

NISTEP 企業名辞書
利用マニュアル
(ver. 2020_1 対応版)

文部科学省科学技術・学術政策研究所
(2020年1月)

目次

はじめに	1
1. 公開するファイルの概要	3
1. 1 NISTEP 企業名辞書	3
(1) 企業名辞書に掲載する条件	4
(2) 企業の名称変遷の取り扱い	6
(3) 企業に関する情報	6
1. 2 外部データ接続テーブル	6
2. 公開するファイルの内容	7
2. 1 企業名辞書	7
(1) テーブルの構成	7
①リレーショナルデータベース型企业辞書	7
②単一テーブル型企业名辞書	8
(2) テーブルのフィールド定義	8
2. 2 外部データ接続テーブル	18
(1) IIP パテントデータベースとの接続テーブル	18
(2) NISTEP 大学・公的機関名辞書との接続テーブル	18
(3) 米国特許との接続テーブル	19
(4) NISTEP 意匠データベースとの接続テーブル	19
(5) NISTEP 商標データベースとの接続テーブル	19
(6) 日本の会社データ 4 万社データとの接続用	20
3. 外部データベース	20
3. 1 IIP パテントデータベース	20
3. 2 NISTEP 大学・公的機関名辞書	20
3. 3 米国特許の出願人 (Applicant 又は Assignee) に記された国内営利企業名英語表 記揺れテーブル	20
3. 4 NISTEP 意匠データベース・NISTEP 商標データベース	21
3. 5 日本の会社データ 4 万社	21
4. 公開ファイルのダウンロードとデータベースの構築	21
4. 1 ダウンロード	21
4. 2 データベースの構築	21
5. 企業名辞書の作成と名寄せ	22
5. 1 企業名辞書の作成	22
5. 2 企業名の変遷を考慮した名寄せ	23
【Appendix】	25
(1) 沿革 (history) を表現するコード	25
(2) 米国特許との接続	25

(2) - 1	接続対象米国特許	25
(2) - 2	日米特許の紐づけ	25
a)	国内特許から米国へパリ優先権を主張して米国へ出願した出願人情報の利用	25
b)	国際出願 (PCT 出願) から米国移行した特許情報の利用	25
c)	日米特許の紐づけ	26
(2) - 3	企業名辞書掲載企業と米国特許出願人 (applicant 又は assignee) の紐づけ	26
(2) - 4	接続テーブルの作成	26

【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2015年6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 2015年6月現在の状況を反映し見直した。

- ・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。

(6) Ver. 2018. 1（2018 年 1 月）：以下の追加・改訂を行った。

- ・NISTEP 企業名辞書；
 - ・企業名辞書登録企業について、外部データとの接続性向上を目的として法人番号の情報を追加した。
 - ・以下の企業状況を 2016 年 6 月現在の状況を基本に確認し、変更のある場合情報の追加、改訂を行った。
 - ・企業名称が変更された場合
 - ・統合・再編等が行われた場合
 - ・株式上場（新規・再）、又は廃止した場合
 - ・倒産、清算等により会社解散が行われた場合
 - ・企業の連結関係に変化がある場合
 - ・その他、所在地、企業 H P、規模、業種等の企業情報に変更がある場合
 - ・IIP パテントデータベース（2015 年版）から、沿革を考慮した企業名寄せを再評価し、未掲載であった特許出願件数（累積 100 件超）の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 7 月から 2016 年 6 月の期間に上場した企業を追加登録した。
 - ・NISTEP 大学・機関名辞書（ver. 2016. 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. 2016. 1）に未掲載の企業を追加登録した。
- ・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（2015 年版）と NISTEP 企業名辞書（Ver. 2018. 1）との接続情報を追加・改訂した。
 - ・NISTEP 大学・公的機関名辞書 ver. 2016. 1 との接続テーブルを更新した。
- ・利用マニュアル；上記に関連する説明の改訂を行った。

(7) Ver. 2018. 11（2018 年 4 月）：以下の改訂追加公開を行った。

- ・NISTEP 企業名辞書（Ver. 2018. 1）と IIP パテントデータベース（2015 年版）との接続テーブルについて、誤記、記載漏れ等のマイナー改訂を行った。
- ・米国特許との接続用テーブル及び米国特許の出願人に記された国内営利企業名英語表記ゆれテーブルの新規公開を行った。

(8) Ver. 2019_1（2019 年 1 月）：以下の追加・改訂を行った。

- ・NISTEP 企業名辞書；
 - ・以下の企業状況を 2017 年 6 月現在の状況を基本に確認し、変更のある場合情報の追加、改訂を行った。
 - ・企業名称が変更された場合

