

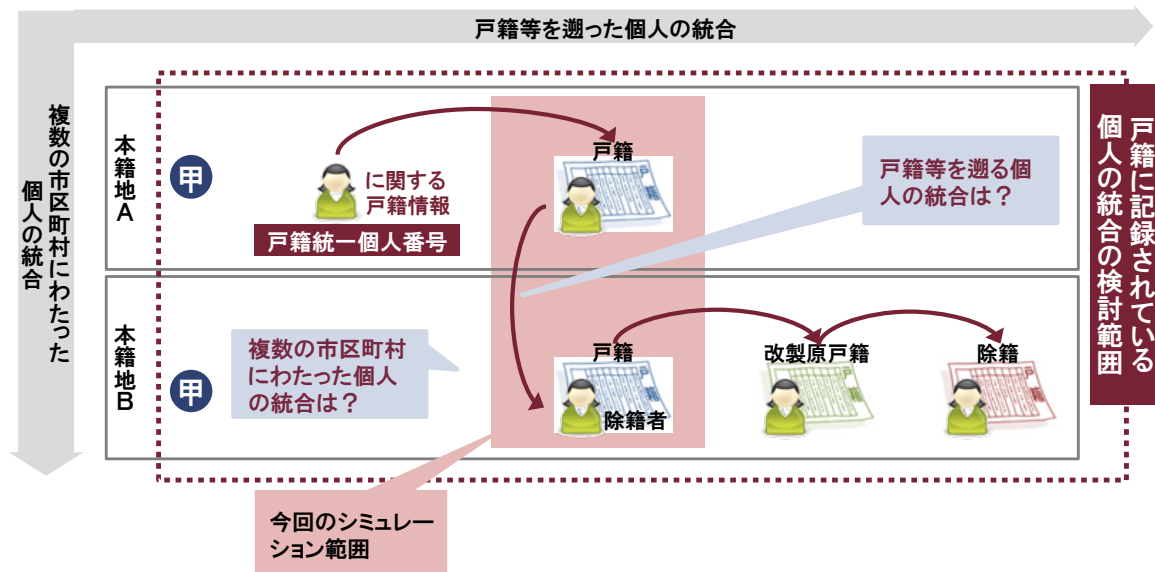
戸籍副本データに関する調査結果の概要について

1 調査概要

(1) 実施した調査

戸籍における一個人の情報は複数の戸籍、除籍、改製原戸籍に記録され、かつ、複数の市区町村にわたることがある。相続関係手続等においては現在戸籍の謄抄本等のみならず、除籍及び改製原戸籍等を多く必要とすることから、戸籍情報へのマイナンバー適用により、国民の利便性向上を実現するためには、その手順として、全国の戸籍を対象に個人を統合した上で、個人ごとに同一の戸籍統一個人番号（戸籍事務処理用のユニークな番号）を付与することが考えられる。

上記について、第5回システム検討WGにて、「戸籍情報へのマイナンバー適用範囲に関する検討における課題」として提示しており、この「戸籍等を遡る個人又は市区町村をわたった個人に対する統合（名寄せ）の実現可能性」について検討するために、戸籍副本データ管理センターの副本情報を元に名寄せのシミュレーションを実施した。なお、シミュレーション範囲としては下図に示すとおり、まずは電算化戸籍部分に絞って実施し、画像データで保存されている戸籍等については、対象外とした。



(2) 戸籍副本データ管理システムについて

戸籍副本データ管理システムでは、各市区町村が保有する戸籍情報の副本として、戸籍情報ファイル（テキスト形式）をLGWAN経由で第1又は第2戸籍副本データ管理センターへ日々送付し、各センターに登録している。第1戸籍副本データ管理センター（以下「第1C」という。）と第2戸籍副本データ管理センター（以下「第2C」という。）については、管轄範囲が2分されており、各センターの管轄範囲は以下となる。

ア 第1戸籍副本データ管理センター（東日本に設置）

名古屋法務局，大阪法務局，広島法務局，高松法務局及び福岡法務局の各管区域内の市区町村

イ 第2戸籍副本データ管理センター（西日本に設置）

東京法務局，仙台北法務局及び札幌法務局の各管区域内の市区町村

複数の市区町村にわたった個人の統合のシミュレーションの実施に当たっては，第1Cと第2Cに分かれている全市区町村の副本情報を1箇所を集約して作業を行うことが望ましいが，情報漏洩等のセキュリティリスクを考慮し，今回は各センター内の副本情報の範囲でのシミュレーションに留めた。

また，未電算化市区町村及び副本情報の送付を開始していない市区町村等，現在は副本情報が戸籍副本データ管理システムに未登録となっている市区町村も存在する。このような市区町村は，今回のシミュレーションの対象とならず，対象となった市区町村数は下表のとおりである。（政令市は行政区ごとに計上）。

表1 対象市区町村数

センター	対象市区町村数	全市区町村数
第1C	925 (98.6%)	938
第2C	912 (95.2%)	958
合計	1,837 (96.9%)	1,896

