NISTEP企業名辞書（Ver.2016.1）
利用マニュアル

文部科学省科学技術・学術政策研究所

（2016年8月）

目次

[はじめに 1](#_Toc459885976)

[１．公開するファイルの概要 1](#_Toc459885977)

[１．１ NISTEP企業名辞書 1](#_Toc459885978)

[（１）企業名辞書掲載企業 2](#_Toc459885979)

[（２）企業名辞書における変遷名称の取り扱い 3](#_Toc459885980)

[（３）企業名辞書に掲載した企業情報 3](#_Toc459885981)

[１．２ 外部データ接続テーブル 4](#_Toc459885982)

[２．公開するファイルの内容 1](#_Toc459885983)

[２．１ 企業名辞書 1](#_Toc459885984)

[（１）テーブルの構成 1](#_Toc459885985)

[①リレーショナルデータベース型企業辞書 1](#_Toc459885986)

[②単一テーブル型企業名辞書 2](#_Toc459885987)

[（２）テーブルのフィールド定義 2](#_Toc459885988)

[２．２ 外部データ接続テーブル 10](#_Toc459885989)

[（１）IIPパテントデータベースとの接続用 10](#_Toc459885990)

[（２）NISTEP大学・公的機関名辞書との接続用 10](#_Toc459885991)

[（３）日本の会社データ4万社データとの接続用 10](#_Toc459885992)

[３．外部データベース 11](#_Toc459885993)

[３．１ IIPパテントデータベース 11](#_Toc459885994)

[３．２ 日本の会社データ4万社 11](#_Toc459885995)

[４．公開ファイルのダウンロードとデータベースの構築 11](#_Toc459885996)

[４．１ ダウンロード 11](#_Toc459885997)

[４．２ データベースの構築 11](#_Toc459885998)

[５．企業名辞書の作成と名寄せ 12](#_Toc459885999)

[５．１ 企業名辞書の作成 12](#_Toc459886000)

[５．２ 企業名の変遷を考慮した名寄せ 13](#_Toc459886001)

[おわりに 14](#_Toc459886002)

[【補足】 16](#_Toc459886003)

[（１）沿革（history）を表現するコード 16](#_Toc459886004)

【改訂履歴】

(1)Ver.2013.1（2013年11月）：NISTEP企業名辞書、各種接続テーブルとともに本書を初公開した。

(2)Ver.2014.1（2014年6月）：以下の改訂を行った。

・NISTEP企業名辞書;重複企業削除、誤記修正等のマイナー改訂を行った。

・IIPパテントデータベースとの接続テーブル; テーブル構造上、iipdb20140417との接続が不可となった旧接続テーブルを改訂し接続可能とした。

・利用マニュアル;上記に関連する記述の改訂を行った。

(3)Ver.2014.2（2014年11月）：以下の追加・改訂を行った。

・NISTEP企業名辞書;

・2013年1月から2014年4月の期間に上場した企業でVer.2014.1に未掲載の企業を追加した。

・変遷情報の掲載未了であった上場企業約2,400社に関し、情報の追加を行った。

・以下の項目を追加し、各企業に関する情報を追加した。沿革年（設立・名称変更・合併）、企業英語名称（上場企業のみ）、所在地大字町丁目コード、過去の上場実績

・掲載全企業に関して、名称変更、合併等2014年4月現在の状況を反映し見直しした。

・IIPパテントデータベースとの接続テーブル; NISTEP企業名辞書（Ver.2014.2）とiipdb20140417とを接続するテーブル可能とした。

・利用マニュアル;上記に関連する記述の改訂を行った。

(4)Ver.2015.1（2015年10月）：以下の追加・改訂を行った。

・NISTEP企業名辞書;

・従来と同様の単一テーブル型企業名辞書に加えて、リレーショナルデータベース型企業名辞書も作成した。

・IIPパテントデータベース（iipdb20140417）から、特許出願件数（累積100件超）の企業を再評価し、新たに100件超となった企業など企業名辞書に未掲載の企業を追加した。

・iipdb20140417から、特許出願件数伸びを3、5、7年の各期間で評価し、企業名辞書に未掲載の伸びの大きい企業を追加した。

・2014年4月から2015年3月の期間に上場した企業を追加した。

・掲載全企業に関して、名称変更、合併等2015年4月現在の状況を反映し見直しした。

・IIPパテントデータベースとの接続テーブル; NISTEP企業名辞書（Ver.2015.1）とIIPパテントデータベース（2015年版）とを接続するテーブルに追加・改訂した。

・利用マニュアル;上記に関連する記述の改訂を行った。

(5)Ver.2016.1（2016年8月）：以下の追加・改訂を行った。

・NISTEP企業名辞書;

・NISTEP大学・機関名辞書（ver.2015.1）に掲載された科学論文著者の所属企業のうち、NISTEP企業名辞書（Ver.2015.1）に未掲載の企業を追加登録（約2,000社）した。

・IIPパテントデータベース（2015年版）から、特許出願件数（累積100件超）の企業を再評価し、新たに100件超となった企業を追加登録した。

・IIPパテントデータベース（2015年版）から、特許出願件数伸びを3、5、7年の各期間で再評価し、企業名辞書に未掲載の伸びの大きい企業を追加登録した。

・2015年4月から2016年3月の期間に上場した企業を追加登録した。

・掲載全企業に関して、名称変更、合併等の沿革を2016年4月現在の状況を反映し見直しした。

・外部データベースとの接続テーブル;

・IIPパテントデータベース（2016年版）とNISTEP企業名辞書（Ver.2016.1）との接続情報を追加・改訂した。

・新たに、NISTEP大学・公的機関名辞書との接続テーブルを作成し、企業レベルで論文（機関IDで接続）と特許（企業idで接続）の両者に接続できる構造とした。

・利用マニュアル;上記に関連する説明の改訂を行った。

# はじめに

文部科学省科学技術・学術政策研究所（以下、NISTEPと呼ぶ）は、NISTEP企業名辞書、および、同辞書と外部データとを繋ぐ接続テーブルの公開を行う。

これらファイルの公開は、本文書と同じwebページに掲載する「産業の研究開発に関するデータ整備について」に記したように、産業セクターの研究開発やイノベーションに関する分析・研究に用いるデータベースの構築に利用し、その分析・研究成果を通じて科学技術イノベーション政策の形成に貢献することを目的とする。

