

外字実態調査の作業内容

平成23年12月15日

作業内容

① 外字情報の整理

字 自治体コード_区分_外字コード 字 字 字

外字情報を1文字単位の画像データに分解



② 外字情報の分類

字 字 字 字 字 字

戸籍業務用の分類基準漢字に分類



③ 外字情報の同定

IPA = 字 自治体コード_区分_外字コード + MJ番号
IPA ≠ 字 自治体コード_区分_外字コード

包摂基準に基づき、文字情報基盤漢字と同定



④ 同定できなかった外字の調査

字 ≡ IPA 類似する文字情報基盤漢字
字 字 字 類似する他の外字との比較

同定できなかった文字をサンプル抽出して調査



⑤ 成果物

全体結果・分析
報告書

包摂基準書

同定結果一覧
(全体及び自治体別)

非同定結果一覧
(全体及び自治体別)

非分類外字
調査報告書

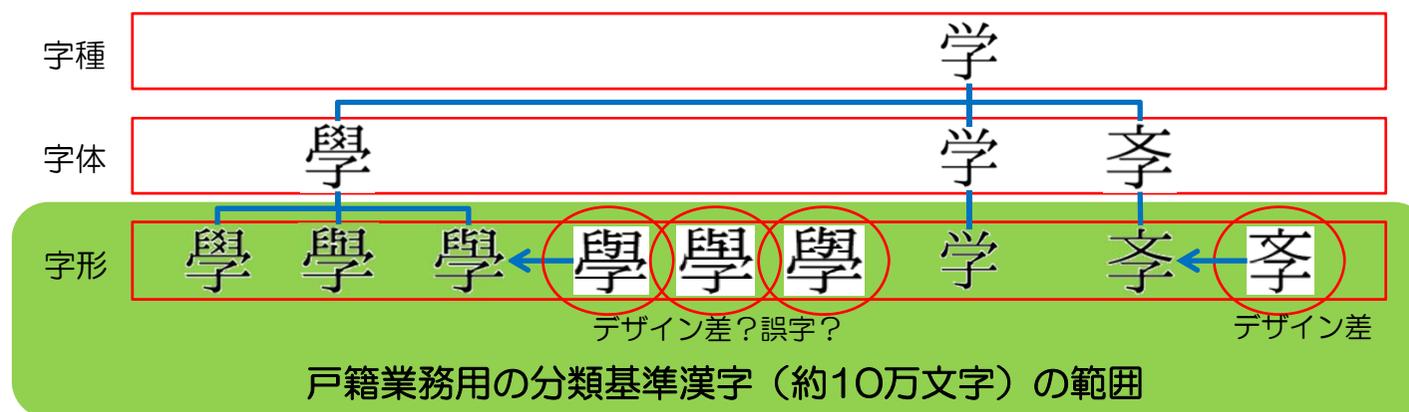
文字情報基盤漢字と戸籍業務用の分類基準漢字との関係

本作業で使用する戸籍業務用の分類基準漢字は、「常用漢字表」の考え方をベースに、戸籍電算化処理において発生したデザイン差や誤字の情報を蓄積したもので、自治体が独自に作成した多種多様な外字に対して高い網羅性を有している。

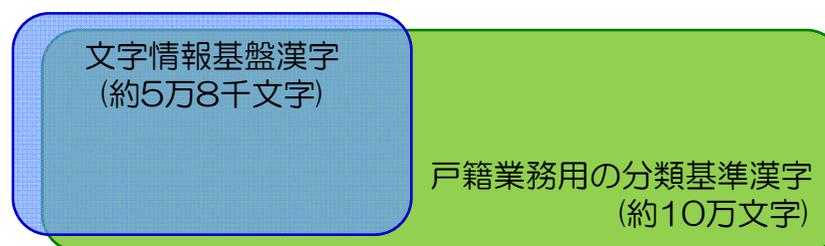
また戸籍業務用の分類基準漢字は、文字情報基盤漢字の大部分を含んでおり、文字情報基盤漢字との対応付けが可能となっている。