- ・ 統合・再編等が行われた場合
 - ・ 株式上場（新規・再）、又は廃止した場合
 - ・ 倒産、清算等により会社解散が行われた場合
 - ・ 企業の連結関係に変化がある場合
 - ・ その他、所在地、企業HP、規模、業種等の企業情報に変更がある場合
- ・ IIP パテントデータベース（2017年版）から、沿革を考慮した企業名寄せを再評価し、未掲載であった特許出願件数（累積100件超）の企業を追加登録した。
 - ・ IIP パテントデータベース（2017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。
 - ・ 2016年7月から2017年6月の期間に上場した企業を追加登録した。
 - ・ NISTEP 大学・機関名辞書（ver.2018.1）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.2018.11）に未掲載の企業を追加登録した。
 - ・ 意匠又は商標の登録件数の多い企業を追加登録した。
 - ・ 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（2017年版）との接続を前提に接続情報を追加・改訂した。
 - ・ NISTEP 大学・公的機関名辞書 ver.2018.1 との接続テーブルを更新した。
 - ・ 利用マニュアル；上記に関連する説明の改訂を行った。
- (9) Ver. 2019_2（2019年11月）：以下の追加・改訂を行った。
- ・ NISTEP 企業名辞書と外部データベースとの接続テーブル；
 - ・ NISTEP で公開する意匠データベース及び商標データベースの出願企業との接続テーブルを新規公開した。
 - ・ 利用マニュアル；上記に関連する説明の改訂を行った。
- (10) Ver. 2020_1（2020年1月）：以下の追加・改訂を行った。
- ・ NISTEP 企業名辞書；
 - ・ 大学発ベンチャー企業の追加を行った。
 - ・ 掲載済企業の状態を2018年6月現在の状態を基本に確認し、変更のある場合情報の追加、改訂を行った。確認項目は（8）記載の場合と同様。
 - ・ 2017年7月から2018年6月の期間に上場した企業を追加登録した。
 - ・ IIP パテントデータベース（2017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。
 - ・ 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（2017年版）との接続テーブルを NISTEP 企業名辞書 Ver.2020_1 との接続を前提としたテーブルに改訂した。
 - ・ NISTEP 大学・公的機関名辞書 ver.2018.1 との接続テーブルを更新した。
 - ・ 利用マニュアル；上記に関連する説明を追加した。

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。



（表示－継承）

又は、



（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. 2020_1
タブ区切り txt 形式の構成テーブル一式【comp_name_dic_rdb_ver2020_1.zip】 **New**
- ・ 単一テーブル型 NISTEP 企業名辞書－Ver. 2020_1
excel 形式【comp_name_dic_ver2020_1.xlsx】 **New**

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との接続テーブル－Ver. 2020_1
txt（タブ区切り）形式【ct_comp_name_dic_vs_iip_ver2020_1.zip】 **New**
- ・ NISTEP 大学・公的機関名辞書との接続テーブル－Ver. 2020_1
txt（タブ区切り）形式【ct_comp_name_dic_vs_organ_name_dic_ver2020_1.txt】 **New**
- ・ 日本の会社データ 4 万社²との接続テーブル－Ver. 2013. 1
csv（カンマ区切り）形式【ct_dic_nkdop_ver2013_1.csv】
- ・ 米国特許との接続テーブル－Ver. 2018. 1
txt（タブ区切り）形式【ct_comp_name_dic_vs_USPT_ver2018_1.zip】

<続く>

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/>）

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。

< 続き >

- ・ NISTEP 意匠データベースとの接続テーブル－Ver. 2019_2
txt (タブ区切り) 形式【ct_comp_name_dic_vs_design_ver2019_2.zip】
- ・ NISTEP 商標データベースとの接続テーブル－Ver. 2019_2
txt (タブ区切り) 形式【ct_comp_name_dic_vs_trademark_ver2019_2.zip】

Ⅲ. 表記揺れテーブル

- ・ 米国特許の出願人に記された国内営利企業名英語表記揺れテーブル
txt (タブ区切り) 形式【Notation Variations table of Corporate applicants_ver2018_1.zip】