なお、以下の【公開するファイル】枠内に記したファイルの利用に際して、「NISTEP企業名辞書」はクリエイティブ・コモンズ・ライセンス（CC ライセンス）の「表示－継承」を、「外部データ接続テーブル」は「表示－非営利」を適用する。

CCライセンスの詳細は、http://creativecommons.jp/licenses/ を参照願いたい。

また、表示するクレジットは下記の通りである。

|  |
| --- |
| BY-SA（表示－継承）　又は、BY-NC（表示－非営利）原作者名:文部科学省科学技術・学術政策研究所 作品タイトル：産業の研究開発に関するデータURL：http://www.nistep.go.jp/research/scisip/data-and-information-infrastructure |
| 【公開するファイル】Ⅰ.NISTEP企業名辞書・RDB型NISTEP企業名辞書－Ver.2016.1 タブ区切りtxt形式の構成テーブル一式（comp\_name\_dic\_rdb\_ver2016\_01.zip）・単一テーブル型NISTEP企業名辞書－Ver.2016.1 excel形式（comp\_name\_dic\_ver2016\_01.xlsb）Ⅱ.外部データ接続テーブル［NISTEP企業名辞書と以下の外部データとの接続テーブル］・IIPパテントデータベース[[1]](#footnote-1)との接続用－Ver.2016.1 txt（タブ区切り）形式（ct\_comp\_name\_dic\_vs\_iip\_ver2016\_01.txt）・NISTEP大学・公的機関名辞書との接続用－Ver.2016.1 txt（タブ区切り）形式（ct\_comp\_name\_dic\_vs\_organ\_name\_dic\_ver2016\_01.txt）・日本の会社データ4万社[[2]](#footnote-2)との接続用－Ver.2013.1 csv（カンマ区切り）形式（ct\_dic\_nkdop\_ver2013\_01.csv）【補足】・()内はファイル名＜続く＞＜続き＞・txtファイルのエンコード形式はUTF-8　BOMなし　である。・各ファイルはzip形式で圧縮して公開しているので、ダウンロード後、解凍願いたい・excel形式のNISTEP企業名辞書は、RDB化された辞書情報を従来と同様の単一テーブルで見通しよく利用したい利用者の利便性を考え提供するものである |

# １．公開するファイルの概要

公開するファイルは、産業セクターのイノベーション分析・研究に用いるデータベースの構築に必要な「NISTEP企業名辞書」を核として、「外部データ接続テーブル」を加えた図1の太い点線で囲った範囲のファイルである。

外部データ接続テーブルとして公開するのは、「IIPパテントデータベース」と「日本の会社データ4万社」を接続対象とする二つの接続テーブルである。

データベースの構築は、MySQL等のデータベース管理システムを用い、txt形式で提供する各ファイルおよび外部データをインポートしてテーブルを構成し、所定のフィールド間を関連付けてRDB化する。

|  |
| --- |
|  |

図表 1　産業セクターのイノベーション分析・研究に用いるデータベース連携の全体図

## １．１ NISTEP企業名辞書

NISTEP企業名辞書（以下、「企業名辞書」と呼ぶ）は、産業セクターのイノベーション分析・研究に用いるデータベースの中心に位置付けられ、特許情報や国内営利企業（以下、「企業」と呼ぶ）に関する各種調査情報など、外部データから指定した企業に関する情報を抽出するためのハブとしての役割を担う。また、企業名の読み、本社所在地、業種など、外部データに含まれる数多くの企業から分析対象である企業を正しく特定しデータ抽出するための支援、および、合併や企業名称の変遷を考慮したデータの収集など、企業を中心としたイノベーション分析・研究における核となる機能を持つ。図表 1で、企業名辞書は一冊の辞書のイラストで単純化し表しているが、前版のVer.2015.1以降、17のテーブルで構成するリレーショナルデータベース形式としている。

### （１）企業名辞書掲載企業

企業名辞書に掲載する企業は、原則、以下の4条件を満足する企業の論理和で構成する。

①特許出願数累積100件以上

②株式上場企業

③特許出願数の伸び率大

④NISTEP大学・公的機関名辞書掲載企業

**①特許出願数累積100件以上**

IIPパテントデータベースの2015年版において、1970年以降の企業の変遷（名称変更、合併）を考慮した特許出願の集約を行い、累積出願数が100件を超える企業を特定したものである。100件を閾値としているのは、企業による特許出願状況の分析から、100件以上の出願実績を持つ企業による出願総数が国内営利企業による全出願数の90%以上を占めていること、また、閾値を下げても、指数関数的に対象企業数が増加する一方で占有率の上昇は僅かであることから決定している。

**②株式上場企業**

株式上場企業（以下、「上場企業」と呼ぶ）は、2012年1月時点の全上場企業を基準として、以降2016年3月までに新規（又は再）上場した企業を追加掲載している。大手や中堅の製造業各社は条件①の基準を満たす確率が高い一方、例えば金融・保険業種などは有力企業であっても条件①からは抽出されず、イノベーション分析適用を考慮した業種ごとの抽出網羅性を高めるためにこの条件を設けている。なお、上場廃止した企業は上場廃止情報を所定のテーブル/フィールドに追加し、辞書から当該企業の抹消は行っていない。

**③特許出願数の伸び率大**

近年起業し活躍するベンチャー企業など、条件①、②では取りこぼす可能性がある企業の抽出を目的とする。ここでは、1970年以降の企業の変遷を考慮した年ごとの出願数を把握し、それらデータを用いて3年、5年、7年の各期間で1年ごと移動させた線形フィットを行い、大きな回帰係数（出願数増分）を持つ企業を抽出する。抽出企業には条件①および②から抽出した企業が含まれるが、それら企業を除いた中から上位500社強を抽出して掲載を行っている。

