

# 文字情報基盤導入ガイド ver. 1.0

2014年3月18日

## 1. 本ガイドの目的

本ガイドは、平成25年6月14日の閣議決定「世界最先端IT国家創造宣言」<sup>1</sup>のⅢ.3.(1)節「利便性の高い電子行政サービスの提供」に示された、「文字の標準化・共通化に関しては、今後整備する情報システムにおいては、国際標準に適合した文字情報基盤を活用することを原則とする」との方針に沿ったシステム構築・運用を行うための具体的な技術的指針を示すとともに、文字情報一覧表<sup>2</sup>、関係国際規格等、導入に当たって参照すべき資料を明確とすることを目的としている。

## 2. 文字情報基盤について

文字情報基盤とは、2002年度～2008年度に実施された「汎用電子情報交換環境整備プログラム」の成果を出発点とし、2010年度に内閣官房情報通信技術（IT）担当室（現IT総合戦略室）、総務省、法務省、経済産業省、文化庁などの関係府省や専門家、産業界関係者が参加する文字情報基盤推進委員会を、独立行政法人情報処理推進機構（以下「IPA」と呼ぶ）を事務局として設置し、検討が開始された事業である。

住民基本台帳ネットワークシステム統一文字（以下「住基統一文字」と呼ぶ）、戸籍統一文字に収録された漢字を整理し、各文字体系との対応関係等をまとめた文字情報一覧表と、文字フォント（IPAmj 明朝フォント）を整備するとともに、それらの文字符号の国際標準化を推進している。

図1に、現在の文字情報基盤検討体制を示す。文字情報基盤ワーキンググループは経済産業省に設置され、IPAが事務局を務めている。

---

<sup>1</sup> 閣議決定「世界最先端IT国家創造宣言」

<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf>

<sup>2</sup> 文字情報基盤文字情報一覧表：<http://mojikiban.ipa.go.jp/1311.html>

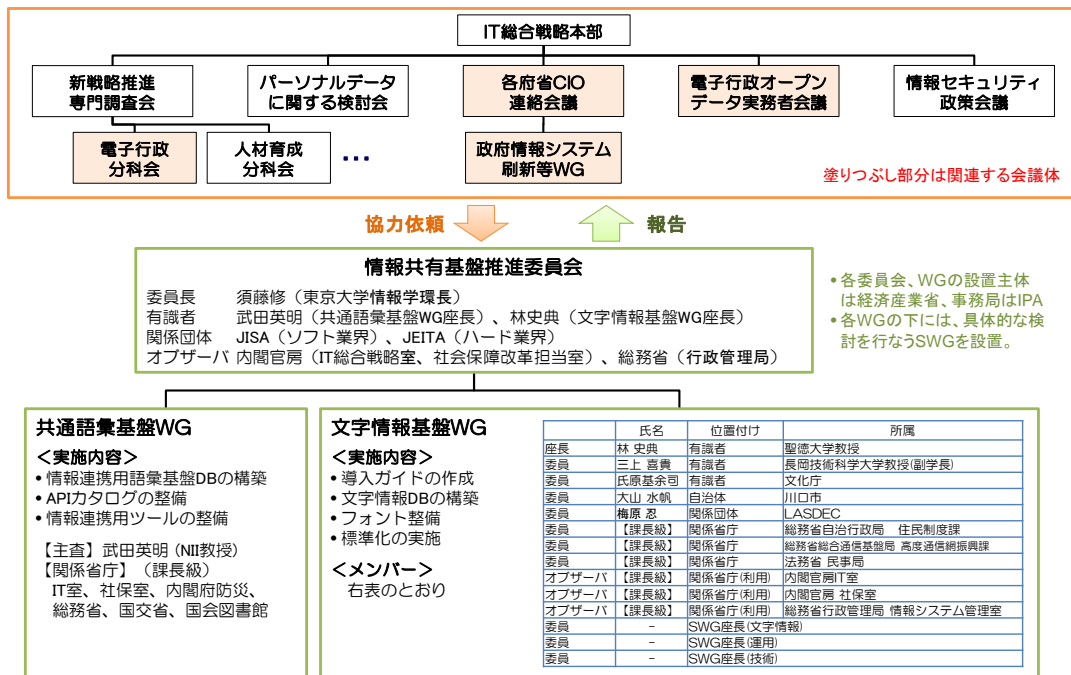


図 1 文字情報基盤の検討体制（2013 年度～）

## 2.1. 文字情報基盤で提供する成果物等

文字情報基盤の事業成果物として、IPA は、以下のものを公開している<sup>3</sup>。

### (1) IPAmj 明朝フォント

文字情報基盤で整備された文字（戸籍統一文字の漢字：55,270 文字、住基統一文字の漢字：19,563 文字を整理した結果）約 60,000 文字を収録したフォントであり、IPA フォントライセンス v1.0 により、無償公開されている。

ISO/IEC 10646 に準拠して符号化されている。符号化方式の詳細は 2.2 節、国際標準化状況は 2.3 節を参照のこと。

### (2) 文字情報基盤 文字情報一覧表（MJ 文字情報一覧表）

文字情報基盤で整備された文字について、

- 文字図形、各種文字コードとの対応関係
- 読み、画数、部首等の漢字属性情報
- SVG フォーマットによる文字図形イメージ情報

等を収録し、文字の「見える化」を図る一覧表である。クリエイティブ・コモンズ・ライセンス 2.1「表示、継承」により無償公開されている。

### (3) 報告書等

- 平成 22 年度調査事業の成果報告書

<sup>3</sup> 文字情報基盤成果物等は <http://mojikiban.ipa.go.jp/> から公開されている。

- 文字情報基盤事業の在り方に係る検討結果等
- 平成 24 年度文字情報基盤導入実証実験成果報告書  
自治体現場への導入実験の結果報告書、実験に用いたデータ（コード変換表等）等を公開
  - 関連委員会の議事概要  
を公開。

