の表記がある。特に、電子公報発行以前の出願人表記には、片仮名、新字旧字、異称、OCRの精度に起因する誤字等が数多く存在し、特許分析の大きな障害となっていた。このため、膨大な数の企業出願人の表記にクレンジングを施し、その企業名と住所コード及び出願人コード等の判別条件を組み合わせた機械的判定による企業の名寄せに加えて、機械的判定では困難な企業の名寄せを目視作業により行い、この問題の解決を図っている。

後者については、後述する企業名辞書において、現行企業、変遷企業及び被合併企業ごとに当該企業の外部データと接続でき、それら企業を個別に、又はグループ化して関係データの取扱いができる構造とすることで問題解決している。

これにより、本データ・情報基盤の利用者は、企

業名辞書と図表 1 で想定する外部データを接続して利用する限り、名寄せという難しい問題から解放される。

### 2-2 企業名辞書

企業名辞書は、企業の規模、業種、所在地、上場及び親会社、子会社などの企業の基本情報の提供、並びに特許データ・財務データなどの外部データを接続するハブとしての役割を担う。

企業名辞書では、以下の条件のいずれかに該当

することを企業掲載の基準としており、最新の企業名辞書<sup>注1</sup>では企業数 10,064 社、それらの変遷企業及び被合併企業も含む全企業レコード数は 20,703 を数える。なお、下記①及び③に該当する企業の抽出には、前述の企業名の名寄せによる成果を使用している。

- ①特許出願数累積 100 件以上 (1970 年以降)
- ②株式上場企業 (2012 ~ 2017 年に上場している企業)
- ③特許出願数の伸び率大 (3, 5, 7 年の各期間で評価)<sup>注2</sup>
- ④「NISTEP 大学・公的機関名辞書」掲載企業 (論文執筆者所属企業)

### 2-3 外部データとの接続テーブル

外部データとの接続テーブル (以下、「接続テーブル」と呼ぶ) は、企業名辞書に登録した企業と特許データ等の外部データに収められた当該企業 (特許の場合は出願企業) との接続を実現するテーブルであり、データ間の架け橋の役割を持つ。また、接続テーブルは、外部データの企業名称の表記揺れを吸収する役割も併せ持つ。例えば、前記の特許データの出願人表記に揺れがあるものの同一企業と判定された企業は、接続テーブルにより企業名辞書の正式企業名称レコードと接続され、表記揺れの吸収が図られている。

注 1 2018 年第 3 四半期に公開予定の企業名辞書 Ver2018\_2 における掲載企業数である。本稿執筆時点において精査中の段階にあり、公開までに若干変動する可能性がある。

注 2 伸び率の評価方法は、以下の URL に掲載する「企業名辞書利用マニュアル」を参照されたい。  
<http://www.nistep.go.jp/research/scisip/rd-and-innovation-on-industry>

### 3. 企業名辞書の利用形態

イノベーション分析・研究における企業名辞書の利用形態は大きく二つに分けられる。

図表 2 に示すように、一つは企業の様々な情報を取得する辞書としての利用であり、もう一つは、企業名辞書の特徴である外部データ接続の容易性を活用し、特許、意匠、商標、論文、財務、株価などの外部データと企業名辞書の企業情報とを連携させ、より高度な分析を行うものである。以下に幾つかの利用方法について説明する。

なお、企業名辞書は、企業情報をパネルデータとして保有することや、データの保守性などの観点からリレーショナルデータベース（以下、「RDB」と呼ぶ）として作成し公開している。しかし、利用者全て

が RDB の操作に慣れている訳ではなく、機能が削られても一つのテーブルで企業情報を見通したいというニーズもあるため、単一表形式（Excel 形式）でも提供している。