自治体外字を文字情報基盤漢字と同定する前段階として、当該分類基準漢字に分類することで、文字情報基盤漢字との同定作業をより効率的かつ緻密に行うことができる。

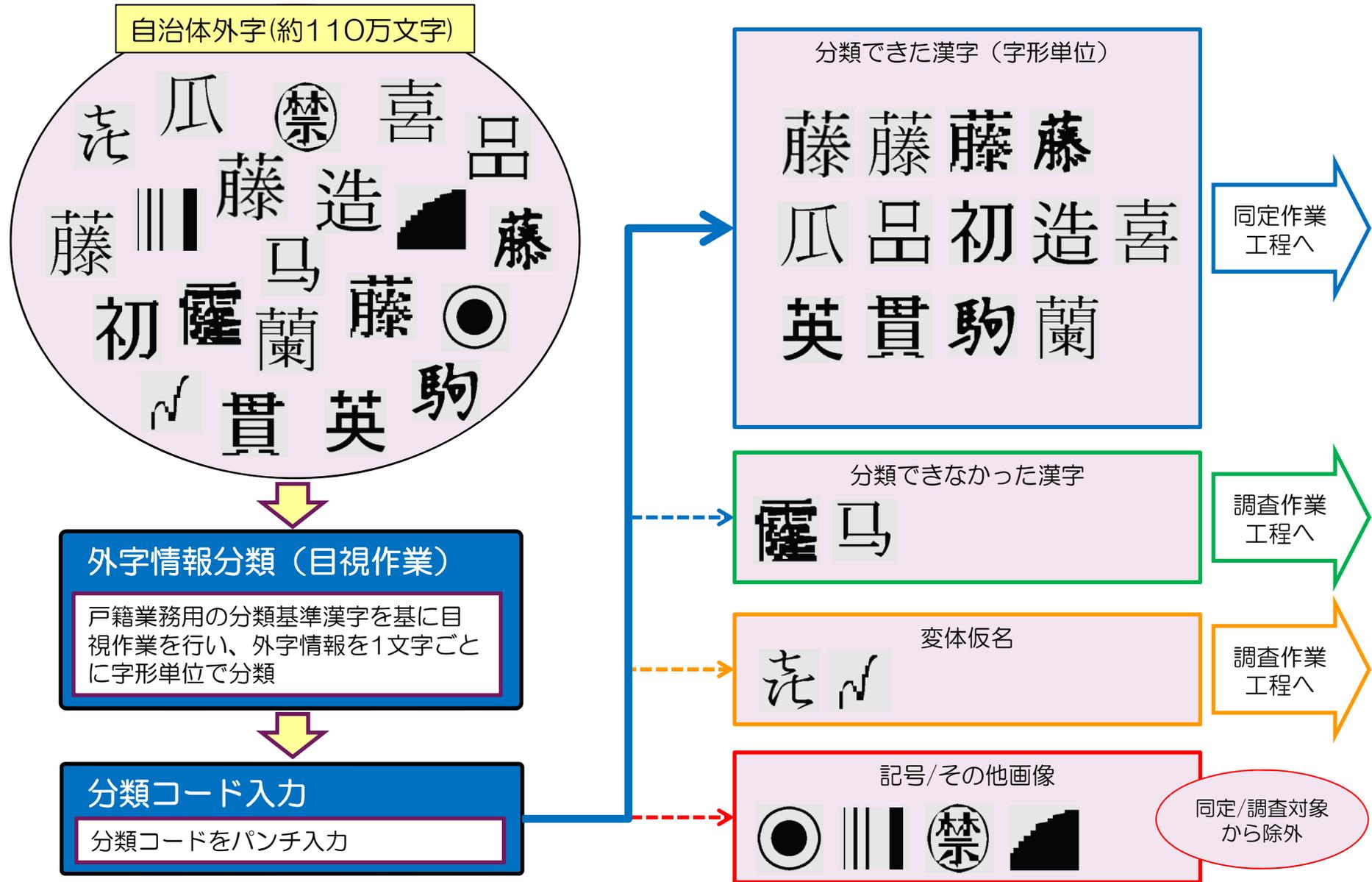
◆文字の形の階層関係に基づく分類 (平成20年度経済産業省委託事業『汎用電子情報交換環境整備プログラム成果報告書』参照)