## 2.2. IPAmj 明朝フォントの符号化方式

IPAmj 明朝では戸籍統一文字、住基統一文字を包含する全約 6 万文字を図 2 のように国際符号化文字集合 ISO/IEC10646<sup>4</sup> Universal Coded Character Set (以下 UCS と呼ぶ) に準拠して符号化している。

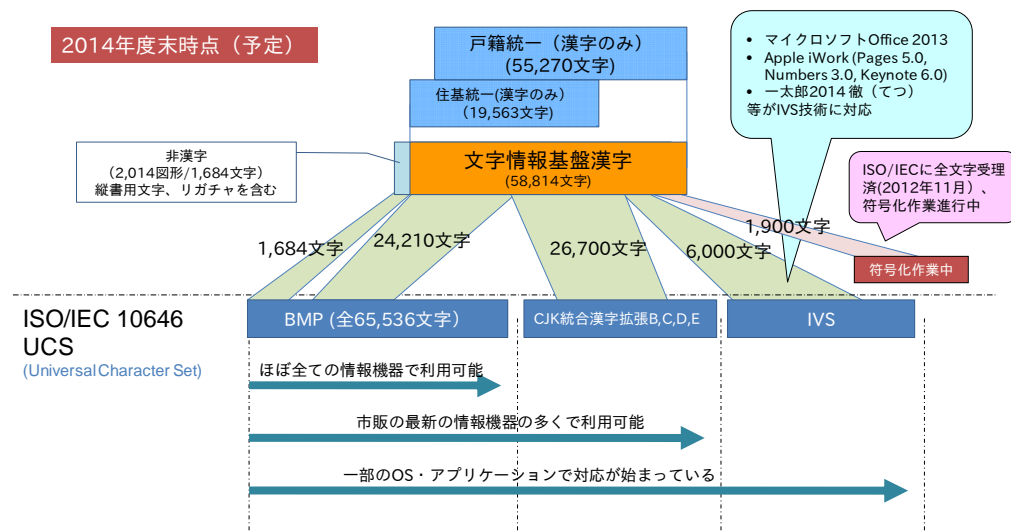


図 2 文字情報基盤の符号化状況

図に示すように、全約 6 万文字は、

- BMP<sup>5</sup>と呼ばれる領域へ符号化された文字：約 26,000 文字
- 拡張領域<sup>6</sup>と呼ばれる領域へ符号化された文字：約 26,000 文字
- IVS<sup>7</sup>と呼ばれる方式で符号化された文字：約 6,000 文字（2014 年中）

<sup>4</sup> ISO/IEC 10646 Universal Coded Character Set(UCS):2014 4th edition。対応する国内規格は、JIS X 0221:2014。(いずれも 2014 年中に発行予定)

<sup>5</sup> BMP (Basic Multilingual Plane): 16 ビットで文字を指定できる範囲で、「基本平面」と呼ばれる。すべての情報機器はこれに対応している。

<sup>6</sup> 文字の指定に 32 ビットを要する領域。一部旧式の情報機器は対応していない場合がある。この領域への符号化は ISO/IEC において順次進められており、「拡張 E」が 2014 年度に出版予定。「拡張 F」の標準化作業が進行中である。

<sup>7</sup> IVS(Ideographic Variation Sequence / Selector) : 「字形選択子」と呼ばれるもので、字形が類似して

- 符号化が未了の文字：約 1,900 文字（2014 年中）  
の各領域へ符号化されている。

これら領域毎に、ソフトウェア（アプリケーションソフトや OS、ミドルウェア等）の対応状況が異なる。特に IVS は、最新の標準に基づくものであり、2014 年 2 月時点で、PC 用の市販ワープロソフト等はほぼ対応が完了しているが、基幹業務系ソフト等の対応はまだ開始されたばかりである。

### 2.3. 国際標準化状況

文字情報基盤で整備された約 6 万文字は UCS に従って符号化を行っているが、図 3 に示すように、2014 年末において約 1,900 文字が国際標準化機構（ISO/IEC JTC1）における符号化作業中のものとなる。これらについては符号化作業完了まで 2～3 年を要するものとみられる（遅くとも 2020 年までには完了の見込みである）。

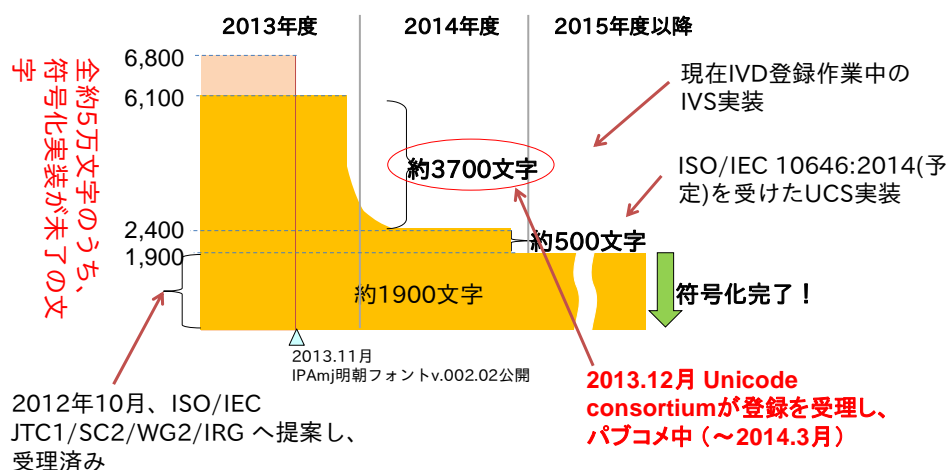


図 3 文字情報基盤の符号化作業状況

### 3. 文字情報基盤導入にあたっての基本方針

文字情報基盤を情報システムへ導入するにあたっては、以下の点に配慮する必要がある。

- 1) 国際標準と矛盾しない技術を用いること
- 2) 既存システムとの共存性、導入コストの低さなど、経済合理性を考慮すること
- 3) 技術の進展、標準化の進展に沿った適切な手順を踏むこと