2 調査結果

本調査で行ったプログラムによる名寄せ作業の流れは下図のとおりであるが，以降，手順ごとに，シミュレーション結果を明示する。

手順1：名寄せに必要な情報の抽出

手順2：電算化戸籍の名寄せ

手順3：電算化戸籍の名寄せ結果の検証

(1) 手順1：名寄せに必要な情報の抽出

戸籍副本データ管理センターのデータベースから市区町村ごとに個人単位で情報を抽出し，「名寄せの基点となる情報」と「従前戸籍に記録される情報」に分類する。なお，名寄せの基点となる情報には，「在籍者」のほか，電算化以降に死亡した「除籍者」も含まれる。

ア 結果

抽出及び分類結果については、表2のとおりである。

なお、No. 1で従前の戸籍を持つとした個人（件数）については、No. 4の電算化後の従前の戸籍を持つ個人（件数）のほか、平成改製原戸籍等のイメージ化した従前の戸籍を持つ個人も含まれている。

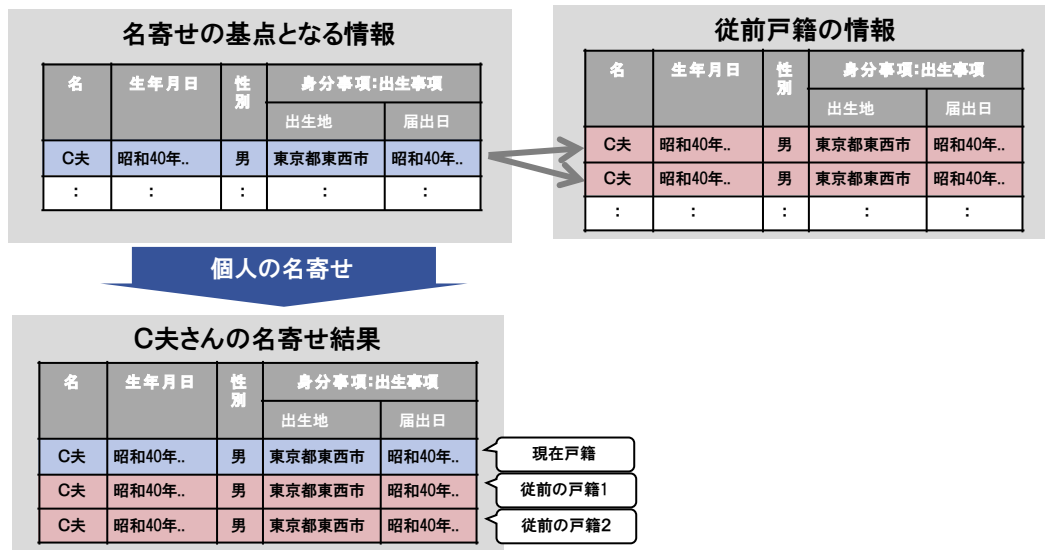
また、No. 2で従前戸籍を持たないとした個人（件数）については、電算化以降に出生し、かつ出生後に他の戸籍への異動が無い個人となる。従前の記録を持たないため、今回の統合シミュレーションの対象外とした。

表2 対象件数

No	データ区分		第1C		第2C		合計	
			件数	割合	件数	割合	件数	割合
1	名寄せの基点となる情報	従前の戸籍を持つ	62,485,420	94.0%	64,176,670	93.9%	126,662,090	93.9%
2		従前戸籍を持たない	3,987,172	6.0%	4,187,967	6.1%	8,175,139	6.1%
3		小計	66,472,592	-	68,364,637	-	134,837,229	-
4	従前の戸籍に記録される情報	-	14,193,437	-	14,372,592	-	28,566,029	-

(2) 手順2：電算化戸籍の名寄せ

電算化戸籍を対象に、「名寄せの基点となる情報」が持つ「生年月日」「性別」「(出生届の)届出日」「名」「出生地」を名寄せ条件として「従前戸籍の情報」を参照し、個人の名寄せを行った。



ア 結果

従前の戸籍に記録される個人に対する名寄せ結果は表3のとおりであり、従前の戸籍に記録される約80%の個人に対し、機械的な名寄せが成功したことが分かる。名寄せされなかった原因等については、後述する今後の課題において述べることとする。

なお、今回の調査において、市区町村が独自に登録している外字については、「ワイルドカード(全ての対象文字にマッチする文字)」として取り扱っているが、その使用率は表4のとおりであり、機械的な名寄せが成功した件数に対する外字の影響度としては全体の3%~7%程度であった。

表3 従前の戸籍に記録される情報に対する名寄せ率

No	内訳	第1C		第2C		合計	
		件数	割合	件数	割合	件数	割合
1	名寄せされるべき件数 (表2のNo. 4)	14,193,437	-	14,372,592	-	28,566,029	-
2	名寄せされた件数	11,365,295	80.1%	11,961,065	83.2%	23,326,360	81.7%
3	名寄せされなかった件数	2,828,142	19.9%	2,411,527	16.8%	5,239,669	18.3%