◆文字情報基盤漢字と戸籍業務用の分類基準漢字の相関図



外字情報の分類

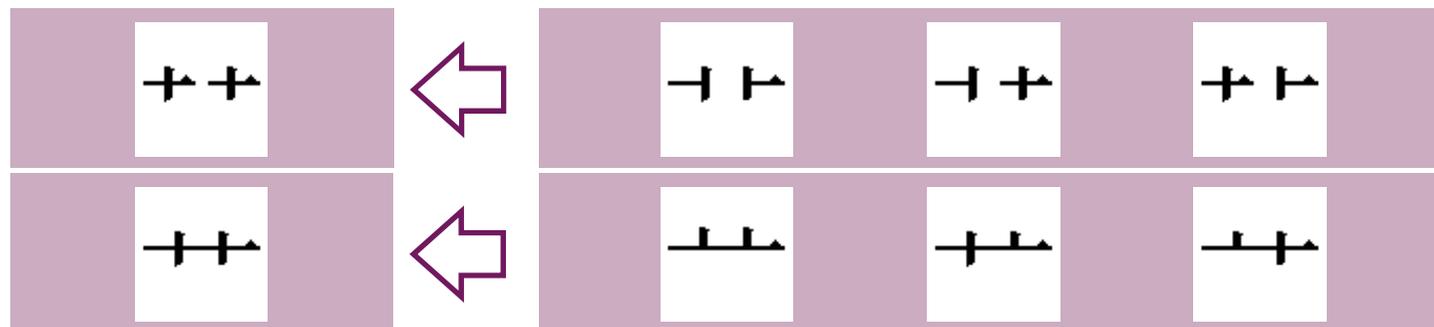


外字情報の分類方法

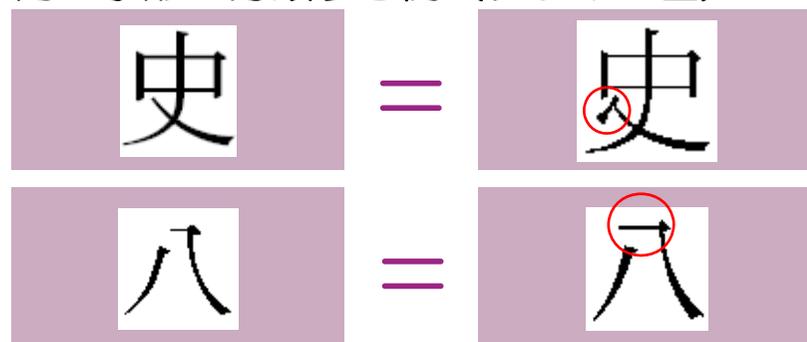
基本的な考え方

- ① 『常用漢字表』（平成22年内閣告示第2号）の考え方をベースにして、戸籍電算化処理にて発生したデザイン差・誤字の情報を蓄積した戸籍業務用の分類基準漢字を基にして分類を行う。
（字形イメージを尊重し、画数は変更しない。部首は漢和辞典の定義に従う。）
- ② 戸籍業務用の分類基準漢字に分類できない漢字は、文字情報基盤漢字を参照し、分類を行う。