上記 1) は、国際標準と矛盾しない技術を用いることにより、市場に在る一般的情報機器の活用が可能となること、機器間、組織間での文字情報の交換に係る変換等のコストが不要となること、等から、電子行政システムの構築・運用に係るコストダウンにつながる事が期待さ

---

おり、符号的には区別されない文字について、その字形を指定するための符号。

れる。国際標準として参照すべき事項については、8章を参照のこと。

上記2)については、行政機関においては人名漢字等を処理するための様々なバージョンの情報機器が稼働しており、これらを一度に置き換えることなく、それらと新規システムとの共存を図る方策をとることにより、移行に係るコストの低減を図ることが必要である。

上記3)については、文字情報基盤の国際標準化が現在進行中であること、また、情報機器の対応についても過渡的状況にあることから、これらの進展の各段階での機器間での相互運用性、共存可能性を十分配慮した導入計画を立てることが求められる。

また、最終的には、行政で用いられる各種情報システム間での文字情報の交換を、UCSに基づく一貫した様式で行うシステムへ移行するべきであることを視野に入れる必要がある。

#### 4. 利用する文字の範囲について

実際に業務で使用する文字の範囲等については、当該業務に係る法令等に従うことが必要であるが、一般的には、業務効率や相互運用性を考慮し、業務内容に応じた適切な大きさの文字範囲を使用するよう検討することが必要である。

表1に、適用場面ごとに想定される文字/文字図形の範囲と符号化方式の例を示す。文字の絞込みの定義に係る国際標準については、8.3節「文字集合の絞込みに係る定義」を参照のこと。

表1 適用業務と文字の範囲（参考）

		文字情報基盤 で整備された 文字図形（字 体レベル <sup>8</sup> ）	文字情報基盤 で整備された 文字図形(UCS レベル <sup>9</sup> )	JIS X 0213	JIS X 0208	常用漢字
文字/文字図形集 合の定義		IVD Moji_Joho collection	MJ 文字情報一 覧表に含まれる UCS 符号位置	8.3 節参照	8.3 節参照	JIS X 0213:2012 附属書 12
組織内 に閉じ た業務	戸籍業務	○ <sup>(A)</sup>	△			
	住民記録業 務	○ <sup>(A)</sup>	△			
	法人関連業 務	○ <sup>(A)</sup>	△			
	生徒名簿等 （字形に正確 さが求められ る場合）	○ <sup>(A)</sup>	△			
	生徒名簿等 （日常業務）			○	△	
	一般的な書			△→○ <sup>(B)</sup>	○→△ <sup>(B)</sup>	

<sup>8</sup> UCSに加え、IVS等を指定し、細かな文字図形を区別する利用法。

<sup>9</sup> UCSのみを指定し、IVSを指定しない利用法。この場合、フォントの当該UCS符号に対するデフォルト図形として実装された文字図形が表示される。

	類作成					
	災害等緊急 時対応シス テム			△→○ <sup>(B)</sup>	○	
外部 への 通知 等	一般向け広 報文書			△	○	○
	正確な人名 表記を要す る通知文等	○ <sup>(C)</sup>	△ <sup>(C)</sup>			
	情報公開等	△ <sup>(C)</sup>	○ <sup>(C)</sup>	○	○	○
外部 から の申 請等	一般的な申 請等			△→○ <sup>(B)</sup>	○→△ <sup>(B)</sup>	
	正確な人名 表記を要す る申請等	○ <sup>(D)</sup>	△ <sup>(D)</sup>			

<sup>(A)</sup>必要に応じ、文字情報基盤で整備された文字図形以外の文字図形を、「外字」等として利用することが有り得るが、できる限り文字情報基盤の文字範囲に収まるよう、調整することが望ましい。

<sup>(B)</sup>市場にある一般的情報機器の対応状況に応じ、変化する。

<sup>(C)</sup>JIS X 0213 範囲以外の文字については、図形イメージによる表示、伝送を検討する。

<sup>(D)</sup>JIS X 0213 範囲以外の文字については、表示を図形イメージによるものとするほか、Web上に漢字入力を支援するサービスを設けることを検討する。

## 5. システム間接続の運用の基本方針

### 5.1. 文字図形の厳密な一致性が不要な場合

一般事務向けのシステム等では、文字図形について厳密な一致性は求められず、JIS X 0213 や UCS で採用されている包摂あるいは統合の考え方（文字図形の差がわずかである文字については同一の符号を与えるという考え方）に従い、文字情報を交換することが適当である。

この場合は、2つのシステムに異なるデザインのフォントが実装されている場合にも、UCSの符号のみを交換すれば十分である。

ただし、同じ UCS を使用しているシステムでも、一方が JIS X 0208 相当の文字集合のみをサポートしており、他方が JIS X 0213 相当の文字集合を使用しているといった場合等、使用する文字図形集合の大きさに差がある場合においては、大きな集合から小さな集合への（N 対 1 の）対応関係をまとめた「縮退マップ」の採用（5.3 あるいは、文字を図形イメージ情報として伝送する（8.4 参照）等）の必要が生じる。

### 5.2. 人名等文字図形の厳密な一致性が求められる場合

人名を扱う行政事務等では、使用する文字図形に対し、戸籍に記載された文字図形との間、あるいは過去に使用した文字図形との間等での高度な図形的 consistency が求められる場合がある。