**④NISTEP大学・公的機関名辞書掲載企業**

科学論文データを中心に個別機関レベルでの体系的な分析を行うための基盤となる大学・公的機関名辞書（http://www.nistep.go.jp/research/scisip/randd-on-university）に含まれる企業である。それらの企業は論文著者所属企業から抽出されたものである。

その他僅かに上記4条件に該当しない企業が存在するが、それらは企業名辞書作成初期における特許出願企業の名寄せ過程で紛れ込んだ同名異企業であり、その後の精査で分離を行ったが、抹消せず残したものである。

以上の条件により企業名辞書Ver.2016.1に掲載した企業数は図表 2に示す通りである。

図表 2　掲載条件と企業数



### （２）企業名辞書における変遷名称の取り扱い

企業は、名称の変更、合併等による企業形態の変化、清算・倒産などが日常的に起こっており、データ分析において、設立から現在までの一気通貫した取り扱いに困難が伴うことが多い。また、同名異企業の存在により、異企業のデータを取り違える恐れや、逆に、本来同一企業のデータと判別されるべきデータが異なる企業のものとして扱われてしまう恐れもある。

このような障害を排除するために、企業名辞書と連携する外部データは、現状の企業名称のみならず、沿革に基づいた旧名称や所在地等を考慮した名寄せを行い、可能な限り正確に情報連携させるよう配慮している。

この前提として、企業名辞書には企業の沿革に沿い遡った旧企業名称も掲載し、外部データの旧企業名称時代の情報は企業名辞書の旧企業名称に紐付けて時代ごとの取り扱いをできるようにしている。逆に、旧企業名称時代から現在までの情報を一括して取り扱いしたい場合は、旧企業名称時代から現在までに付与された共通のid（沿革id）を用いて取り扱いできるようにするなど、分析要求にフレキシブルに対応できるよう配慮した。

企業名辞書Ver.2016.1に掲載した企業数は8,388社であり、これらの変遷名称も含む全企業名称の数は18,650件である。

また、持株会社の取り扱いとして、企業が○○ホールディングス等に名称変更し持株会社化され、同時に旧名称を引き継いだ事業子会社を新設分割する場合（抜け殻方式）が数多くあるが、企業名辞書では実際に合わせて新設事業会社は別法人として企業番号を付与し掲載している。

### （３）企業名辞書に掲載した企業情報

企業名辞書は、企業の主業に関する産業分類（日本標準産業分類（Japan Standard Industrial Classification）、証券コード協議会業種）、企業規模、沿革、連結関係、所在地、URLなど分析者を支援する各種の企業情報を含んだテーブルで構成している。

但し、財務データや株価データなどは、EDINET等を通じた有価証券報告書からのデータ収集が比較的容易であることや、市販データも存在することから、それらデータを収集・公開することは止め、データを接続するキーとして証券コードやEDINETコード情報を付加している。

## **１．２** 外部データ接続テーブル

外部データ接続テーブル（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、データ間の架け橋となる機能を有する。

また、接続テーブルには、外部データに含まれる企業名称の表記揺れを吸収する役割も持たせている。例えば、IIPパテントデータベースの出願人テーブルに記載された旧い年代の出願人表記には、カタカナやOCR読み取りに起因すると推測される多くの誤記や異称などが存在する。データ間の同一企業情報を接続する作業過程において、表記は揺れているが同一企業と判定できる企業は可能な限り接続テーブルに接続情報を含めている。データ間の同一企業の判別には、原則として企業名称、法人格コード（前株後株等をコード化）、住所コード（市区町村レベル）の三つの要素を使用するが、ここでは目視作業により機械的なマッチングでは不可能な接続を補完している。これにより、企業名辞書の正式企業名称から、外部データの表記揺れした企業のデータ取得を可能としている。

# ２．公開するファイルの内容

ここでは、公開するファイルのテーブル構成、及び各テーブルのフィールドについて説明する。

## ２．１ 企業名辞書

### （１）テーブルの構成

企業名辞書は、単一テーブル型の企業名辞書（comp\_name\_dic\_ver2016\_01.xlsb）と、データの共通性や管理に配慮し正規化し複数のテーブル構成としたリレーショナルデータベース型の企業辞書（comp\_name\_dic\_rdb\_ver2016\_01.zip）の2種類を提供する。

#### ①リレーショナルデータベース型企業辞書

企業名辞書は、前版のVer.2015.1より図表 3に示す17のテーブルで構成するリレーショナルデータベース構造で提供することとした。（Entity Relationship Diagramは図表 24参照)

このリレーショナルデータベース化はデータの正規化によるメンテナンス性の向上や企業データをパネル化できるようにする要望が大きかったことによる。

各テーブルの概要は図表 3に一覧としてまとめておく。なお、図表中の論理名は利用者に意味が通じる名称であるのに対して、物理名はコンピューターが識別する名称を意味している。