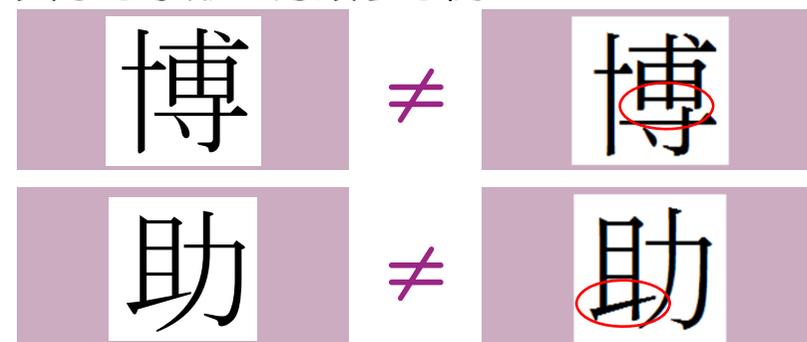
部首の判断例（草冠）



同じ字形と分類する例（デザイン差）

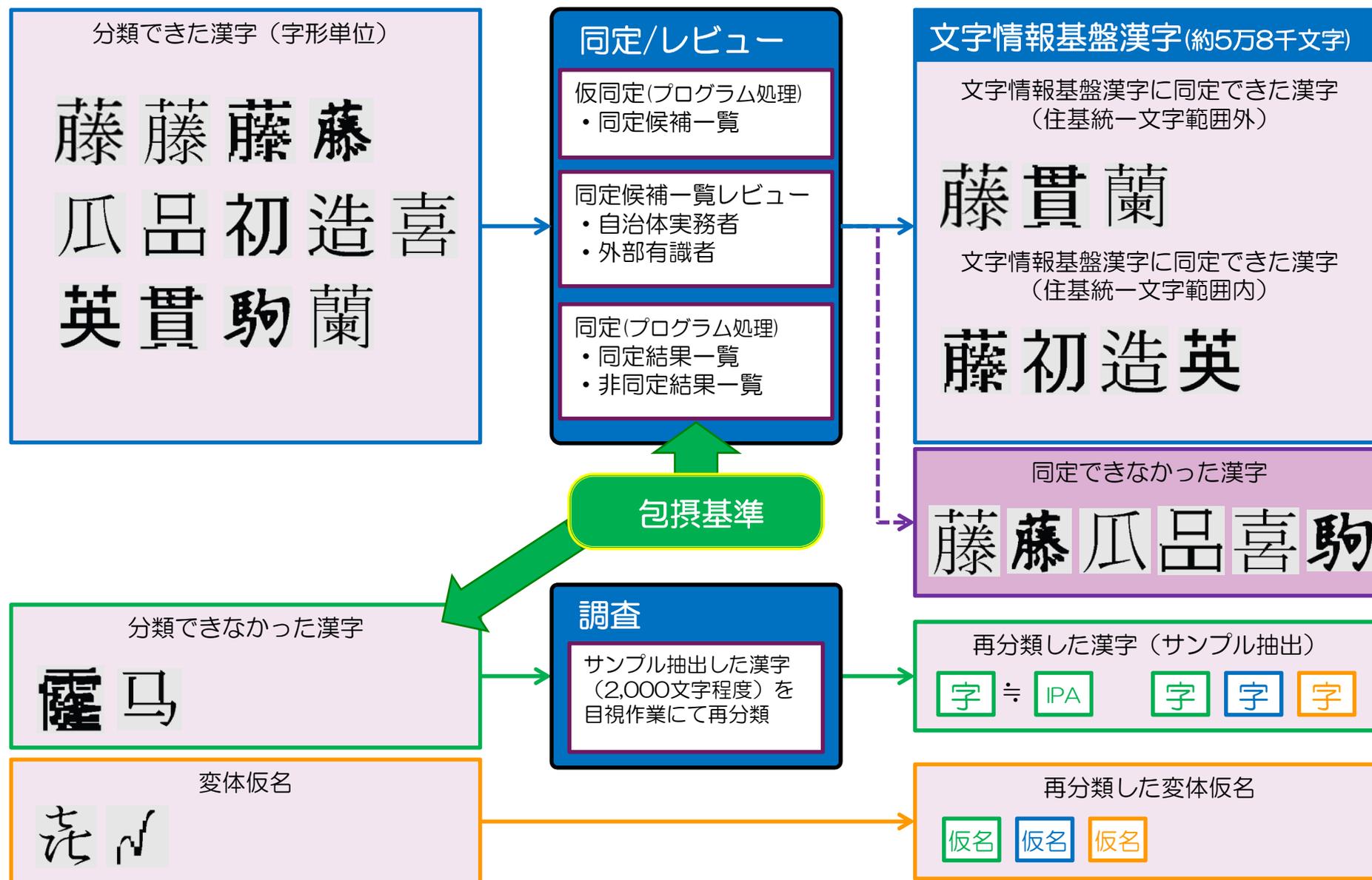


異なる字形と分類する例



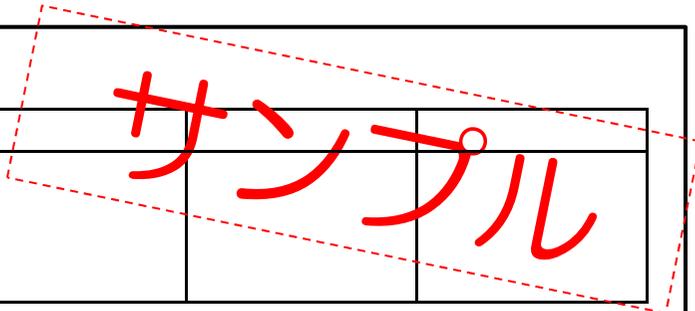
※「常用漢字表」（平成22年内閣告示第2号）より、一部内容を引用

外字情報の同定／同定できなかった外字の調査



包摂基準の策定方法（イメージ）

同定候補一覧							
MJ023082	字形一致	XXXXX_1_E33A					
藤		藤					
↑ 包摂の強弱を調整 ↓	デザイン差	XXXXX_1_E020	XXXXX_2_EE00	XXXXX_2_E200	XXXXX_1_E064	XXXXX_1_E007	
		藤	藤	藤	藤	藤	
	誤字扱い	XXXXX_3_F022	XXXXX_1_EAC5	XXXXX_1_EBCC	XXXXX_1_EFDD	XXXXX_1_E679	XXXXX_1_E465
		藤	藤	藤	藤	藤	藤