この場合、異なるシステム間での文字情報交換について、以下のような手法が考えられる。

(1) 文字情報基盤文字情報一覧表を媒介とした対応付け

文字情報基盤で整備された文字図形は、戸籍統一文字及び住基統一文字の文字図形との間で 1 対 1 の対応関係を持つよう作られている。人名に用いられる漢字を処理するための既存の情報システムは、人名が戸籍由来であることから、戸籍統一文字、住基統一文字に同定可能な字形を持つ独自フォントを採用していることが多い。

その結果、これらシステムで用いられている多くの文字図形は、文字情報基盤が整備した文字図形に同定可能である。

そのため、固有の文字体系を持つシステムの持つ文字図形を、一旦文字情報基盤の文字図形に同定し、その対応関係を明らかにしておけば、これを媒介にして、他のシステムで用いられている文字図形と、容易に対応することが可能となる。

このような対応法を用いることにより、N 種類の文字体系について、それらと文字情報基盤の文字図形との対応関係を記述した N 通りのマップを作るだけで、N×N のシステム間を接続することが可能となり、対応付けの効率が飛躍的に高められる。

また、ある文字体系と文字情報基盤との対応関係のマップは、当該文字体系の提供者がその責任において作成することができ、そのマップと他の体系について他の体系の提供者が作ったマップを組み合わせることで、最終的な文字対応をとることができることから、文字の対応マップ作成に係る責任分界が明確となる。結果、責任分界の面でも、文字情報交換に係るシステム構築の容易性が拡大する。

このような対応付け方法の有効性は、IPA の実施した実証実験においても確認されている<sup>10</sup>。

将来的には文字情報基盤の文字図形は、全て UCS および IVS によって参照できるようになる予定であるが、現在は ISO/IEC JTC1 での規格化手続きの進行中であり、2014 年度末においても約 1,900 文字については国際標準化された符号による指定はできない。そのため、変換マップ等において文字情報基盤で整備された文字図形を識別するためには、将来的には UCS+IVS を用いることが推奨されるが、当面は「MJ 文字図形名」を用いることが必要となる。

相互に合意した組織の間では、これらの文字を UCS で規定された「私用面」へ割り当てた、いわゆる「外字」のコードを交換することも有り得る。ただし、その場合においては、「合意した組織」外へ、当該外字コードが流出しないよう、十分な配慮を行う必要がある。また、外字を使用した文書を保存する際には、外字コードが不変なものでないことを十分配慮する必要がある。

(2) その他の方法（図形イメージ情報の交換）

文字情報基盤の文字へ同定できない文字図形については、字形そのものを図形（イメージ）情報として交換する必要がある、その技術的方法は、8.4 に示した。

各交換方式の特徴を表 2 に示す。

対象業務における文字図形の厳密性への要求、アプリケーションの対応状況等に応じ、適切な方式の採用を検討することが求められる。

---

<sup>10</sup> 文字情報基盤を自治体等のシステムで活用する実証実験の結果については、<http://mojikiban.ipa.go.jp/category/03/03002> に掲載されている。

表2 交換手段の特徴

交換方式	相互運用性	対応ソフトの豊富さ(一般アプリケーション)	対応ソフトの豊富さ(基幹システム)	伝えられる文字図形の厳密さ
UCS(BMP)	◎	◎	◎	△
UCS(BMP+拡張領域)	◎	◎	○	○
UCS+IVS	◎	△→○	× <sup>(B)</sup>	◎
文字図形情報の交換 (MJ 文字図形名の活用)	○	△(ブラウザ等)	× <sup>(B)</sup>	◎
文字図形情報の交換 (独自文字図形の利用)	○ <sup>(A)</sup>	△ <sup>(A)</sup> (ブラウザ等)	× <sup>(B)</sup>	◎
外字の利用(UCS 私用面 <sup>11</sup> の利用)	◎/× <sup>(C)</sup>	◎	○	◎/× <sup>(C)</sup>

<sup>(A)</sup>文字の図形的形状を伝えることができるが、その文字を特定するための情報を伝えることはできず、検索や意味の解釈に支障をきたす場合がある。

<sup>(B)</sup>システム内では外字へ変換するなどの作りこみは可能。

<sup>(C)</sup>当該外字の利用について相互に合意した組織においては厳密な文字図形情報を伝達できるが、当該外字コードを当該組織外に流出させた場合においては、文字化け等の障害を起こす。また、外字コードを含んだ文書を保存するにあたっては、外字コードが不変なものでないことを十分配慮する必要がある。

### 5.3. 縮退変換の活用

2つのシステムの扱える文字集合の大きさが異なる場合、大きな文字集合を小さな文字集合にマップさせる、「縮退変換」を行うという考え方がある。

例えば、災害時の対応等、被災者支援等のために、文字図形の厳密な識別よりも、市販の情報機器等を活用する即応性、利便性、検索性などが優先される場合がある。

市販情報機器では行政機関等が用いている人名漢字を扱えない場合も多く、変換に大きな手間を要したり、文字化け等が起きることが東日本大震災の経験でも多く指摘された。

このような場合に備え、人名などを扱うために用意された大きな文字セットを、JIS X 0208等の最小限の文字セットへ縮退変換するためのテーブルをあらかじめ用意することが求められる。

また、災害対策といった特殊な場合でなくとも、通常業務においても、業務効率やシステム構築コストの関係から、文字図形の数を抑制することは重要である。

文字情報基盤が整備した文字図形と1対1に同定できない文字のうち、通常業務で多く用いられている文字等については、8.4のような図形情報による交換だけでなく、文字情報基盤が整備した文字への縮退変換マップの整備等も検討する(7.1参照)。

なお、一旦、小さな文字図形集合に縮退された情報を大きな文字図形集合に再変換した場合、

<sup>11</sup> UCSの私用面(15面、16面 : 0F0000-10FFFF)には、相互に合意した組織の間で「外字」を登録し、私的に(外部へ流出させずに)使用することが認められている。