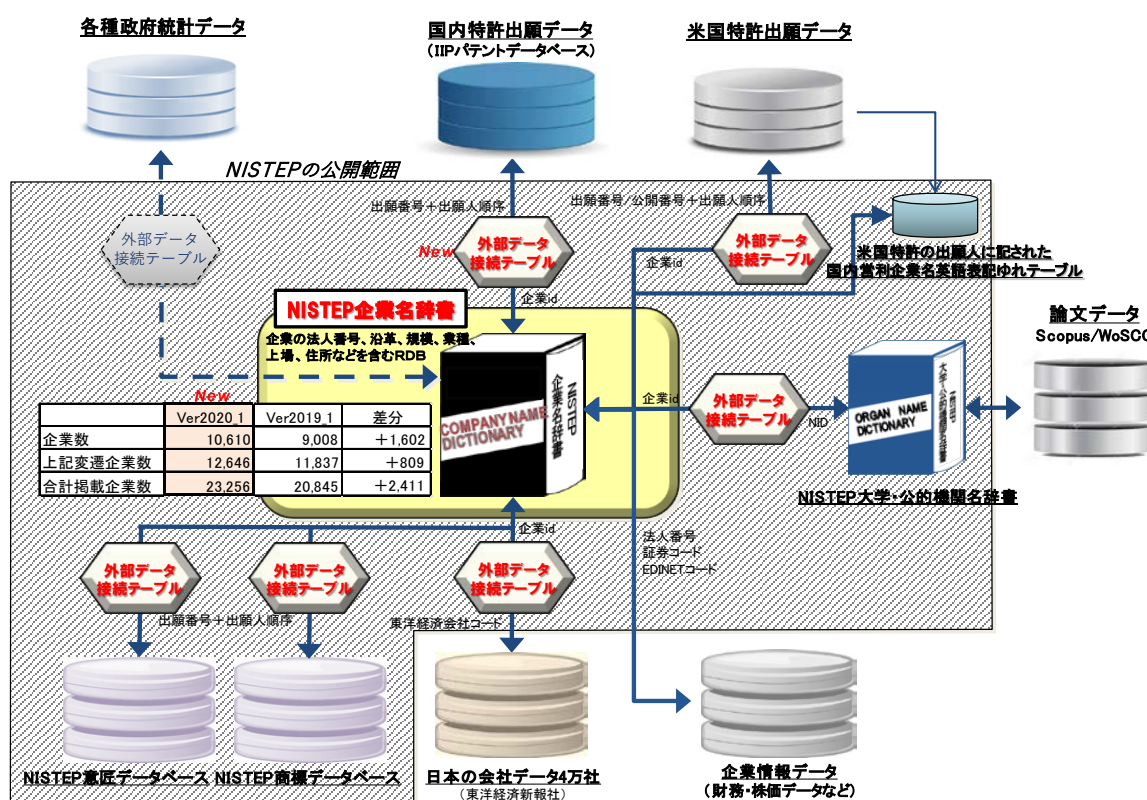
【補足】

- ・ 【】内はファイル名
- ・ **New** の付けられたファイルは新規に作成 (又は改訂) され公開するファイルである。
- ・ txt ファイルのエンコード形式は UTF-8 BOM なし である。
- ・ ファイルの圧縮は zip 形式で行っている
- ・ Excel 形式の NISTEP 企業名辞書は、RDB 化された辞書情報を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「外部データ接続テーブル」を加えた図表 1 に示す範囲である。

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、国内営利企業（以下、「企業」と呼ぶ）に関し分析するための基本情報の提供や特許情報を始めとする各種の外部データベースを接続するための橋渡しなどハブとしての役割を担う。論文データとは NISTEP 大学・公的機関名辞書を経由して接続が可能であり、企業単位で特許情報と論文情報とを接続して分析することができる。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、正しく企業情報に関する辞書として活用できる。

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、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（以下、「RDB」と略す）形式に改め公開している。

なお、全ての利用者がRDBの操作に慣れている訳ではなく、機能が削られても良いから一つのテーブルで企業情報を見通せるように欲しいというニーズもあることからエクセル形式のファイルも公開している。

(1) 企業名辞書に掲載する条件

企業名辞書に掲載する企業は、原則、次の5つの条件の何れか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
- ④ NISTEP 大学・公的機関名辞書掲載企業
- ⑤ 意匠・商標登録数
- ⑥ 大学発ベンチャー企業

① 特許出願数累積 100 件以上

IIP パテントデータベースの 2017 年版により、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 件を超えた企業を掲載している。ちなみに、100 件を閾値としているのは、別途実施した企業による特許出願数の分析結果から、100 件以上の出願実績を持つ企業の出願数合計が全企業の出願数の 90%以上を占めていること、また、この閾値を下げると、指数関数的に該当企業が増える一方で占有率は僅かしか上昇しない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 2017 年 10 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、その後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わず、他の掲載企業と同様に企業情報の更新を行っている。

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。線形フィットの傾きの閾値は、それぞれ 8、21/5、75/28 としている。この閾値は、当初、傾きの上位社から掲載した経緯があり、その後、最下位社の傾きを閾値として採用したことによる。

④ NISTEP 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。

⑤ 意匠・商標登録数累積 100 件以上

NISTEP 意匠・商標データベース³にて、意匠・商標登録数累積 100 件以上の企業である。

⑥ 大学発ベンチャー企業

企業名辞書－Ver. 2020_1 で新たに追加した掲載条件である。それ以前の版においても、①～⑤の条件に該当する一部の大学発ベンチャー企業は掲載されていたが、大学発ベンチャー企業として新たに掲載属性を設け、該当企業の調査を行い追加している。