図表 3　企業名辞書の構成テーブル

| 番号 | テーブル名称 | 概要 |
| --- | --- | --- |
| 論理名 | 物理名 |
| 1 | 企業名辞書メインテーブル | 1\_comp\_name\_main\_TBL | 企業名、企業id等のメインの情報、およびパネルデータとして整備をする必要がなく、最新の情報のみ保持すればよいデータを保管 |
| 2 | 沿革テーブル | 2\_comp\_history\_TBL | 名称変更や吸収合併などの事象が発生した際に発生した年、事象の種類を保管 |
| 3 | 所在地テーブル | 3\_address\_TBL | 企業の所在地に関する情報を保管本社、本店、移転など複数の住所情報の保管、パネル化が可能 |
| 4 | 企業規模テーブル | 4\_comp\_size\_TBL | 資本金、従業員数、中小企業基本法による企業規模情報を保管規模測定年ごとのパネル化が可能 |
| 5 | 業種（証券コード協会）テーブル | 5\_ind\_class\_tse\_TBL | 証券コード協議会の定める当該企業の業種区分を保管属する分類が変更された際のパネル化が可能 |
| 6 | 業種（日本標準産業分類）テーブル | 6\_ind\_class\_jsic\_TBL | 主業の日本標準産業分類を保管属する分類が変更された際のパネル化が可能 |
| 7 | EDINETコードテーブル | 7\_edinet\_code\_TBL | EDINETのコードを保管コードが変更された際のパネル化が可能 |
| 8 | 証券コードテーブル | 8\_sec\_code\_TBL | 証券コードを保管コードが変更された際のパネル化が可能 |
| 9 | 連結企業テーブル | 9\_consolidate\_TBL | 連結子会社である場合の親企業情報を保管連結関係の変化のパネル化が可能 |
| 10 | データ登録条件マスターテーブル | 10\_reg\_reason\_MTBL | 企業が企業名辞書に登録された理由に関するマスターテーブル |
| 11 | 企業名称使用開始事象マスターテーブル | 21\_use\_name\_start\_event\_MTBL | 新設、旧名称からの名称変更等、企業名称の使用が開始された場合の使用開始事象に関するマスターテーブル |
| 12 | 企業名称使用終了事象マスターテーブル | 22\_use\_name\_end\_event\_MTBL | 名称変更、吸収合併など、企業名称の使用が終了した場合の使用終了事象に関するマスターテーブル |
| 13 | 事業所区分マスターテーブル | 31\_office\_class\_MTBL | 住所情報の本社、本店、事業所等を判定するためのマスターテーブル |
| 14 | 業種（証券コード協会）マスターテーブル | 51\_tse\_MTBL | 証券コード協議会の定める業種区分に関するマスターテーブル |
| 15 | 業種（日本標準産業分類）マスターテーブル | 61\_jsic\_MTBL | 日本標準産業分類に関するマスターテーブル平成25年10月改定・平成26年4月1日施行に準拠 |
| 16 | 企業連結事象発生マスターテーブル | 91\_consolidate1\_MTBL | 連結事象が発生した場合の発生理由(子会社化等)に関するマスターテーブル |
| 17 | 企業連結事象終了マスターテーブル | 92\_consolidate2\_MTBL | 連結事象が終了した場合の発生理由(他社の子会社となった、独立した等) |

#### ②単一テーブル型企業名辞書

企業名辞書をリレーショナルデータベース化した副作用もあり、企業ごとのデータ参照に一手間かかり辞書として如何なものかという意見が存在することや、特にSQLのスキルを持たない利用者にとっては扱い辛いことも確かである。このため、一部情報の制約はでるが従来と同様にエクセル形式の単一テーブル型の企業名辞書を提供することにした。

### （２）テーブルのフィールド定義

図表 3の企業名辞書の構成テーブルに関して、フィールド（項目）を図表 4～図表 20に定義する。

なお、表中のY/Nは以下を意味する。

重複 Y=重複を許す、N=重複を許さない

NULL Y=NULLを許す、N=NULLを許さない

主キー データを一意に決定する事が出来る項目にY

外部キー 外部データベースとの接続に利用するキーの場合、当該テーブルの対応

図表 4　企業名辞書メインテーブルのフィールド定義

|  |
| --- |
| 企業名辞書メインテーブル　[1\_comp\_name\_main\_TBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 説明 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | N | N | Y |  | 企業（企業名称ごと）に固有に付与した番号 |
| 沿革番号 | history\_id | 数値（整数） | Y | N |  |  | 同一企業の変遷レコードをグループ化して扱うための番号 |
| 企業名称 | comp\_name | 文字列 | Y | Y |  |  | 企業の名称（変遷名称も含む） |
| ふりがな | read | 文字列 | Y | Y |  |  | 上記企業名称のふりがな |
| 法人格コード | comp\_code | 文字列 | Y | Y |  |  | 企業の法人格を表すコード（下表参照） |
| 英語名称 | e\_name | 文字列 | Y | Y |  |  | 企業の英語名称 |
| URL | url | 文字列 | Y | Y |  |  | 企業のウェブページのURL |
| データ登録理由番号 | reg\_reason\_id | 数値（整数） | Y | Y |  | データ登録理由マスターテーブルの理由番号 | 当該企業の辞書掲載条件 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

|  |  |
| --- | --- |
| 法人格コード | 例 |
| KB1 | 株式会社○○○ |
| KB2 | ○○○株式会社 |
| KB3 | ○○○株式会社□□□ |
| YG1 | 有限会社○○○ |
| YG2 | ○○○有限会社 |
| GD1 | 合同会社○○○ |
| GD2 | ○○○合同会社 |
| SG1 | 相互会社○○○ |
| SG2 | ○○○相互会社 |
| GS1 | 合資会社○○○ |
| GS2 | ○○○合資会社 |
| GM1 | 合名会社○○○ |
| GM2 | ○○○合名会社 |

図表 5　沿革テーブルのフィールド定義

|  |
| --- |
| 沿革テーブル　[2\_comp\_history\_TBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 説明 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | N | N | Y | 企業名辞書メインテーブルの企業番号 | 企業（企業名称ごと）に固有に付与した番号 |
| 企業名称使用開始年 | uns\_year | YEAR | Y | Y |  |  | 企業名称の使用を開始した年 |
| 企業名称使用開始事象id | uns\_event\_id | 数値（整数） | Y | Y |  | 企業名称使用開始事象マスターテーブルの事象番号 | 企業名称の使用を開始した事由 |
| 事象発生前企業id | pre\_comp\_id | 数値（整数） | N | Y |  | 企業名辞書メインテーブルの企業番号 | 事象発生前に名乗っていた企業名称につけられた企業番号 |
| 企業名称使用終了年 | une\_year | YEAR | Y | Y |  |  | 企業名称の使用が終了した年 |
| 企業名称使用終了事象id | une\_event\_id | 数値（整数） | Y | Y |  | 企業名称使用終了事象マスターテーブルの事象番号 | 企業名称の使用を終了した事由 |
| 事象発生後企業id | post\_comp\_id | 数値（整数） | Y | Y |  | 企業名辞書メインテーブルの企業番号 | 事象終了後の新しい名称につけられた企業番号 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