大きな文字図形集合における識別粒度が保証されないことに留意する必要がある。

## 6. 導入の手順

最終的には、文字情報基盤が整備した文字図形の全てが UCS および IVS により扱えることを目指している。

また、究極的には、文字情報基盤が整備した文字図形を使う全てのシステムが IVS に対応することが望まれる。

しかし、これらは、標準化およびソフトウェアの対応状況の進展に応じ、段階的に進めてゆく必要がある。

当然、その過程のどの段階においても、既存システムと新規システムとの間の相互運用性は維持しなくてはならない。

また、将来的には文字情報基盤が整備した文字図形以外の文字図形を全く使用しないことにするという選択肢も有り得るが、本ガイドでは、そのような選択肢は想定せず、こういった文字については、図形（イメージ）情報として 8.4 に示したような方式に則って交換可能とすることを目標とする。

システム構築は、同一組織内に閉じたシステム間における情報交換と、組織を超えた情報交換とに分けて考える必要がある。「組織を超えた」情報交換においては、例えば「外字」フォントを共有する（8.5 参照）といったことを想定しない方法を検討する必要がある。

これらを一覧にまとめると表 3 の様になる。

表 3 各組織・各段階における導入の方針

	短期	中・長期
組織内部（内部実装）	特に内部実装を変更する必要は無い。	IVS を含む UCS が使用可能なシステムの導入を図る
行政組織間・自治体間（文字情報の交換）	図 4 に示す「交換表現」により、状況と必要に応じ、UCS、IVS、図形情報等を交換する。システムには、「交換表現」に対応した変換システムを装備し、交換された情報を内部表現へ変換する。特に相互に合意のとれた組織間では、私用面に割り当てたコードの交換も検討する。	文字情報一覧表に記載された文字図形は、すべて IVS を含む UCS 表現を用い、その他の図形文字は図形（イメージ）情報として交換する
組織外（開放系への情報発信）	使用する部分文字レパートリ（8.3 参照）を明記する。 符号化方式（Shift-JIS など）を明記する。 符号化範囲外にある文字図形は図形（イメージ）情報とし	使用する部分文字レパートリ（8.3 参照）を明記する。 IVS を含む UCS の符号化方式を用いる IVS が使用できない環境へ向けた発信のため、必要に応じ

	て発信する。 文字図形を厳密に識別する必要のない外部組織等への情報公開、災害対応等のために縮退変換マップを活用する。縮退のためのマッピングテーブルは原則公開する。	て図形（イメージ）情報の伝送を併用する。 文字図形を厳密に識別する必要のない外部組織等への情報公開、災害対応等のために縮退変換マップを活用する。縮退のためのマッピングテーブルは原則公開する
組織外（開放系からの入力）	使用可能な部文字レパートリ（8.3 参照）を制御する。 別途、手動による入力手段（GUI による選択入力等）を提供する。	MJ 文字図形のすべての図形文字を識別入力可能とする。 別途、手動による入力手段（GUI による選択入力等）を提供する。