#### 3-1 企業情報を知る辞書として利用

##### (1) 企業の変遷

企業名辞書はレコード単位で現行、過去及び被合併法人の企業名称とそれらと関係付ける沿革データを保有しており、容易に企業の変遷を知ることができる。

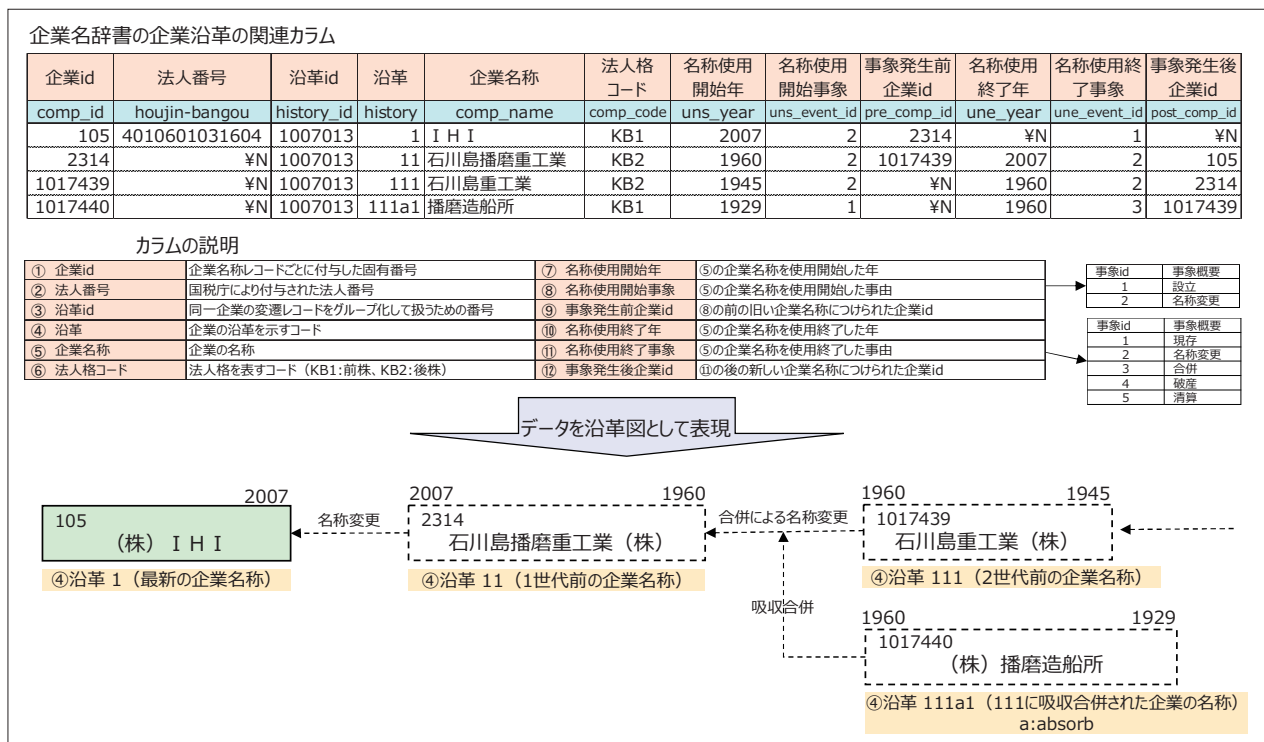
図表 3 は（株）IHI を例に企業名辞書の沿革データを取り出したものである。例えば、データテーブルのカラム'企業 id'<sup>注3</sup>の値が 105 である（株）IHI のレコードは、「（株）IHI という名称は石川島播磨重工（株）〔'事象発生前企業 id'=2314〕が名称変更〔'名称使用開始事象'=2〕し、2007 年〔'名称使用開始年'=2007〕から使用した名称であり、その名称で企業活動中〔'名称使用終了事象'=1〕である」ことを意味する。

現行、過去及び被合併法人の複数レコードは同一番号

図表 2 企業名辞書の利用形態

(1) 企業名辞書を単独で利用	①法人番号・EDI ネットコード・証券コードなど企業を一意的に特定するコード ②企業の変遷名称、合併等の沿革情報 ③住所・住所コード・緯度経度などの所在地情報 ④企業規模・業種などの企業属性情報 ⑤企業の親子関係 ⑥上場企業・上場日・上場廃止日など株式上場情報 ⑦1970年以降特許出願累積数100件以上の企業 など
(2) 企業名辞書と他のデータと連結して利用	①特許出願データと接続して企業の特許出願状況を分析する ②意匠・商標登録データと接続して企業の意匠・商標登録状況を分析する ③財務諸表・株価データ等の企業情報と接続して企業分析する ④論文データと接続して企業の論文算出状況を分析する など