- ① 分類できた漢字は、戸籍電算化処理に用いられている包摂基準を参考に、プログラムにて同定処理を行い、文字情報基盤漢字(MJ番号)候補文字との「同定候補一覧」を作成
- ② 「自治体の実務者」及び「外部有識者」でレビューを行い、検討結果（包摂の強弱等）をプログラムに反映し、再度、同定処理を実施
- ③ 上記の経過及び検討内容を取りまとめた「包摂基準書」を作成

全体成果物（イメージ）

同定結果一覧	
MJ023082	
字形一致	XXXXX_1_E33A XXXXX_2_E09A XXXXX_1_E211
	藤 藤 藤
	サンプル
デザイン差	XXXXX_1_E020 XXXXX_2_EE00 XXXXX_2_E200 XXXXX_1_E064 XXXXX_1_E007
	藤 藤 藤 藤 藤
非同定結果一覧	
誤字扱い	XXXXX_3_F022 XXXXX_1_EAC5 XXXXX_1_EBCC XXXXX_1_EFDD XXXXX_1_E679 XXXXX_1_E465
	藤 藤 藤 藤 藤 藤

- ① 「全体結果・分析報告書」 (PDF)
- ② 「包摂基準書」 (PDF)
- ③ 「同定結果一覧」 (PDF)
- ④ 「非同定結果一覧」 (PDF)
- ⑤ 「非分類外字 調査報告書」 (PDF)

※外字情報の分類及び同定作業、レビュー内容を反映

※ MJ番号単位／文字情報基盤漢字に同定できた漢字

※分類できたが、文字情報基盤漢字と同定できなかった漢字

※分類できなかった漢字(2,000文字程度)及び変体仮名

自治体ごとの成果物（イメージ）

〇〇自治体（自治体コード + システム区分） 同定結果一覧（字形一致）		〇〇自治体（自治体コード + システム区分） 同定結果一覧（デザイン差）		〇〇自治体（自治体コード + システム区分） 非同定結果一覧	
00001 (E2E5)	MJ029181	00552 (E2E6)	MJ023082	01222 (E211)	
鯉	鯉	藤	藤	(財)	
00002 (E2ED)	MJ059295	00553 (E33A)	MJ019245	01223 (E212)	
才	才	窗	窗	(名)	
00003 (E32B)	MJ024670	00554 (E355)	MJ029073	01224 (E332)	
謎	謎	鯰	鯰	あ	
00004 (E33A)	MJ023082	00555 (E33A)	MJ023046	01225 (EE88)	
藤	藤	藏	藏	佞	
⋮	⋮	⋮	⋮	⋮	

〇〇自治体（自治体コード + システム区分）
~~非同定結果一覧~~

サンプル

各自治体ごとに「同定／非同定結果一覧」(PDF)、「対応テーブル」(CSV)を電子媒体で提供
 ※住基統一文字に含まれる漢字にはフラグを設定

品質向上策

外字情報の整理

外字情報を1文字単位の画像データに分解

1) プログラム処理によるエラーチェック

2) 作業用PDFリストの目視確認

包摂基準の策定

文字情報基盤漢字に同定する際の包摂基準を策定

1) 外部有識者による包摂基準のレビュー

2) 自治体の実務者による包摂基準のレビュー

外字情報の分類

戸籍業務用の分類基準漢字に分類

分類コードのなぞり&再確認入力（戸籍経験者）

外字情報の同定作業

包摂基準に基づき、文字情報基盤漢字に同定

1) プログラム変換による文字同定処理

2) 同定結果リストの目視照合（戸籍経験者）

3) 外部有識者による同定結果一覧のレビュー

4) 自治体実務者による同定結果一覧のレビュー

同定できなかった外字の調査

同定できなかった文字をサンプル抽出して調査

外部有識者による調査報告書のレビュー