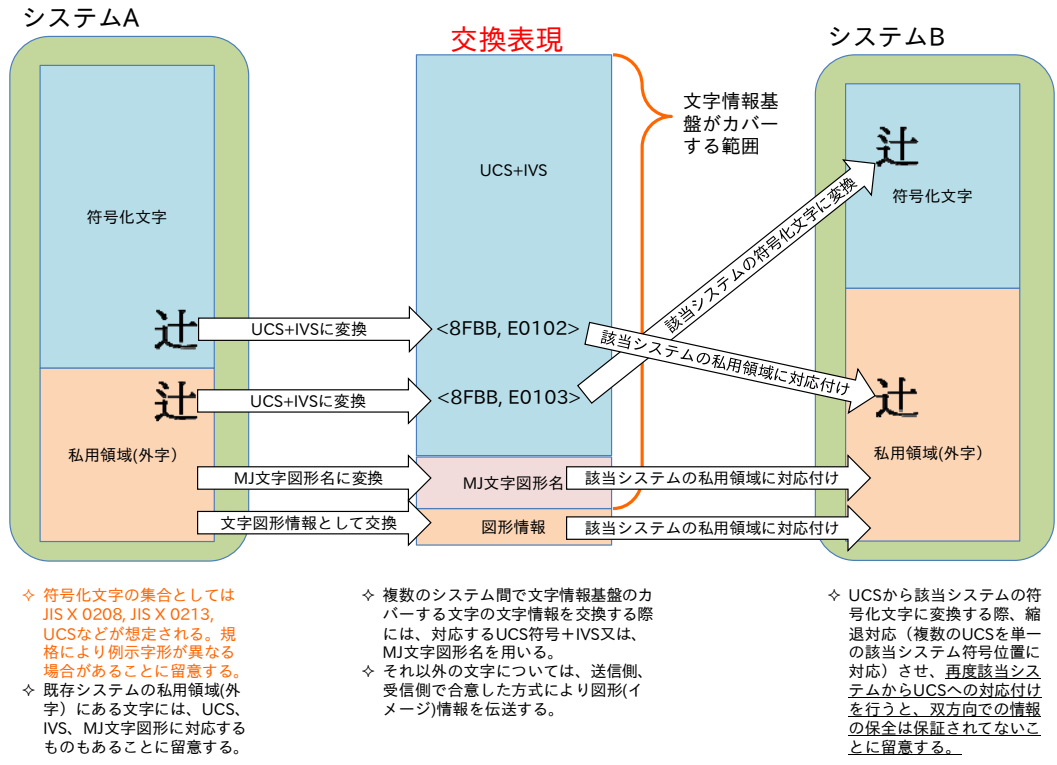


図 4 文字情報交換のイメージ

## 7. 今後整備すべき事項

### 7.1. 行政システムの構築等で参照するための一覧表等

文字情報の円滑な交換を支援するため、文字情報基盤として、以下のような情報を検討し、整備、公開するものとする。

- (1) 文字情報基盤文字情報一覧表に含まれる UCS 符号位置を列挙した一覧表
- (2) 文字情報基盤が整備した文字図形のうち、対応する UCS 符号位置を持たないものすべての MJ 文字図形名の一覧表
- (3) 「交換表現」(図 4 参照) のリファレンスとなるスキーマ類
- (4) 文字情報基盤が整備した文字図形を、以下のように分類して掲載した一覧表および、それらの文字図形を直接参照するための URI
  - A) UCS に対応するデフォルトグリフ<sup>12</sup>  
MJ 文字図形の通常のオープンタイプフォント実装において、IVS を用いなくても該当文字図形が表示される範囲の文字図形
  - B) UCS に対応するデフォルトグリフ以外の文字図形  
MJ 文字図形で対応する UCS が存在するもののうち、IVS を用いなければ該当文字図形が表示できない文字図形
  - C) 対応する UCS 符号位置を持たない MJ 文字図形  
MJ 文字図形一覧表において、現在、対応する UCS 符号位置を持たない文字図形 (ISO における標準化作業過程にあるもの。「拡張 F」として規定されることが想定される)。これらの一覧表を参照することにより、IVS に対応しないシステムでの内部表現のために IVS 文字を外字として作成し、使用する場合、符号位置を持たない文字図形をシステム内で外字として使用する際等の参考とすることができる。
- (5) 変体仮名の文字図形
- (6) 行政で多く用いられるにもかかわらず、文字情報基盤で整備した対象となっていない文字図形について、文字情報基盤で整備された文字図形に縮退変換するための変換テーブル。
- (7) 文字情報基盤で整備された文字をより小さな文字集合へ縮退対応させるための対応表。

### 7.2. 国際標準化活動

現在、Unicode Consortium が管理する、IVD に登録されている Moji\_Joho Collection を拡充し、文字情報基盤が整備したすべての文字図形に対して IVS が一意に定まるよう拡張登録を行う作業を進めている。

また、残された 1,900 文字については、その符号化提案が国際標準化団体に受理され、符号化作業が進行中である。

さらに、以下のような標準化を計画している。

- (1) ISO/IEC JTC1/SC2/WG2 に対し、文字情報基盤で整備された全ての文字図形に対応する UCS の符号位置を、UCS の部分レパートリとして追加提案する。
- (2) 公用文への適用が義務づけられている常用漢字表に含まれる漢字及び、日本語のローマ字

---

<sup>12</sup> デフォルトグリフは、IVS を明に指定しない場合に表示される文字図形である。

<http://mojikiban.ipa.go.jp/1292.html> 参照

表記に必要となる文字（長音記号付きのローマ字等）についても、併せて部分レパートリとしての提案を行う。

- (3) JIS X 0221 の附属書 JA において、Moji\_Joho collection を参照することの提案を対応する国内委員会を行う。
- (4) 変体仮名の符号化につき、ISO/IEC JTC1 WG2 へ提案する。

UCS の CJK 統合漢字拡張 F の規格化及び IVD Moji\_Joho collection の拡張登録が終了すれば、文字情報基盤文字情報一覧表に記載されている文字図形のすべてを IVS で記述することが可能となり、文字図形集合として、Moji\_Joho Collection を指定するだけで、文字図形集合を特定することが可能となり、当該文字図形セットに対応する民間のフォントの活用等を含め、情報システムの利便性が、いちじるしく向上することが期待される。

### 7.3. 最終目標に至るまでのロードマップ

	国際標準化活動	文字情報基盤事業	推奨される新規システム 内部実装
2012 年	IRG で拡張 F 審議開始 (済)		UCS (UTF-16) UCS 実装文字以外の私用 面へのマッピング
2013 年		文字情報一覧表 003.01 公開 IPAmj 明朝 002.02 公 開	UCS (UTF-16) UCS 実装文字以外の私用 面へのマッピング
2014 年	IVD Moji_Joho collection (第 1 次) 公開 MJ 文字図形相当の部分文 字レパートリの UCS Annex A への追加を ISO/IEC JTC1 に提案 IVD Moji_Joho collection の拡張登録 (第 2 次)	文字情報一覧表バージ ョンアップ IPAmj 明朝バージョ ンアップ	UCS (UTF-16) UCS 実装文字以外の私用 面へのマッピング UCS+IVS の実装の推奨
2018 年	拡張 F の標準化完了 IVD Moji_Joho collection 追加登録 (第 3 次) JIS X 0221 改正 (Annex JA の増補改正)	文字情報一覧表バージ ョンアップ IPAmj 明朝バージョ ンアップ 文字情報基盤事業の成 果物と国際標準の整合 性の達成	UCS+IVS の実装促進
2020 年			UCS+IVS の一層の促進

## 8. (参考) 文字情報の交換に係る標準規格等

### 8.1. 符号化文字集合と文字図形集合<sup>13</sup>

図 5 に示すように、UCS では細かな字形の差異を区別しないことを原則としており、その結果、行政の業務上で区別されるべき複数の文字図形に対して同一の符号（文字コード）が与えられる場合がある。

UCS では同一符号を与えられた複数の文字図形から一つを指定する手段として「字形選択子」(Variation Selector : VS) という特別な符号列を定義しており<sup>14</sup>、この VS を「文字符号」に付記することにより、同一の文字符号を持つ文字図形から一つの文字図形を選択して指定することができる。この符号と字形選択子 (VS) の列を、Ideographic Variation Sequence (IVS) と呼ぶ。