大学発ベンチャー企業を抽出するデータソースとして、企業名辞書－Ver. 2020_1 では経済産業省の大学発ベンチャーデータベース、大学の HP で公開する大学認定ベンチャー企業、大学発ベンチャーに関する各種調査報告書に掲載された企業などを用いている。抽出企業の中には既に企業活動を中止している企業や、公開された情報から企業属性を取得するのが困難な企業も多く、企業名辞書－Ver. 2020_1 では同版公開までの期間において抽出努力した結果を掲載している。

その他、上記条件に該当しない企業として、次の事由による掲載企業がある。

⑦ 持株会社制移行に伴い設立された事業会社

前記条件に合致し掲載された企業が純粋持株会社に移行し、同時に事業を承継するために新規に設立された会社である。ある事業会社が〇〇ホールディングス等に名称変更して純粋持株会社に移行し、同時に旧名称を引き継いだ事業子会社が新設分割される場合（抜け殻方式）が数多くある。世間的には、それら新旧の同一名称の会社は区別せず同じ企業として扱われている。このため、企業名辞書では新設事業子会社を追加掲載し、法人登記に合わせて純粋持株会社となった同一名称の旧会社と別法人として企業番号を付与している。

⑧ 一部事業の譲渡に伴い設立された会社

前記条件に合致し掲載された企業がその事業の一部を承継させるために新たに設立した会社。

⑨ 名称変更又は吸収合併した企業が登録事由に該当

現在の企業は登録事由に該当しないが、同企業の沿革の中で登録事由に該当する企業を吸収合併した事象がある。企業の登録事由確認時点で、登録事由に該当する企業を吸収合併した当時の名称から変更されている企業も含む。

⑩ その他

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、その後の精査で分離したものなどが少数存在する。

³ NISTEP で公開する NISTEP 意匠データベース及び NISTEP 商標データベースを利用した。このデータベースには 2000 年から 2013 年までに登録された意匠・商標に関するデータが収録されている。
<http://www.nistep.go.jp/archives/27214>

(2) 企業の名称変遷の取り扱い

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貫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て使う恐れや、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られるようにし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名称まで共通の id (沿革 id) を用いて取り扱えるようにするなど、分析要求にフレキシブルに対応できるよう配慮した。

(3) 企業に関する情報

企業名辞書は、企業の主業に関する産業分類として、日本標準産業分類 (Japan Standard Industrial Classification) 及び証券コード協議会業種のほか、企業規模、沿革、連結関係、所在地、URL など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、それらデータの収集・公開は行っていない。市販データ等と接続する場合は、法人番号、証券コード及び EDINET コード情報を利用されたい。

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_ver2019_1.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_ver2019_1.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、2015年にリリースした Ver. 2015.1以降は、図表2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。(Entity Relationship Diagramは図表30/P22参照)

これ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2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2 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_M_TBL	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_M_TBL	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

番号	テーブル名称		概要
	論理名	物理名	
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）
18	大学発ベンチャー企業テーブル	100_Venture_TBL	関連大学、研究者、設立区分など大学発ベンチャーに特化した情報を保管
19	大学発ベンチャー設立区分（文科省）マスターテーブル	101_Venture_type1_MTBL	大学発ベンチャー組織の設立区分に関するマスターテーブル（文科省区分）
20	大学発ベンチャー設立区分（経産省）マスターテーブル	102_Venture_type2_MTBL	大学発ベンチャー組織の設立区分に関するマスターテーブル（経産省区分）

②単一テーブル型企业名辞書

企業名辞書をリレーショナルデータベース化したことによる副作用もあり、企業ごとのデータ参照に手間が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、一部の情報に制約がでるが、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。

(2)テーブルのフィールド定義

図表 2 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 3～図表 19 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 3 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
法人番号	corp_num	数値 (整数)	Y	Y			国税庁により付与された法人番号

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						
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	株式会社〇〇〇
KB2	〇〇〇株式会社
KB3	〇〇〇株式会社□□□
YG1	有限会社〇〇〇
YG2	〇〇〇有限会社
GD1	合同会社〇〇〇
GD2	〇〇〇合同会社
SG1	相互会社〇〇〇
SG2	〇〇〇相互会社
GS1	合資会社〇〇〇
GS2	〇〇〇合資会社
GM1	合名会社〇〇〇
GM2	〇〇〇合名会社

図表 4 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_com p_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 5 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 6 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 7 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 8 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 9 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 10 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 11 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

連結企業テーブル [9_consolidate_TBL]							
フィールド名		データ	重複	NULL	主キ	外部キー	詳細
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 12 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_M_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						
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

条件番号	登録条件
1	特許出願件数
2	特許出願件数伸び
3	上場
4	大学公的機関名辞書
5	意匠または商標出願件数
6	持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
7	名称変更又は吸収合併した企業が登録事由に該当
8	大学発ベンチャー
12	特許出願件数・特許出願件数伸び
13	特許出願件数・上場
14	特許出願件数・大学公的機関名辞書
15	特許出願件数・意匠または商標出願件数
17	特許出願件数・特許出願件数伸び・名称変更又は吸収合併した企業が登録事由に該当
23	特許出願件数伸び・上場
24	特許出願件数伸び・大学公的機関名辞書
25	特許出願件数伸び・意匠または商標出願件数
26	特許出願件数伸び・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
27	特許出願件数伸び・名称変更又は吸収合併した企業が登録事由に該当
28	特許出願件数伸び・大学発ベンチャー
34	上場・大学公的機関名辞書
35	上場・意匠または商標出願件数