図表 3 企業の沿革を知る



注 3 シングルクォーテーションで囲った文字列はデータテーブルのカラム名を示す。



の'沿革id'でグループ化されており、図表3の'沿革id'=1007013である4レコードを沿革図として表現すると、同図表の下図ようになる。

また、'沿革'の値である沿革コードは変遷を簡易的にコードとして表現したものであり、その意味は図表4を参照されたい。

## (2) 企業の規模・業種

図表5に示すように、企業名辞書は企業の規模や業種のデータも保有しており、目的とする企業の情報取得や分析におけるカテゴリーデータとして利用できる。

企業の規模は、中小企業基本法の定義に基づく業種・資本金・従業員数を考慮した分類、資本金階級による分類及び従業員数階級による分類の3種類のデータを保有している。また、業種は、企業の主業について、日本標準産業分類に基づいた大・中・小の3分類に加えて、上場企業については証券コード協議会による業種分類を保有している。

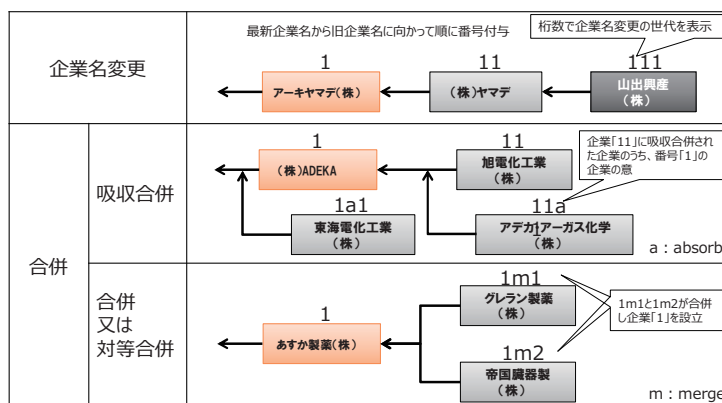
企業の規模や主業は企業活動とともに変化する場合もあり、RDB形式の企業名辞書では調査日ごとにデータをパネル化して保有できるよう考慮している。簡易版の単一表形式では最新調査日のデータが表示される。

## (3) 上場企業

企業名辞書には2012年時点で株式上場していた全企業と、以降、企業名辞書の改訂ごとに新規又は再上場した企業を追加掲載している。その間、上場を廃止した企業もあることから、RDB形式の企業名辞書では証券コード(4桁の銘柄コード)、上場市場、上場日、上場廃止日及び証券コード協議会による業種分類の各情報を提供し、ある時点における上場企業を絞り込んで取扱いすることができる。

図表6はExcelの単一表形式の企業名辞書にオートフィルタ機能を使い、現行活動を行っている企業

図表4 沿革コードの説明



- 注:
- 企業名変遷の世代番号は、現在の名称から遡る形態で付与しているため、通常の世代番号とは逆になっている
  - 複数企業の合併において、存続・消滅企業が明確な場合、存続企業が世代番号の継承を行う

図表5 企業の規模・業種を知る

企業名辞書の企業規模・業種の関連コラム										
企業id	法人番号	企業名称	法人格コード	企業規模	資本金階級	従業員数階級	日本標準産業分類コード	日本標準産業分類(大分類)	日本標準産業分類(中分類)	日本標準産業分類(小分類)
comp_id	houjin-bangou	comp_name	comp_code	comp_size_law	comp_size_cap	comp_size_emp	jsic_code	jsic_l	jsic_m	jsic_s
1	5120001158218	パナソニック	KB2	大企業	10億円以上	5,000人以上	jsic293	製造業	電気機械器具製造業	民生用電気機械器具製造業