図表 6　所在地テーブルのフィールド定義

| 所在地テーブル　[3\_address\_TBL] |
| --- |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | Y | N | Y | 企業名辞書メインテーブルの企業番号 | 企業（企業名称ごと）に固有に付与した番号 |
| 所在地利用開始年 | ads\_year | YEAR | Y | Y |  |  | 所在地で活動を開始した年 |
| 所在地利用終了年 | ade\_year | YEAR | Y | Y |  |  | 所在地で活動を終了した年 |
| 事業所区分コード | office\_code | 数値（整数） | Y | N | Y | 事業所区分マスターテーブルの区分コード | 本店、本社、事業所等の区分コード |
| 所在地 | address | 文字列 | Y | N | Y |  | 所在地住所 |
| 都道府県コード | pref\_code | 数値（2桁整数） | Y | Y |  |  | 所在地の都道府県を示すコード |
| 地方自治体コード | city\_code | 数値（5桁整数） | Y | Y |  |  | 所在地の市区町村レベルまでの住所コード |
| 住所コード | add\_code | 数値（12桁整数） | Y | Y |  |  | 所在地の大字・町丁目を示す国道交通省が定める住所コード |
| 緯度 | latitude | 数値（実数） | Y | Y |  |  | 所在地の北緯 |
| 経度 | longitude | 数値（実数） | Y | Y |  |  | 所在地の東経 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

図表 7　企業規模テーブルのフィールド定義

|  |
| --- |
| 企業規模テーブル　[4\_comp\_size\_TBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | Y | N | Y | 企業名辞書メインテーブルの企業番号 | 企業（企業名称ごと）に固有に付与した番号 |
| 企業規模測定年 | judg\_year | YEAR | Y | N | Y |  | 企業規模を確認した年 |
| 中小企業基本法 | comp\_size\_law | 文字列 | Y | Y |  |  | 中小企業基本法に準拠し判定した企業規模 |
| 資本金階級 | comp\_size\_cap | 文字列 | Y | Y |  |  | 資本金の該当階級 |
| 従業員数階級 | comp\_size\_emp | 文字列 | Y | Y |  |  | 従業員数の該当階級 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

企業規模（中小企業基本法）の区分

大企業者

中小企業者

小規模企業者

資本金階級

100万円未満

100万円以上

1000万円以上

2000万円以上

5000万円以上

1億円以上

10億円以上

従業員階級

5人未満

5～29人

30～99人

100～299人

300～999人

1,000～4,999人

5,000人以上

図表 8　業種（証券コード協会）テーブルのフィールド定義

| 業種（証券コード協会）テーブル　[5\_ind\_class\_tse\_TBL] |
| --- |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | Y | N | Y | 企業名辞書メインテーブルの企業番号 | 企業（企業名称ごと）に固有に付与した番号 |
| 業種分類開始年 | inds\_year | YEAR | Y | Y |  |  | 証券コード協会の業種分類の確認初年 |
| 業種分類終了年 | inde\_year | YEAR | Y | Y |  |  | 証券コード協会の業種分類の確認最終年 |
| 業種分類コード | ind\_code | 数値（4桁整数） | Y | N | Y | 業種（証券コード協会）マスターテーブルの分類コード | 証券コード協会の分類該当業種 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

図表 9　業種（日本標準産業分類）テーブルのフィールド定義

|  |
| --- |
| 業種（日本標準産業分類）テーブル　[6\_ind\_class\_jsic\_TBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | Y | N | Y | 企業名辞書メインテーブルの企業番号 | 企業に固有のid |
| JSIC開始年 | jsics\_year | YEAR | Y | Y |  |  | JSIC分類を確認した最初の年 |
| JSIC終了年 | jsice\_year | YEAR | Y | Y |  |  | JSIC分類を確認した最後の年 |
| JSIC分類番号 | jsic\_code | 文字列 | Y | N | Y | 業種（日本標準産業分類）マスターテーブルのJSIC分類番号 | JSIC分類該当業種 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

図表 10　 EDINETコードテーブルのフィールド定義

|  |
| --- |
| EDINETコードテーブル　[7\_edinet\_code\_TBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | Y | Y | Y | 企業名辞書メインテーブルの企業番号 | 企業（企業名称ごと）に固有に付与した番号 |
| EDINETコード確認年 | check\_year | YEAR | Y | Y | Y |  | EDINETコードを確認した年 |
| EDINETコード | edinet\_code | 文字列 | Y | Y |  |  | 電子開示システムEDINETにおけるコード |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

図表 11　証券コードテーブルのフィールド定義

|  |
| --- |
| 証券コードテーブル　[8\_sec\_code\_TBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | Y | Y | Y | 企業名辞書メインテーブルの企業番号 | 企業（企業名称ごと）に固有に付与した番号 |
| 証券コード | sec\_code | 数値（4桁整数） | Y | Y | Y |  | 証券コード協議会が付与する証券コード |
| 上場市場 | stock\_market | 文字列 | Y | Y |  |  | 企業が上場した東証等の市場 |
| 上場日 | listed\_date | 年月 | Y | Y |  |  | 企業が上場した日 |
| 上場廃止日 | delisted\_date | 年月 | Y | Y |  |  | 企業が上場を廃止した日 |
| ISINコード | isin\_code | 文字列（12桁） | Y | Y |  |  | 国際的に統一された証券コード |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

図表 12　連結企業テーブルのフィールド定義

|  |
| --- |
| 連結企業テーブル　[9\_consolidate\_TBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | Y | Y | Y | 企業名辞書メインテーブルの企業id | 企業（企業名称ごと）に固有に付与した番号 |
| 連結事象発生年 | cons\_year | YEAR | Y | Y |  |  | 連結事象が発生した年 |
| 連結事象発生事象id | cons\_id | 数値（整数） | Y | Y |  | 連結事象発生マスターテーブルの事象id | 連結事象発生前の当該企業の連結状況 |
| 連結事象発生前連結企業id | post\_parent\_compid | 数値（整数） | Y | Y |  | 企業名辞書メインテーブルの企業id | 連結事象発生前の親会社の企業id |
| 連結先企業id | parent\_compid | 数値（整数） | Y | N | Y | 企業名辞書メインテーブルの企業id | 連結の親会社の企業id |
| 連結事象終了年 | cone\_year | YEAR | Y | Y |  |  | 連結事象が終了した年 |
| 連結事象終了事象id | cone\_id | 数値（整数） | Y | Y |  | 連結事象終了マスターテーブルの事象id | 連結事象終了後の当該企業の連結状況 |
| 連結事象終了後連結企業id | new\_parent\_compid | 数値（整数） | Y | Y |  | 企業名辞書メインテーブルの企業id | 連結事象終了後の親会社の企業id |
| データ登録日 | reg\_year | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_year | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