36	上場・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
37	上場・名称変更又は吸収合併した企業が登録事由に該当
38	上場・大学発ベンチャー
45	大学公的機関名辞書・意匠または商標出願件数
47	大学公的機関名辞書・名称変更又は吸収合併した企業が登録事由に該当
48	大学公的機関名辞書・大学発ベンチャー
56	意匠または商標出願件数・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
123	特許出願件数・特許出願件数伸び・上場
124	特許出願件数・特許出願件数伸び・大学公的機関名辞書
125	特許出願件数・特許出願件数伸び・意匠または商標出願件数
126	特許出願件数・特許出願件数伸び・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
128	特許出願件数・特許出願件数伸び・大学発ベンチャー
134	特許出願件数・上場・大学公的機関名辞書
135	特許出願件数・上場・意匠または商標出願件数
145	特許出願件数・大学公的機関名辞書・意匠または商標出願件数
156	特許出願件数・意匠または商標出願件数・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
234	特許出願件数伸び・上場・大学公的機関名辞書
235	特許出願件数伸び・上場・意匠または商標出願件数
238	特許出願件数伸び・上場・大学発ベンチャー
245	特許出願件数伸び・大学公的機関名辞書・意匠または商標出願件数
248	特許出願件数伸び・大学公的機関名辞書・大学発ベンチャー
345	上場・大学公的機関名辞書・意匠または商標出願件数
348	上場・大学公的機関名辞書・大学発ベンチャー
1234	特許出願件数・特許出願件数伸び・上場・大学公的機関名辞書
1235	特許出願件数・特許出願件数伸び・上場・意匠または商標出願件数
1245	特許出願件数・特許出願件数伸び・大学公的機関名辞書・意匠または商標出願件数
1246	特許出願件数・特許出願件数伸び・大学公的機関名辞書・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
1248	特許出願件数・特許出願件数伸び・大学公的機関名辞書・大学発ベンチャー
1256	特許出願件数・特許出願件数伸び・意匠または商標出願件数・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
1258	特許出願件数・特許出願件数伸び・意匠または商標出願件数・大学発ベンチャー
1345	特許出願件数・上場・大学公的機関名辞書・意匠または商標出願件数
2345	特許出願件数伸び・上場・大学公的機関名辞書・意匠または商標出願件数
12345	特許出願件数・特許出願件数伸び・上場・大学公的機関名辞書・意匠または商標出願件数
12348	特許出願件数・特許出願件数伸び・上場・大学公的機関名辞書・大学発ベンチャー
12356	特許出願件数・特許出願件数伸び・上場・意匠または商標出願件数・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
12456	特許出願件数・特許出願件数伸び・大学公的機関名辞書・意匠または商標出願件数・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社

図表 13 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_M_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						
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 14 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_M_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						
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 15 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_M_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						
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 16 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_M_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						
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 17 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〇〇〇〇
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 18 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 19 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

事象番号	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

図表 20 大学発ベンチャーテーブルのフィールド定義

大学発ベンチャーテーブル [100_Venture_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）	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
区分・文科省	type1	文字列（1～4の結合）	Y	Y			大学発ベンチャーの区分（文科省区分）。複数に該当する場合は列挙する（例えば、人材移転型かつ特許型の場合は12とする等）
区分・経産省	type2	文字列（1～5）	Y	Y			大学発ベンチャーの区分（経産省区分）。現時点ではいずれか一つが付与されているため、列挙はしないが、将来的に複数に該当するものが現れた場合は列挙する。
関連大学1	related_univ1	文字列	Y	Y			ベンチャー設立関連大学
関連大学2	related_univ2	文字列	Y	Y			ベンチャー設立関連大学
関連大学3	related_univ3	文字列	Y	Y			ベンチャー設立関連大学
関連研究者1	related_researcher1	文字列	Y	Y			ベンチャー設立関連大学研究者
関連研究者2	related_researcher2	文字列	Y	Y			ベンチャー設立関連大学研究者
関連研究者3	related_researcher3	文字列	Y	Y			ベンチャー設立関連大学研究者
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 21 大学発ベンチャー設立区分（文科省）マスターテーブルのフィールド定義

大学発ベンチャー設立区分（文科省）マスターテーブル [101_Venture_type1_M_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						
ベンチャー設立区分番号	type_id	数値（整数）	N	N	Y	ベンチャー区分に付与するid	ベンチャー区分
設立区分名称	name	文字列	N	N		ベンチャー区分の概要	区分名称