セル値の選択肢		
大企業	100万円未満	5人未満
中小企業	100万円以上	5~29人
小規模企業	1000万円以上	30~99人
	2000万円以上	100~299人
	5000万円以上	300~999人
	1億円以上	1,000~4,999人
	10億円以上	5,000人以上

総務省の日本標準産業分類を記した下記URLを参照  
[http://www.soumu.go.jp/toukei\\_toukatsu/index/seido/sangyo/02toukatsu01\\_03000044.html#](http://www.soumu.go.jp/toukei_toukatsu/index/seido/sangyo/02toukatsu01_03000044.html#)

図表6 最新の上場企業を知る

沿革id	沿革	企業名称	法人格コード	名称使用終了	証券コード	証券コード	証券コード	証券コード	企業規模	資本金階級	従業員数階級	日本標準業種	日
you	history_id	history	comp_name	comp_cc	une_event	sec_code	class	s_class	comp_si	comp_si	comp_si	jsic_code	jsic
17640	8002467	1 メディカル・データ・ビジョン	KB2		1	3902	運輸・情報通情報・通信業	大企業	1億円以上	100~299人	jsic391	情報	
6966	8000679	1 K Hネオケム	KB2		1	4189	製造業	化学	大企業	10億円以上	300~999人	jsic163	製
1392	8000448	1 オプティム	KB1		1	3694	運輸・情報通情報・通信業	大企業	1億円以上	100~299人	jsic391	情報	
5907	8000359	1 エムケイシステム	KB1		1	3910	運輸・情報通情報・通信業	中小企業	1億円以上	30~99人	jsic391	情報	
1476	8000277	1 エクストリーム	KB1		1	6033	サービス業	サービス業	大企業	1億円以上	100~299人	jsic391	情報
1663	1009997	1 ヘルーナ	KB1		1	9997	商業	小売業	大企業	10億円以上	300~999人	jsic611	卸
5153	1009996	1 サトー商会	KB1		1	9996	商業	卸売業	大企業	10億円以上	300~999人	jsic522	卸
6937	1009995	1 ルネサスイーストン	KB1		1	9995	商業	卸売業	大企業	10億円以上	300~999人	jsic541	卸
4536	1009994	1 やまや	KB1		1	9994	商業	小売業	大企業	10億円以上	300~999人	jsic585	卸
2171	1009993	1 ヤマザワ	KB1		1	9993	商業	小売業	大企業	10億円以上	300~999人	jsic581	卸
4707	1009992	1 理研グリーン	KB1		1	9992	商業	卸売業	大企業	10億円以上	300~999人	jsic532	卸
		1 マックス	KB1		1	9991	商業	卸売業	大企業	10億円以上	300~999人	jsic532	卸

（‘名称使用終了事象’=1）と証券コードの記載あり（‘証券コード’<>¥N）を条件として最新の上場企業<sup>注4</sup>を抽出した例である。

#### (4) 企業の所在地情報の利用

NISTEP では、国立大学の研究者による発明に基づく特許出願の状況を網羅的に調査した結果を報告<sup>1)</sup>している。

図表7は、その報告から、名古屋大学を例とした特許出願に至る共同発明を行った特許出願累積数（2004～2012年度）の上位企業の情報を使い、その所在地を地図上に表示したもので、地域内の産学連携状況についても直感的に知ることができる。

図表7では、企業名辞書の企業所在地の緯度・経度情報を利用して描画<sup>注5</sup>しているが、企業名辞書ではこのほかにも、企業の所在地情報として、住所、住所

コード（都道府県レベル（2桁）、市区町村レベル（6桁）、国土交通省コード（12桁））が提供される。

なお、企業の所在地は、原則本社の住所としているが、本店、事業所等の住所も提供される。また、移転により所在地が変わることも数多くあり、所在地を企業名寄せにおける同一企業判定の要素とするためには複数の所在地情報を必要とする。このため、RDB形式の企業名辞書では所在地情報もパネル化できる構造としている。