図表 13　データ登録条件マスターテーブルのフィールド定義

|  |
| --- |
| データ登録理由マスターテーブル　[10\_reg\_reason\_MTBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 条件番号 | re\_id | 数値（整数） | N | N | Y | 企業名辞書メインテーブルの企業id | 企業が企業名辞書に登録された抽出条件を示すid |
| 登録条件 | reason | 文字列 | N | Y |  |  | データが企業名辞書に抽出登録された条件 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

図表 14　企業名称使用開始事象マスターテーブルのフィールド定義

|  |
| --- |
| 企業名称使用開始事象マスターテーブル　[21\_use\_name\_start\_event\_MTBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 事象番号 | uns\_event\_id | 数値（整数） | N | N | Y |  | 企業名称使用開始事象のid（下表参照） |
| 事象概要 | uns\_event | 文字列 | N | Y |  |  | 企業名称使用開始事象の概要（下表参照） |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

|  |  |
| --- | --- |
| 事象id | 事象概要 |
| 1 | 設立 |
| 2 | 名称変更 |

図表 15　企業名称使用終了事象マスターテーブルのフィールド定義

|  |
| --- |
| 企業名称使用終了事象マスターテーブル　[22\_use\_name\_end\_event\_MTBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 事象番号 | une\_event\_id | 数値（整数） | N | N | Y |  | 企業名称使用終了事象のid（下表参照） |
| 事象概要 | une\_event | 文字列 | N | Y |  |  | 企業名称使用終了事象の概要（下表参照） |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

|  |  |
| --- | --- |
| 事象id | 事象概要 |
| 1 | 現存 |
| 2 | 名称変更 |
| 3 | 合併 |
| 4 | 破産 |
| 5 | 清算 |

図表 16　事業所区分マスターテーブルのフィールド定義

|  |
| --- |
| 事業所区分マスターテーブル　[31\_office\_class\_MTBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 事業所区分コード | office\_code | 数値（整数） | N | N | Y |  | 本店、本社、事業所等の区分コード（下表参照） |
| 事業所区分 | class | 文字列 | N | N |  |  | 本店、本社、その他の事業所等の区分（下表参照） |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

|  |  |
| --- | --- |
| 事業所区分コード | 概要 |
| 1 | 本社 |
| 2 | 本店 |
| 3 | 研究所 |
| 4 | 事業所 |
| 5 | その他 |

図表 17　業種（証券コード協会）マスターテーブルのフィールド定義

|  |
| --- |
| 業種（証券コード協会）マスターテーブル　[51\_tse\_MTBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 業種コード | class\_code | 数値（4桁整数） | Y | N | Y |  | 証券コード協会の業種分類によるコード |
| 版 | tse\_ver | 年月日 | Y | N | Y |  | 証券コード協会の業種分類の版 |
| 大分類 | l\_class | 文字列 | Y | N |  |  | 証券コード協会の業種大分類 |
| 小分類 | s\_class | 文字列 | Y | N |  |  | 証券コード協会の業種小分類 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

図表 18　業種（日本標準産業分類）マスターテーブルのフィールド定義

|  |
| --- |
| 業種（日本標準産業分類）マスターテーブル　[61\_jsic\_MTBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| JSIC分類番号 | jsic\_code | 文字列 | Y | N | Y |  | 日本標準産業分類の分類番号－JISC○○○○ |
| JSIC版 | jsic\_ver | 日付 | Y | N | Y |  | 日本標準作業分類の版（改訂日） |
| JSIC大分類 | jsic\_l | 文字列 | Y | N |  |  | 日本標準産業分類大分類 |
| JSIC中分類 | jsic\_m | 文字列 | Y | N |  |  | 日本標準産業分類中分類 |
| JSIC小分類 | jsic\_s | 文字列 | Y | N |  |  | 日本標準産業分類小分類 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

図表 19　企業連結事象発生マスターテーブルのフィールド定義

|  |
| --- |
| 企業連結事象発生マスターテーブル　[91\_consolidate1\_MTBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 詳細 |
| 論理名 | 物理名 |
| 事象番号 | cons\_id | 数値（整数） | N | N | Y |  | 連結事象発生事象の番号 |
| 事象概要 | cons\_event | 文字列 | N | Y |  |  | 連結事象発生事象の概要 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

|  |  |
| --- | --- |
| 事象id | 事象概要 |
| 1 | 連結子会社として設立 |
| 2 | 名称変更以前から連結子会社であった |
| 3 | 独立企業から連結子会社になった |
| 4 | 他社の連結子会社から当社の連結子会社になった |

図表 20　企業連結事象終了マスターテーブルのフィールド定義

|  |
| --- |
| 企業連結事象終了マスターテーブル　[92\_consolidate2\_MTBL] |
| フィールド名 | データ型 | 重複 | NULL | 主キー | 外部キー | 説明 |
| 論理名 | 物理名 |
| 事象番号 | cone\_id | 数値（整数） | N | N | Y |  | 連結事象終了事象の番号 |
| 事象概要 | cone\_event | 文字列 | N | Y |  |  | 連結事象終了事象の概要 |
| データ登録日 | reg\_date | 年月日 | Y | N |  |  | データを本テーブルに登録した日 |
| データ更新日 | up\_date | 年月日 | Y | N |  |  | 既登録データの情報更新した日 |

|  |  |
| --- | --- |
| 事象id | 事象概要 |
| 1 | 現在も連結子会社である |
| 2 | 名称変更 |
| 3 | 連結子会社から独立企業になった |
| 4 | 当社の連結子会社から他社の連結子会社になった |
| 5 | 連結先に合併されて解散した |
| 6 | 連結先以外に合併されて解散した |
| 7 | 合併以外の理由により解散した |