ベンチャー設立区分番号	設立区分名称
1	特許
2	技術移転
3	人材
4	出資
12	特許・技術移転
13	特許・人材
23	技術移転・人材

24	技術移転・出資
34	人材・出資
123	特許・技術移転・人材
124	特許・技術移転・出資
134	特許・人材・出資
234	技術移転・人材・出資
1234	特許・技術移転・人材・出資

図表 22 大学発ベンチャー設立区分（経産省）マスターテーブルのフィールド定義

大学発ベンチャー設立区分（経産省）マスターテーブル [102_Venture_type2_M_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						
ベンチャー設立区分番号	type_id	数値（整数）	N	N	Y	ベンチャー区分に付与する id	ベンチャー区分
設立区分名称	name	文字列	N	N		ベンチャー区分の概要	名称

ベンチャー設立区分番号	設立区分名称
1	研究成果ベンチャー
2	共同研究ベンチャー
3	技術移転ベンチャー
4	学生ベンチャー
5	関連ベンチャー
6	その他

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テーブル

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 23 IIP パテントデータベースとの接続テーブルのフィールド定義

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
出願日	adate	日付	特許の出願日

(2) NISTEP 大学・公的機関名辞書との接続テーブル

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 24 NISTEP 大学・公的機関名辞書との接続テーブルのフィールド定義

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) 米国特許との接続テーブル

企業名辞書メインテーブルと外部データである米国特許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米国特許出願番号又は米国特許公開番号にて接続を行う。

接続テーブル作成の詳細は、「【Appendix】(2) 米国特許との接続」を参照されたい。

図表 25 米国特許との接続テーブルのフィールド定義

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 特許出願番号 +記載順序	iip_ida	文字列	上記企業番号の企業が出願人である国内出願特許
米国特許(公開番号)	uspt_pub	文字列	上記企業番号の企業が出願人に含まれる米国出願特許
米国特許(出願番号)	uspt_app	文字列	上記企業番号の企業が出願人に含まれる米国出願特許

(4) NISTEP 意匠データベースとの接続テーブル

企業名辞書メインテーブルと外部データである NISTEP 意匠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番号と NISTEP 意匠データベースの意匠出願番号及び連番にて接続を行う。また、NISTEP 意匠データベースでは、出願以降の出願人・権利者の変更に伴って、最旧(登録直後)、最新、それぞれの時点のデータを保有しており、どちらか一方との接続は同データベースの新旧フラグを条件に加えて接続する。

図表 26 NISTEP 意匠データベースとのフィールド定義

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
意匠出願番号	application_number	文字列	上記企業番号の企業の意匠出願番号
連番	sequence	数値(整数)	上記企業番号の企業の出願人記載順序

(5) NISTEP 商標データベースとの接続テーブル

企業名辞書メインテーブルと外部データである NISTEP 商標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番号と NISTEP 商標データベースの商標出願番号及び連番にて接続を行う。また、NISTEP 商標データベースでは、出願以降の出願人・権利者の変更に伴って、最旧(登録直後)、最新、それぞれの時点のデータを保有しており、どちらか一方との接続は同データベースの新旧フラグを条件に加えて接続する。

図表 27 NISTEP 商標データベースとのフィールド定義

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
商標出願番号	application_number	文字列	上記企業番号の企業の商標出願番号

連番	sequence	数値(整数)	上記企業番号の企業の出願人記載順序
----	----------	--------	-------------------

(6) 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 28 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次の 3. 1 及び 3. 4 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 NISTEP 大学・公的機関名辞書

研究活動を行っている我が国の機関(約 1 万 9 千機関)を掲載した機関名辞書であり、論文の著者企業も含まれている。企業名辞書はそれらの企業と接続でき、その結果、特許と論文を企業レベルで接続することができる。

<http://www.nistep.go.jp/research/scisip/randd-on-university>

3. 3 米国特許の出願人 (Applicant 又は Assignee) に記された国内営利企業名英語表記揺れテーブル

企業名辞書と米国特許との接続情報の生成の中で整理した米国特許の出願人 (Applicant 又は Assignee) に記された国内営利企業名英語表記揺れテーブルである。

図表 29 米国特許の出願人に記された国内営利企業名英語表記揺れテーブル
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		
米国特許出願企業名	uspt_comp_name	文字列	米国特許出願に記載された企業名表記
記載出願件数順位	rank	数値(整数)	上記企業表記の出願件数順位
企業名称	comp_name	文字列	企業名辞書記載の企業名表記
法人格コード	comp_code	文字列	企業の法人格を表すコード
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
法人番号	corp_num	数値(整数)	国税庁により付与された法人番号
沿革番号	history_id	数値(整数)	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. history_id=1の企業が最新企業となる