#### (5) 企業の変遷に沿った特許出願状況の再現

外部データである特許データを用いなくても、企業名辞書と国内特許出願データ対応の接続テーブルのみで企業の変遷に沿った特許出願状況が再現できる。

接続テーブルは、‘企業id’、‘出願番号\_出願人順序’、‘出願日’の3つのカラムを有し、‘企業id’と‘出願番号\_出願人順序’は

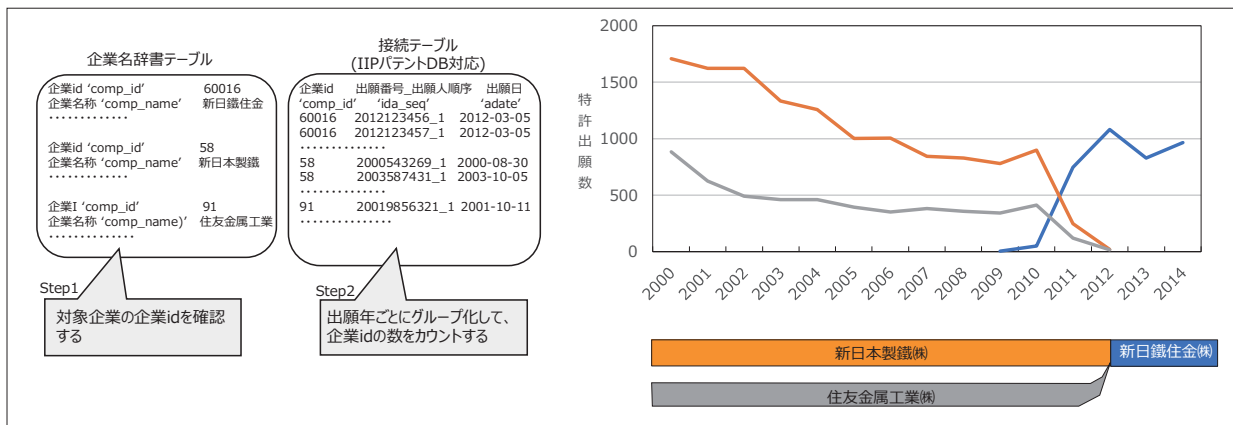
それぞれ企業名辞書と特許出願人テーブルとの接続情報となる。特許の出願人は、原則、公表の公開時点の出願人であるため、企業名辞書の変遷に沿った企業名と接続され、出願日を考慮して接続テーブルの‘企業id’を数えることにより企業の変遷に沿った特許出願状況の再現が可能となる。

図表8は新日鐵住金

図表7 国立大学と共同発明を行った企業の所在地（例：名古屋大学）



図表8 企業の変遷に沿った特許出願状況の再現



注4 厳密には、単一表形式に上場廃止日を含めておらず、上場廃止企業が若干数混在する。

注5 埼玉大学教育学部谷謙二研究室で公開する「Google Maps API v3 を使ったポイント地図化」を用いて描画した。  
<http://ktgis.net/gcode/lonlatmapping.html>

(株) について、2000 年からの特許出願状況を示したものである。同社は 2012 年に新日本製鐵(株)と住友金属工業(株)が合併し発足した。図表 8 には発足以前から新日鐵住金(株)名義の出願が見られるが、それらは設立時点で公開未了であった出願について、遡って出願人の名義変更が行われた結果と考えられる。

### 3-2 外部データと接続して利用

企業名辞書と外部データとの接続は接続テーブルを経由して、又は、市販の財務データやその他の企業データに法人番号、EDINET コード又は証券コードが付されている場合は企業名辞書と直接接続することができる。

法人番号は、国税庁により法人ごとにユニークな番号(13桁)が指定され、個人番号(マイナンバー)と異なり原則公表されるので自由に利用できる。このため、同一企業のデータを接続して利用する際の標準番号として今後利用範囲が広がるであろうが、番号法施行日(平成 27 年 10 月 5 日)以前に消滅した企業は、当然、番号指定はなく、過去に遡った企業データを取り扱う場合はそれだけで十分とはならず他の方法が必要になる。

企業名辞書は、外部データと接続して両者のデータを分析に利用するだけでなく、外部データ接続のハブとして適用し、特許データ、財務データ、意匠・商標データなどを組み合わせて分析することもできる。