## ２．２ 外部データ接続テーブル

### （１）IIPパテントデータベースとの接続用

企業名辞書と外部データであるIIPパテントデータベースを連携させるための接続テーブルであり、企業名辞書メインテーブルの企業idとIIPパテントデータベースの出願人テーブルのida\_seqフィールドを関係付け接続する。

図表 21　接続テーブル（IIPパテントデータベースとの接続用）のフィールド定義

|  |  |  |
| --- | --- | --- |
| フィールド名 | データ型 | 説明 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | 企業（企業名称ごと）に固有に付与した番号 |
| IIPパテント出願番号＋記載順序 | ida\_seq | 文字列 | 上記企業番号の企業が出願人である特許 |

### （２）NISTEP大学・公的機関名辞書との接続用

企業名辞書とNISTEP大学・公的機関名辞書、又は企業名辞書と科学論文-NISTEP大学・公的機関名辞書対応テーブルと連携させるための接続テーブルであり、企業名辞書メインテーブルの企業idとNISTEP大学・公的機関名辞書の機関番号を関係付け接続する。

図表 22　接続テーブル（日本の会社データ4万社データとの接続用）のフィールド定義

|  |  |  |
| --- | --- | --- |
| フィールド名 | データ型 | 説明 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | 企業（企業名称ごと）に固有に付与した番号 |
| 機関id | n\_id | 文字列 | NISTEP大学・公的機関名辞書に掲載する機関の識別のために付与した番号 |
| 下部組織フラグ | low\_org | 数値（整数） | NISTEP大学・公的機関名辞書に掲載する企業には、部署、社内カンパニー、企業立病院など企業の下部組織が一部含まれている。下部組織フラグはそれらを識別するためのフラグである。 |

|  |  |
| --- | --- |
| フラグ番号 | 下部組織フラグ説明 |
| 1 | 部署 |
| 2 | 社内カンパニー |
| 3 | 企業立病院 |

### （３）日本の会社データ4万社データとの接続用

企業名辞書メインテーブルと外部データである東洋経済新報社の「日本の会社データ4万社」を連携させるための接続テーブルであり、企業名辞書メインテーブルの企業番号と東洋経済企業コードを対応させ関係付けしている。

なお、日本の会社データ4万社は、2011年4月版を対象に接続テーブルを作成している。その後、新版が発行されているが、東洋経済企業コードに変更がない限り、2011年版で対応付けされた企業の範囲で接続可能である。

また、この接続テーブルにおける東洋経済企業コードの使用と公開は株式会社東洋経済新報社より許諾を受けている。本稿執筆時点では、許諾権の制約から新版対応の接続テーブルを作成する計画はない。

図表 23　接続テーブル（日本の会社データ4万社データとの接続用）のフィールド定義

|  |  |  |
| --- | --- | --- |
| フィールド名 | データ型 | 説明 |
| 論理名 | 物理名 |
| 企業番号 | comp\_id | 数値（整数） | 企業（企業名称ごと）に固有に付与した番号 |
| 東洋経済企業コード | code\_tk | 文字列 | 東洋経済新報社が企業の識別のために付与した独自の番号 |

# ３．外部データベース

産業セクターのイノベーション分析・研究に用いるデータベースのうち、外部データに関しては「NISTEP大学・公的機関名辞書」を除きNISTEPの公開対象外である。

以下に、参考用として、外部データベースのうちIIPパテントデータベース、日本の会社データ4万社について簡単な説明を加えておく。

## ３．１ IIPパテントデータベース

一般財団法人知的財産研究所（略称IIP）が提供する特許データベースであり、特許庁の特許整理標準化データをリレーショナル型のデータベースに変換し、イノベーション研究に必要と思われる項目を集め公開している。

データベースは、出願テーブル、出願人テーブル、発明者テーブル等に構造化されている。企業名辞書との接続は、このうち、出願人テーブルと行う。

IIPのURL：http://www.iip.or.jp/

## ３．２ 日本の会社データ4万社

東洋経済新報社が独自調査で収集した上場会社および非上場会社の合計4万社弱の企業情報を収録した商業データベースである。

1社1レコード形式で、csvファイルで提供されている。

# ４．公開ファイルのダウンロードとデータベースの構築

## ４．１ ダウンロード

公開したファイルは、任意のディレクトリにダウンロードし、圧縮形式のファイルは解凍する。

## ４．２ データベースの構築

企業名辞書を構成する各テーブルファイルは、MySQL等のデータベース管理システムを用いてインポートし、図表 24に示すテーブル間のリレーションを定義する。

データのインポート方法や構築方法は利用するデータベース管理システムの説明書を参照されたい。

IIPパテントデータベースと接続する場合は、図表 25に示すように接続テーブルを介して接続を行う。また、市販の財務データベースやその他企業情報と接続する場合は、証券コードやEDINETコードを利用して接続することができる。

|  |
| --- |
|  |

図表 24　企業名辞書ＥＲＤ



図表 25　企業名辞書とＩＩＰパテントデータベース／NISTEP大学・公的機関名辞書との接続

# ５．企業名辞書の作成と名寄せ

## ５．１ 企業名辞書の作成

（１）企業名辞書は、ＩＩＰパテントデータベースを用いて、特許出願実績の多い企業、及びある時期に特許出願数が増加した企業の特定を行い、加えて、そこに含まれない上場企業を追加する手順で作成している。

（２）特許出願数に関する企業の特定において、企業名称の変遷、合併、及びパテントデータベースの出願人表記揺れ等を考慮し出願人名寄せを行い算出した特許出願数を用いている。

（３）企業名辞書では、企業名称の変更や合併前の企業名称など沿革に基づく一連の変遷名称をグループ化して取り扱いできるように沿革コードを付与している。また、企業名辞書改訂時点における現存企業は、企業ホームページからＵＲＬ、本社所在地等の最新情報を取得し収録している。なお、企業名称の表記は、有価証券報告書、又は企業がインターネット等で用いられている表記としており、登記上の表記とは異なる場合もある。