IVS によって指定することのできる文字図形は、Unicode Consortium に登録するむねが UCS で規定されている。Unicode Consortium ではこの文字図形を IVD (Ideographic Variation Database)<sup>15</sup> として管理している。

文字情報基盤で用いている IVS は、現在、この IVD 中の “Hanyo-Denshi collection” という集合を用いているが、情報規格調査会は現在、文字情報基盤が整備した文字図形のうち未だ IVD に登録されていない文字図形を含めた文字図形全体を、改めて “Moji\_Joho collection” として、新規登録する手続きを進めており (図 3 参照)、2014 年中にはその手続きが完了する見込みである。その結果、文字情報基盤で整備された文字図形のうち、約 1,900 文字図形を除くすべての文字図形が、国際標準に従って指定できるようになる<sup>16</sup>。

2014 年度以降に出荷される一般 PC 向けの OS、Web ブラウザ、ワープロ等のソフトウェアのほとんどが IVS に対応しており、今後、IVS に対応した環境は急速に含むと考えられる。

### 8.2. エンコーディング方式 (encoding scheme)

8.1 に述べた規定に従って符号化された文字を、通信時やファイル格納時等に実際にコンピュータが利用できる形式に変換する方式。どのような方式でビット列化するかを定めた規定であり、UCS の場合、UTF-8、UTF-16、UTF-32 等が用いられる。

---

<sup>13</sup> 本ガイドでは、《符号化文字》という用語を UCS で定義されている “Character” の訳語として、《文字図形》という用語を ISO/IEC TR 15285 “Information technology – An operational model for characters and glyph” で定義されている glyph の訳語として用いる。なお、UCS の翻訳規格である JIS X 0221 では、《文字》を “Character” の訳語として用いていることに留意する必要がある。

<sup>14</sup> IVS の概略については、<http://mojikiban.ipa.go.jp/1292.html> を参照のこと。

<sup>15</sup> <http://www.unicode.org/ivd/>

<sup>16</sup> 文字情報基盤が参照する IVD の collection は、2014 年度中に “Hanyo-Denshi collection” から “Moji\_Joho collection” へ変更されることになるが、多くの IVS 値は新 collection へ継承されており、また、旧フォントを前提にして作成された文書ファイル等に対する、新フォントの上位互換性は保たれる。

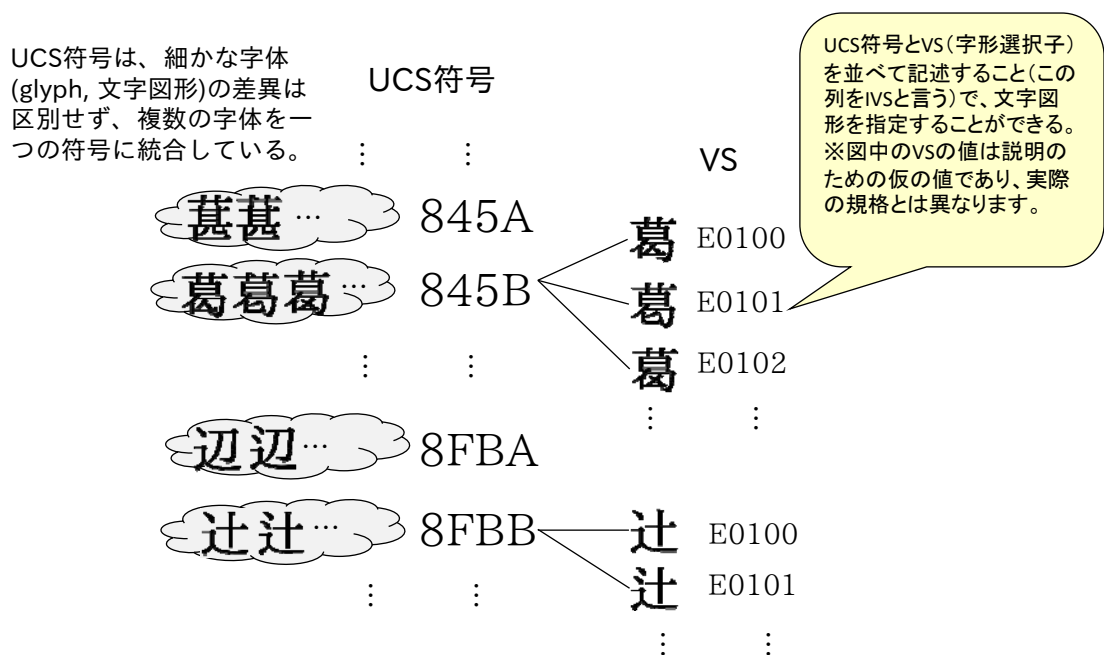


図 5 IVS (字形選択子) について

### 8.3. 文字集合の絞込みに係る定義

UCS には、ローマ字、漢字、ハングル、アラブ系文字、インド系文字、その他多様な文字が規定されている。その中で、日本の行政業務で必要とするのは一部であり、UCS で規定された文字全体の中の、どの様な部分集合を、行政向け情報システムの調達等において指定するかを意識する必要がある<sup>17</sup>。

UCS では、用途に応じて絞り込まれた文字集合を「UCS の部分文字レパートリ」と呼び、その Annex A の章において多数規定している。これらのうち、日本の行政現場の実務に関わるものとしては、以下のようなものがある。

- BASIC JAPANESE (基本日本文字集合)  
JIS X 0201, ISO/IEC 646 の IRV 及び JIS X 0208 に相当する基本部分集合。
- JIS2004 IDEOGRAPHICS EXTENSION (JIS2004 拡張漢字集合)  
JIS X 0213 の第 3 水準及び第 4 水準漢字部分に相当する拡張用の部分集合。
- JAPANESE NON IDEOGRAPHICS EXTENSION (拡張非漢字集合)  
JIS X 0213 の非漢字部分 (ただし, JIS X 0208 にも含まれるものを除く。)に相当する拡張用の部分集合。