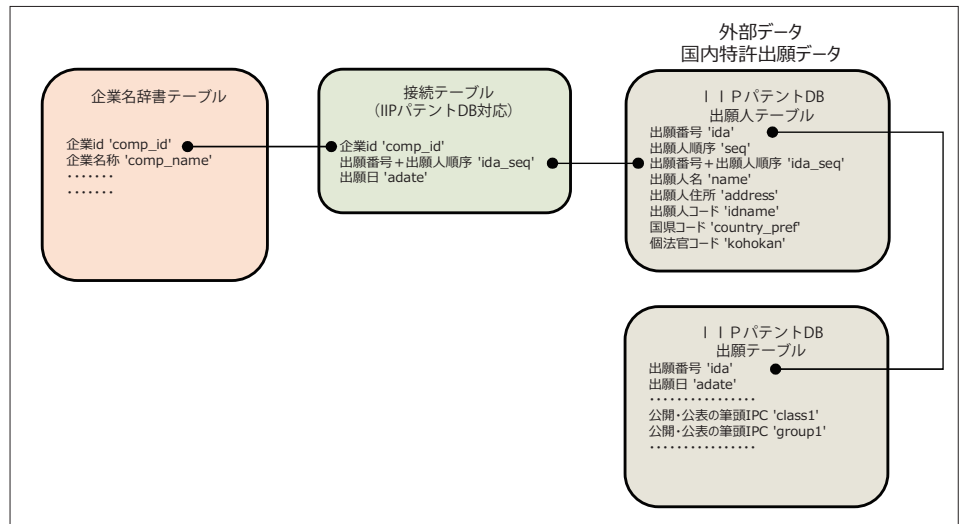
ここでは、国内特許出願データである IIP パテント DB<sup>注 6</sup>を外部データとして企業名辞書と接続し分析を行った事例を以下に示

す。IIP パテント DB との接続は、図表 9 に示すように専用の接続テーブルを介して、IIP パテント DB の出願人テーブル(applicant table)と接続する。大規模データである特許データを前処理なしに直接分析に供することは本来容易な作業ではないが、企業名辞書が介在することで極めて容易に分析が行える。

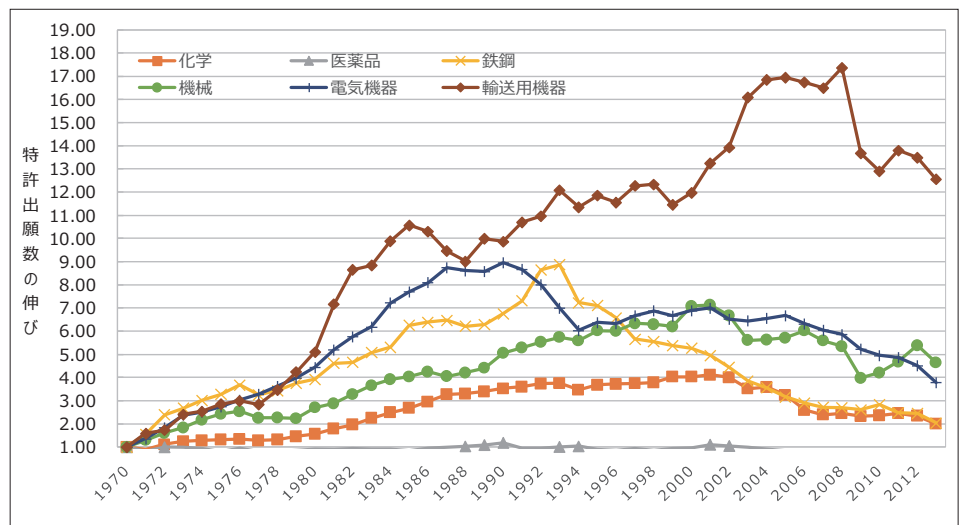
#### (1) 分析事例：業種別の特許出願数の伸び

図表 10 は、3-1 (3) で例示した最新の株式上場企業を対象に、証券コード協会の業種分類(小分類)をカテゴリーデータとして、1970 年の業種ごとの特許出願数を基準に 6 業種について年ごとの出願数の伸びを表示したものである。