表4 最終的に一致した個人に対する名寄せ時の外字使用率

No	名寄せ項目	外字有無		第1C件数		第2C件数	
		比較元	比較先	件数	割合	件数	割合
1	名	無	無	11,017,784	96.9%	11,541,423	96.5%
2		有	無	79,157	0.7%	102,654	0.9%
3		無	有	83,593	0.7%	116,113	1.0%
4		有	有	184,761	1.6%	200,875	1.7%
5		合計		11,365,295	-	11,961,065	-
6	出生地	無	無	10,520,069	92.6%	11,371,452	95.1%
7		有	無	96,258	0.8%	136,767	1.1%
8		無	有	114,400	1.0%	111,032	0.9%
9		有	有	634,568	5.6%	341,814	2.9%
10		合計		11,365,295	-	11,961,065	-

イ 結果の検証

名寄せされた結果に対し、名寄せ元である現在戸籍と名寄せ先である従前の戸籍について、「本籍」「筆頭者」「父の名及び母の名」で突合を行い、名寄せ結果に対する検証（正しく名寄せが行えたかどうかの検証）を行った。その結果は表5のとおりである。

これによると、検証が成功した割合については99%に近い値が出ており、正しく名寄せが行われた（誤った個人を名寄せしていない）ことを示している。

表5 検証結果

No	内訳	第1C		第2C		合計	
		件数	割合	件数	割合	件数	割合
1	名寄せされた件数	11,365,295	-	11,961,065	-	23,326,360	-
2	検証が成功した件数	11,257,410	99.1%	11,818,613	98.8%	23,076,023	98.9%
3	検証が失敗した件数	107,885	0.9%	142,452	1.2%	250,337	1.1%

3 今後の課題

(1) 名寄せされなかった原因の究明

本来、電算化情報である従前の戸籍に記録される個人については、100%名寄せされることが望ましいが、機械的な名寄せ成功率は80%程度に留まっている。その原因等については、下記の点の影響が想定されるところ、今後、追加調査等を検討することとした。

ア 第1Cと第2Cにまたがって記録されている個人による影響

前述したとおり、今回は各センター内の副本情報の範囲でのシミュレーションに留めており、名寄せの基点となる個人情報と従前の戸籍に記録される個人情報が第1Cと第2Cにまたがって記録されている個人については、名寄せができない。

イ 副本情報の未登録市区町村による影響

戸籍副本データ管理システムに副本情報の登録が完了していない市区町村（未電算化市区町村を含む。）に名寄せの基点が記録されている個人については、名寄せができない。

ウ 例外データ等による影響

名寄せにおいては、文字列比較を行うところ、名寄せ元又は名寄せ先のデータにおいて誤った形式（例外形式）でデータ設定がされている場合、当然のことながら名寄せに失敗するため、例外データを排除して名寄せを行わなければならない。

また、市区町村をまたがった名寄せを行う場合は、名寄せ元と名寄せ先で戸籍情報システムの提供元（開発事業者）が異なるケースも多く存在する。戸籍副本データ管理システムにおいては、標準統一様式を定めてはいるが、各戸籍情報システム依存の項目等があった場合は、それを排除した上で名寄せを行わなければならない。

本調査において発見された例外等のデータ例については、表6ないし8のとおり。

表6 実在日以外の生年月日

No	内訳	第1C		第2C		合計	
		件数	割合	件数	割合	件数	割合
1	実在日	87,086,598	99.99%	86,936,857	99.99%	174,023,455	99.99%
2	実在日以外	10,944	0.01%	7,593	0.01%	18,537	0.01%
3	合計	87,097,542	-	86,944,450	-	174,041,992	-

表7 性別項目と父母との続柄項目の不一致件数

No	内訳	第1C		第2C		合計	
		件数	割合	件数	割合	件数	割合
1	一致	87,091,735	99.99%	86,925,586	99.98%	174,017,321	99.99%
2	不一致	5,807	0.01%	18,864	0.02%	24,671	0.01%
3	合計	87,097,542	-	86,944,450	-	174,041,992	-

表8 在籍者、除籍者の判定用コードにおける不明コード使用件数

No	内訳	第1C		第2C		合計	
		件数	割合	件数	割合	件数	割合
1	在籍者	61,599,412	70.7%	61,737,915	71.0%	123,337,327	70.9%
2	除籍者、消除者等	25,330,715	29.1%	24,826,829	28.6%	50,157,544	28.8%
3	不明	167,415	0.2%	379,706	0.4%	547,121	0.3%
4	合計	87,097,542	-	86,944,450	-	174,041,992	-

(2) 画像データの戸籍の名寄せ

画像データの戸籍については、今回の調査ではシミュレーションの対象外としたが、戸籍等を遡る個人の統合はどこまで可能かを見極めるため、今後、シミュレーションを実施していく予定である。