3. 4 NISTEP 意匠データベース・NISTEP 商標データベース

意匠・商標各整理標準化データをデータソースとして、2000年から2013年までに登録された意匠・商標に関するデータを収録したデータベースである。NISTEPで構築し、下記URLで公開されている。

なお、NISTEP 意匠・商標データベースの利用はライセンス条件に従い、また、報告書等を引用する場合には出典の明記をお願いする。

https://nistep.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=4774&item_no=1&page_id=13&block_id=21

3. 5 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。

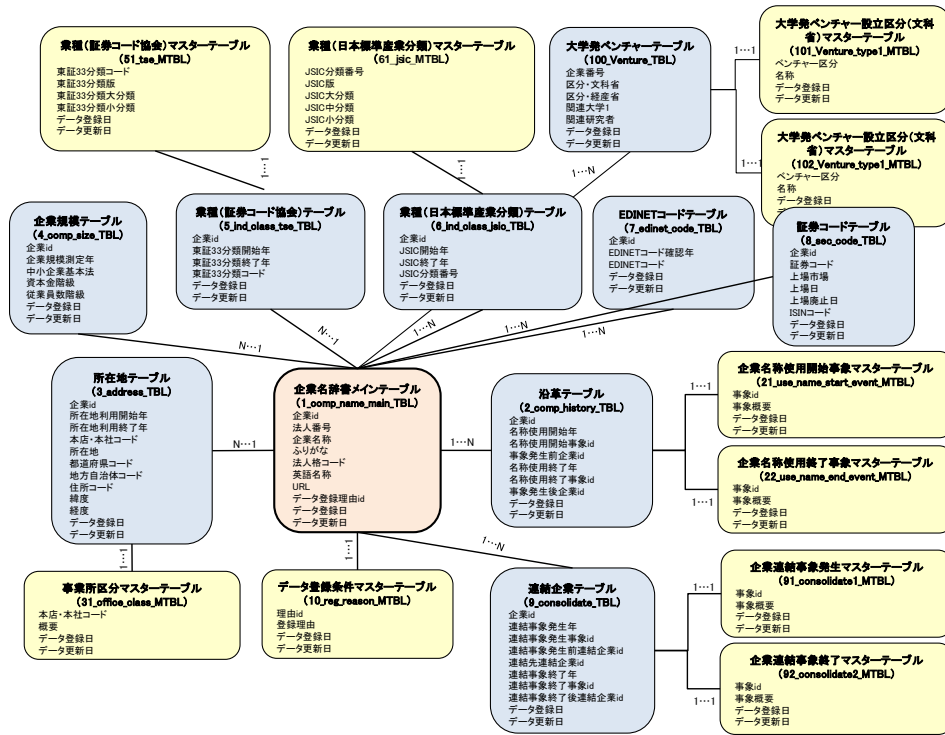
4. 2 データベースの構築

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 30 に示すテーブル間のリレーションを定義する。

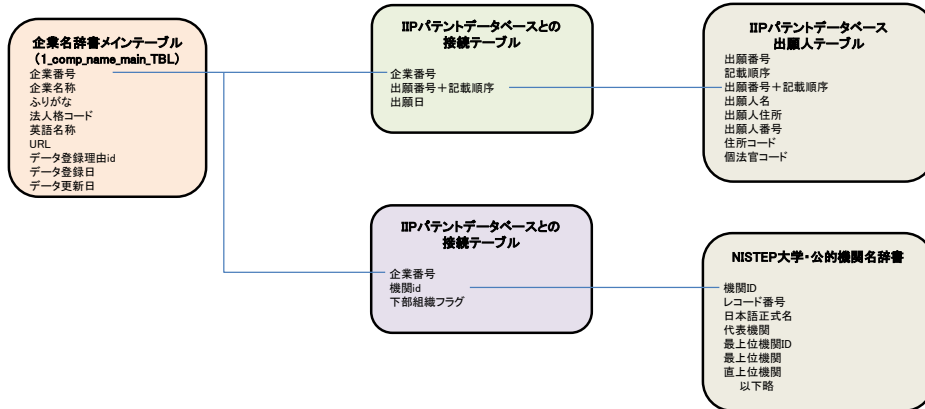
データのインポート方法や構築方法は利用するデータベース管理システムの説明書を参照されたい。

例えば、IIP パテントデータベースと接続する場合は、図表 31 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 30 企業名辞書 E R D



図表 31 企業名辞書と I I P 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、I I P 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え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⁴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⁵、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。

⁴ 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。

⁵ 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.