図表 9 企業名辞書と IIP パテント DB との接続方法



図表 10 業種別の特許出願数の伸び

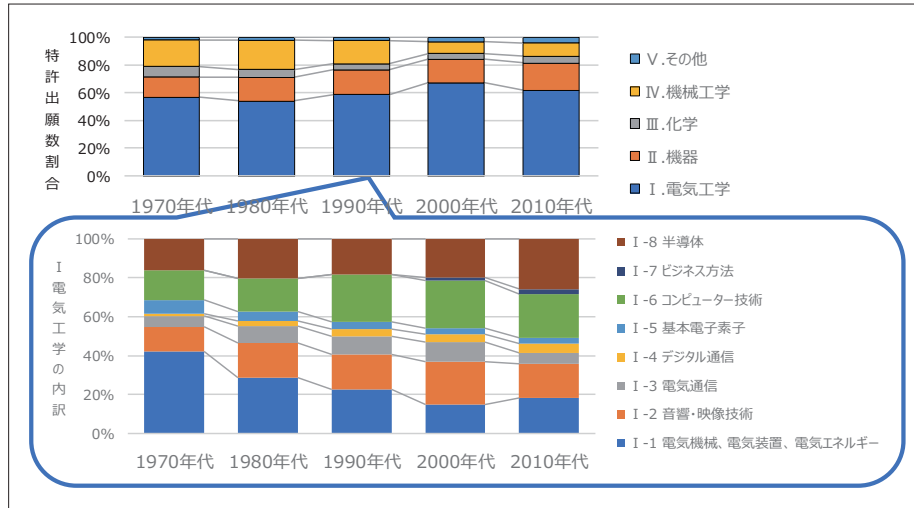


注：図表 10 は 6 業種について 1970 年の特許出願数を 1 と表示した

注 6 (一財) 知的財産研究所より公開される特許庁の整理標準化データをもとに特許統計分析用に開発されたデータベースである。本稿執筆時点の IIP パテントデータベースの最新版は 2017 年版であり、同研究所のホームページより入手できる。  
(<http://www.iip.or.jp/>)



図表 11 ある電気機器メーカーの特許出願技術領域の変化



こうした図表は、例えば景気動向、株価等の経済指標と組み合わせて考察するなど、新たな知見を取得するための材料としても使える。

## (2) 分析事例：発明の技術領域の変化

図表 11 は、ある電気機器メーカーが特許出願した発明について、年代ごとに技術領域がどのように変化したかを示している。ここでは、図表 9 のように、出願人テーブルに加えて、発明の技術分類である IPC (International Patent Classification：国際特許分類) データが記載された出願テーブル (ap table) を接続して利用している。図表 11 下図注<sup>7</sup> は、I 電気工学分野の細目を示しているが、いわゆる、重電領域から半導体、計算機領域へと比重が移っている様子がうかがえる。

## 4. おわりに

企業名辞書と接続テーブルは、以下の NISTEP のホームページにて公開しており、記載する条件で使用する限り、誰でも利用することができる。

<http://www.nistep.go.jp/research/scisip/rd-and-innovation-on-industry>

企業情報は常に変化しており、企業名辞書が陳腐化しないよう情報を収集し、一定の品質を担保するには多くの手間を必要とする。外部データも特許データのように年間数十万件のデータが追加されるものもあり、逐次接続情報を追加することが必要になる。企業名辞書の維持に行うべきことはたくさんあり、そのフォローは大変ではあるが、科学技術イノベーション政策の形成に企業名辞書が活用され、貢献することができれば幸甚である。

注 7 IPC を WIPO (World Intellectual Property Organization：世界知的所有権機関) の「IPC - Technology Concordance Table3」で変換した技術分野で描画している。

## 参考文献

- 1) 科学技術・学術政策研究所、『国立大学の研究者の発明に基づいた特許出願の網羅的調査』, 調査資料 -266, 科学技術・学術政策研究所, 2017 年 12 月.  
DOI: <http://doi.org/10.15108/rm266>