（４）企業名称の読みに関し、「日本」の読みは、下記ＵＲＬ掲載の「日本（にほん・にっぽん）の読み方」に掲載の情報を参考とした。「にっぽん」として掲載のない企業名称は、原則、「にほん」としている。

　　　　　http://hiramatu-hifuka.com/onyak/nippon.html

（５）緯度・経度

所在地の緯度・経度の算出には、東京大学空間情報科学研究センターが下記ＵＲＬにて利用提供する「CSVアドレスマッチングサービス」を利用した。

http://newspat.csis.u-tokyo.ac.jp/geocode/

（６）企業の産業分類（日本標準産業分類）は、@niftyビジネスの企業検索、Wikipediaの日本の企業一覧、当該企業のホームページなど各種のインターネット情報から取得した情報を参考として、総務省統計局の日本標準産業分類五十音索引表（分類検索を簡便化するための細分類各項目の具体的業種が例示された表）と引き当てした産業分類としている。

## ５．２ 企業名の変遷を考慮した名寄せ

（１）IIPパテントデータベースの特許出願人データは、整理標準化データ[[3]](#footnote-3)をそのまま採用しているが、同一出願人に対して異なる表記（表記揺れ）が存在し[[4]](#footnote-4)、同一出願人単位の集約を行う場合には表記揺れをクレンジングした上で「名寄せ」する作業が必要となる。
また、整理標準化データでは出願人固有のIDとして出願人番号が付与されているものの、年代を遡るほど出願人番号が付与されていない例も多数存在する。このため、出願人を適切に同定するために以下のような手順で名寄せ作業を実施している。

①企業名の修正と標準化

特許出願人名について、目視による修正を行った。目視による修正は、類似表記が多数ある場合は、同一年における出願数の多い表記を正として、その表記に統一した。次に、「株式会社」「有限会社」などの法人格を表す表記を標準コード化し、企業名の表記法として＜企業名＋法人格コード＞の形式を本整備の標準形式とした。

②所在地表記の修正と標準化

多様な表記のゆれが混在するIIPパテントデータベースの出願人所在地表記を対象に、漢数字、ローマ数字、番地町名形式等の表記の標準化を行った。次に市町村合併、町名変更などの情報を別の情報源から収集し、これを用いて、過去の所在地情報も最新のものに統一した。その上で、国土交通省が定義する12桁の住所コード（街区レベル位置情報参照情報）を各出願人住所（所在地）に割り当てた。以上の作業により、所在地情報の表記法として＜住所コード（12桁）＞形式を本整備の標準形式とした。

③企業名＋法人格コード＋住所コードの組み合わせによる企業の名寄せ

上記作業で修正した企業名、法人格コードおよび住所コードを利用し、3項目が一致するレコードを同一企業であるとみなし、名寄せを行った。

④名寄せ結果の評価と修正

名寄せ結果として、異企業を同一企業と見なすエラーは見当たらない。しかし、同一企業を異企業と見なしてしまうエラーが一部発見された。長期間のデータには、所在地表記に番地などの僅かな違いが存在し、12桁コードでは細密過ぎることが判明した。結論として、5桁の市区町村レベルのコード[[5]](#footnote-5)で精度高く名寄せできることが判明し、これによる名寄せを実施した。

また、長期間に渡るデータであるため、多くの企業がその間に所在地移転している。このため、同一企業でありながら寄せができない事例も多々生じたため、移転状況のチェックを行いつつ人手で名寄せを行った。

（２）企業の変遷名称を考慮した同一企業の名寄せは、企業名辞書Ver.2014.1までは大西宏一郎氏（大阪工業大学）、西村陽一郎氏（神奈川大学）らの研究グループにより整備された特許出願人の名寄せに関するデータベース（ONAPPdata)を参考とさせて頂いた。収録年の制約もあり、その後の版では、有価証券報告書、および企業ホームページの沿革等から得た情報をソースとしている。

# おわりに

企業名辞書の改訂公開も今回で5回目となり、特許と接続し企業セクターの各種分析を行うという視点では必要十分な企業を網羅したと考えている。

ただし、企業情報も特許情報も絶えず変化又は増加しており、今は良くても直ぐに陳腐化し、更新のために常に走り続けることが求められる。

この点で、今後は情報の品質を維持しつつ、改訂負担を最少化するよう心掛けている。

また、各種の企業情報データベースを接続した分析に国税庁から指定・通知される法人番号を利用することが考えられ、次版では法人番号の付加を予定している。

企業名辞書が各種分析や教育等に活用される事例も増えており、こうしたデータ整備事業の成果が科学技術イノベーション政策の形成を始めとする多くの事柄に貢献できれば幸いである。

以上

# 【補足】

単一テーブル型の企業名辞書にのみ存在するフィールドについて以下に説明を加える。

## （１）沿革（history）を表現するコード

沿革（history）に記されたコードは、企業名称変更時の形態（合併のような形態が変化が伴う場合も含む）を簡易的に示したものである。

沿革は、図表 26に示すような合併による形態の変更も含めた企業名称の変遷をコードで表している。

図表 26　沿革コードの説明

|  |
| --- |
| 【補足説明】・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 |

1. IIPパテントデータベースは、一般財団法人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。
IIPパテントデータベースの入手は同研究所のホームページより行い、定められた利用規約に従って利用願いたい。（http://www.iip.or.jp/） [↑](#footnote-ref-1)
2. 日本の会社データ4万社は、市販データベースである。必要に応じて、発売社（東洋経済新報社）に確認願いたい。 [↑](#footnote-ref-2)
3. 特許情報を、XMLやSGMLといった標準的なデータの記述方法を用いて記述することで整理したデータである。 [↑](#footnote-ref-3)
4. 以下の文献に詳述されている。

Onishi, K., Y. Nishimura, N. Tsukada, I. Yamauchi, T. Shimbo, M. Kani and K. Nakamura (2012) "Standardization and Accuracy of Japanese Patent Applicant Names", IIPR Working Paper No.2012-001. [↑](#footnote-ref-4)
5. JIS市区町村コードと同じコード [↑](#footnote-ref-5)