<sup>17</sup> もし、「UCS に対応する」とのみ仕様書に書いてしまうと、仕様に合致した製品の入手は極めて困難なものになってしまう。

- COMMON JAPANESE (通用日本文字集合)

a)に示した BASIC JAPANESE に、日本の市販 PC 等で広く用いられている拡張を加えた部分集合。Shift-JIS を用いるシステムで、しばしば用いられる。

日本文字部分レパートリのうち、BASIC JAPANESE 及び COMMON JAPANESE は、それぞれを単独で又は他の日本文字部分レパートリと組み合わせて用いることが想定されている。

他の三つのレパートリは、単独で用いることは想定されていない。

JIS X 0213 2004 に規定された文字の集合を、UCS における「部分レパートリ」の定義の言葉で表現した場合、

- (1) “BASIC JAPANESE, JIS2004 IDEOGRAPHICS EXTENSION 及び JAPANESE NON IDEOGRAPHICS EXTENSION の組合せ”

と表現することとなる。

その他、非漢字等の必要性や、市販パソコンにおける拡張文字との互換性への必要等に応じ、

- (2) “COMMON JAPANESE 及び JIS2004 IDEOGRAPHICS EXTENSION の組合せ”
- (3) “COMMON JAPANESE, JIS2004 IDEOGRAPHICS EXTENSION 及び JAPANESE NON IDEOGRAPHICS EXTENSION の組合せ”

といった指定が有り得る。

これ以外の組み合わせは、通常業務での必要性が考えにくく、また、市販システム等などの互換性の観点から推奨されない。

文字情報基盤で整備した文字図形全体の集合を指定する際は、上記 3 種類のいずれかの組み合わせ指定に加え、における” Moji\_Joho collection” の文字図形集合を参照することとなる。

Moji\_Joho collection については、ISO/IEC 10646 の国際一致規格である JIS X 0221 の次期以降の版において参照するようにすることを検討する (7.2 参照)。

ただし、文字情報基盤で整備された全 6 万文字図形のうち、約 1,900 文字については、ISO での標準化作業が進行中であることから、これらが UCS および IVS で指定できるようになるにはしばらく時間がかかる。それまでの間、この範囲の文字図形については、文字情報基盤で整備した文字情報一覧表で示された MJ 文字図形名の集合として指定する必要がある (7.1 参照)。

#### 8.4. 文字図形イメージの交換

UCS や IVS を持たない文字図形については、情報システム間で、それを図形 (イメージ) として交換する必要がある。図形 (イメージ) を交換するための規格として広く知られているものに、SVG<sup>18</sup>、PNG<sup>19</sup>がある。

文字図形に係るデータを、XML 文書等に埋め込む場合、以下のような記法が考えられる。本記法は JIS X 4166 “XML 文書へのグリフ識別子埋め込み”の記法を参考にしたものである<sup>20</sup>。

---

<sup>18</sup> SVG: Scalable Vector Graphics. W 3 C で標準化された図形表記のフォーマット  
(<http://www.w3.org/standards/techs/svg>)

<sup>19</sup> PNG: Portable Network Graphics. ISO/IEC 15948:2004 として規定されている。

<sup>20</sup> JIS X 4166 は、文字図形の参照を ISO/IEC 10036 に基づいて登録されたグリフ名を参照するものとし

XML 一般においては、

```
<glyph uri= “[文字図形に係る URI]” />
```

(ここで、「glyph」は、指定したグリフをこの場所へ表示することを示すタグとして別途定義するもの。タグ名を含め、別途、リファレンスとなる定義を規定する (7.1 参照)。)

一般の html ブラウザ向けの記述では、

```
<span glyph:uri= “[文字図形に係る URI]” >代</span>
```

(ここで「代」は、URI で指定された当該グリフを表示できない環境で代替えとして置き換えられる文字(列)。“glyph”は文字図形を指定するための属性として別途リファレンスとなる定義を規定する (7.1 参照)。)

ここで参照する文字図形が、文字情報基盤で整備された図形であれば、[文字図形に係る URI]の部分に

```
glyph.mojikiban.ipa.go.jp/<MJ 文字図形名>
```

のように記述することにより、指定可能である<sup>21</sup>。「MJ 文字図形名」は、文字情報基盤で整備された文字図形を一意に識別するために設けられた、MJxxxxxx のような、MJ+6 桁の数字から構成される識別名である。

文字情報基盤で整備した対象となっていない文字図形を指定する必要がある場合、SVG、PNG 等の標準規格で表現した図形を別途独自に用意し、その URI を指定することで図形を交換することが可能である。この場合、当該 URI は長期に渡って内容を維持管理する必要がある。また、図形情報をデータファイルの形で別途伝送した上で、そのファイル名を指定するといった手段も有る。この場合は、文字図形に係る情報は、受け手が管理する必要がある。

## 8.5. 外字の利用

UCS では、特定の範囲(私用面: Private Use Plane, PUP)の文字符号に、利用者が独自の文字図形を割り振って使うことが認められている。このようにして使用する文字を一般に、「外字」と呼ぶ。

「外字」の符号は、利用者が任意に与えるものなので、異なる利用者間で符号のみを交換すると、同一符号が両者で異なる文字図形に対応することとなり、いわゆる「文字化け」の原因となり、情報交換上の混乱を来す。そのため、外字の利用は特定組織の内部あるいは、密接に連携した組織間のみにとどめ、それら組織外への送信には用いるべきでない。

---

ているが、ここでは一般的 URI で参照するものとして記載した。

<sup>21</sup> 同 URI は近く IPA から公開予定である (7.1 参照)。