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、名寄せを行っている。加えて、近年の出願分では出願人番号を加えて名寄せを行っている。

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⁶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、企業名辞書作成初期の Ver. 2014.1 までは大西宏一郎氏（早稲田大学）、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。

以上

⁶ JIS 市区町村コードと同じコード

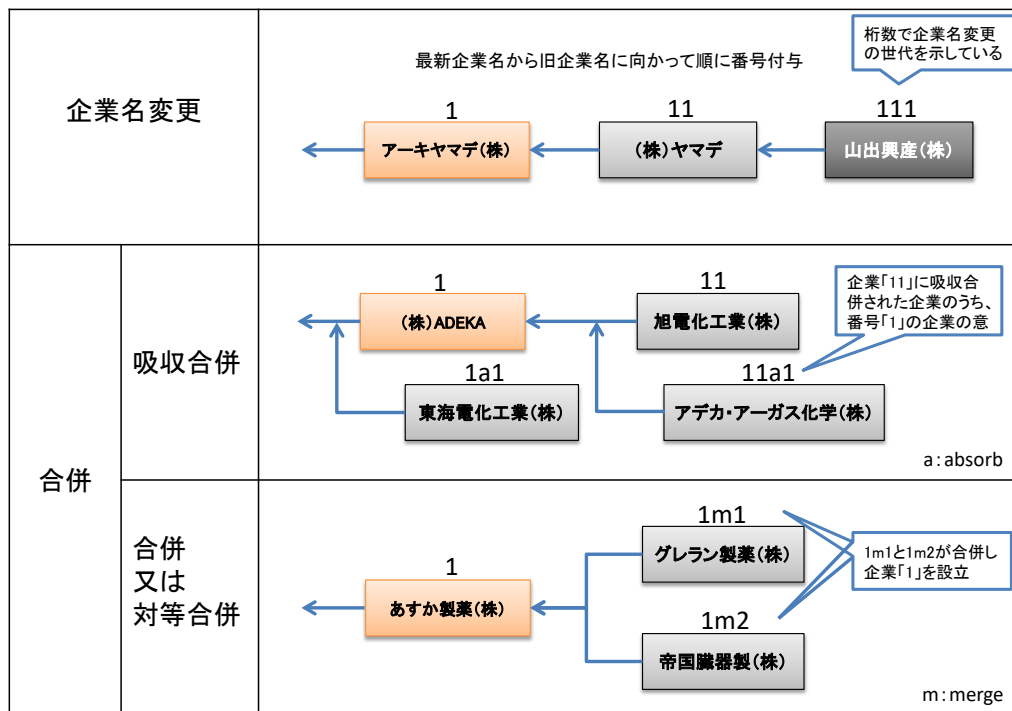
【Appendix】

(1) 沿革 (history) を表現するコード

単一テーブル型の企業名辞書には、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）を簡易的に示す沿革 (history) フィールドが付加されている。

沿革 (history) に記されたコードの意味は、図表 32 に示す通りである。

図表 32 沿革コードの説明



注: 上記に例示した企業は、企業名辞書の「あ」行掲載企業から例示に適した企業を選択した

【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

(2) 米国特許との接続

米国特許と企業名辞書に掲載する企業との接続は以下のように行っている。

(2) - 1 接続対象米国特許

米国特許の母集団は、2001年以降に出願された公開特許公報である。

(2) - 2 日米特許の紐づけ

a) 国内特許から米国へパリ優先権を主張して米国へ出願した出願人情報の利用

企業が特許出願を行う場合、自国への特許出願後に、パリ優先権を主張して、海外に権利取得のために特許出願を行うことが多い。米国公報に付与された日本からの優先権主張を抽出することで、国内から米国へ移行された特許を判別できる。優先権主張番号は、国内出願番号であるので、この二つを日米特許の紐づけキーとする。

b) 国際出願 (PCT 出願) から米国移行した特許情報の利用

近年、多数の外国へ特許出願することが多く、その際には国際出願が利用される。そ

ここで、企業が国際出願した後、米国へ移行した特許情報を JP の国際優先権情報が付与された米国公報を抽出することで、日米特許の紐づけキーとする。

c) 日米特許の紐づけ

2001 年以降の米国公開公報の XML データ 4,714,118 件から、米国出願番号、公開番号、パリ優先権主張番号、国際優先権主張番号、米国出願人名 (applicant 又は assignee) と、発明者名を取得する。

これらの情報と、(2) - 2 a) 及び b) で取得した優先権番号すなわち国内出願番号をキーとして紐づけする。

紐づけの結果、529,316 組の日米特許のペアが生成された。

(2) - 3 企業名辞書掲載企業と米国特許出願人 (applicant 又は assignee) の紐づけ

529,316 組の日米特許出願のペアから、日米ともに単願である 354,698 組及び日米ともに 2 出願人である 5,216 組について企業名辞書掲載企業について紐づけを行った。なお、2 出願人の共願のペアにおいては、出願人名の順番が入れ替わっているものがあり、入れ替え修正を行っている。

(2) - 4 接続テーブルの作成

(2) - 3 の結果を利用して、企業名辞書と米国特許との接続テーブルを作成した、なお、米国特許には米国出願番号及び米国公開番号を付与した。本来、米国出願番号のみで十分なはずであるが、米国出願番号のシリーズコードには不備があることが既知であり、公開番号の付与も行っている。