

文字情報基盤検討状況について

2013 年9月

独立行政法人情報処理推進機構

文字情報基盤の概要

IPA[®]

- 行政の実務で必要とされる約6万文字を提供
 - IPAmj明朝フォント : 約6万文字を国際標準に準拠して実装
 - 文字情報一覧表 : 他の体系の文字との対応関係などの情報を提供

文字情報基盤

邊邊辺 …

邊邊邊 …

戸籍統一文字 (漢字55270字)

戸籍のオンライン手続に使用することを目的として整理した文字 (辞書をベースに整理)

邊邊邊 …

邊邊邊 …

住民基本台帳ネットワーク

システム統一文字 (漢字19563字)

多くの住民が氏名に使う文字を整理

邊邊

JIS漢字コード (10050字)

実用上の情報交換の必要性から、出現頻度などを元に文字を選定 (JIS X 0213:2004)

辺

常用漢字 (2136字)

法令、公用文書、新聞、雑誌、放送など、一般の社会生活において、現代の国語を書き表す場合の漢字使用の目安を示す。
(個々人の表記にまで及ぼそうとするものではない。)

邊
邊
：

行政事務の効率向上、外字管理コストやシステム構築コストの削減を目的に、人名表記に使える共通的な漢字文字フォントと文字情報を整備した。

IPAmj 明朝フォント

- 無償で誰でも利用可能
- 文字の対象は、戸籍統一文字、住民基本台帳ネットワークシステム統一文字の58,813文字 (Ver.002.01)
- OpenTypeフォーマット
- 文字情報一覧表 (文字のコード間の関連や文字の持つ情報) 整備。

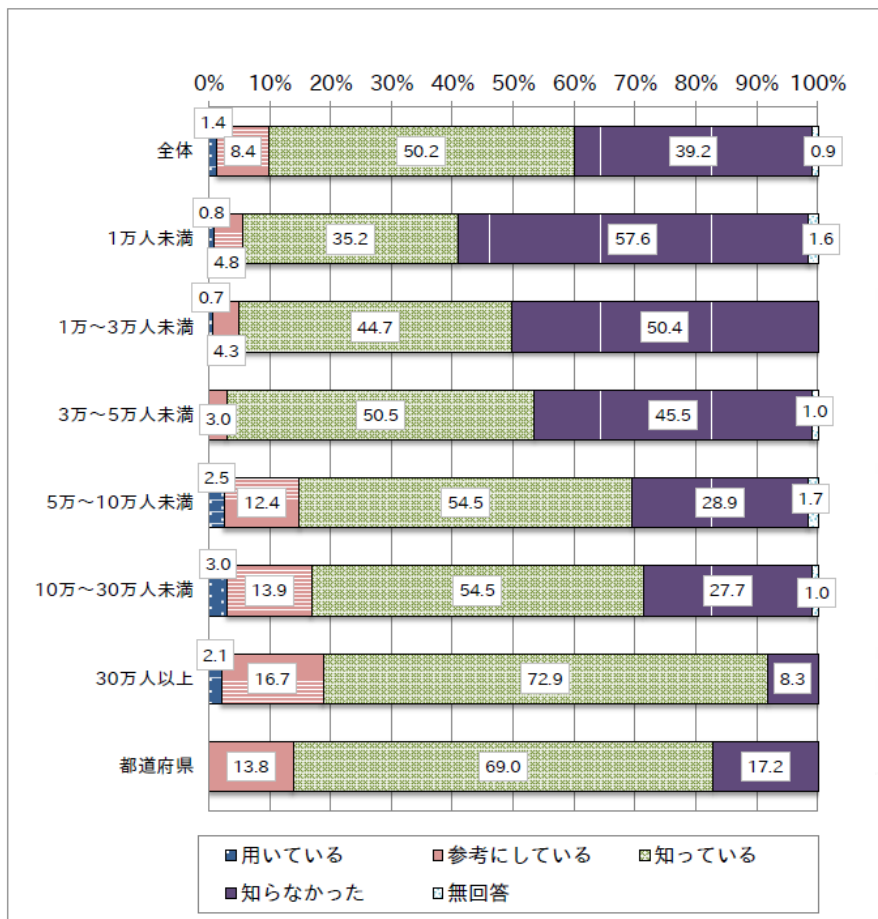
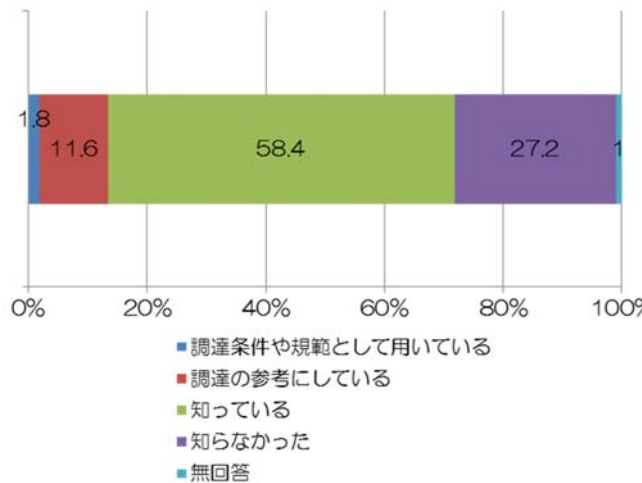


図 3.20 「文字情報基盤」の認知度（人口規模別、N=666）

IPAmj明朝フォント

累計4万件以上のダウンロード
(民間サイトからの配布を除く)

文字情報基盤の活用

H24実証実験の実施

	実証実験テーマ名（対象自治体／実施企業）	概要
1	多種文字コードと文字情報基盤文字コードとのコード変換ライブラリを介したデータ連携（石巻市／株式会社日立製作所）	異なる文字体系を持つ戸籍系システムと住基系システムの文字連携を実験
2	基幹系住民情報の汎用的連携（藤沢市／日本電気株式会社）	住基系システムと被災者支援システムの接続を実験
3	文字情報基盤を活用した自治体間での字形情報の交換（福岡県、北九州市、粕屋町／株式会社エヌ・ティ・ティ・データ）	異種システム、複数自治体間において自動車税等人名漢字を含む情報の交換を実験
4	文字情報基盤を適用したプリントアウトソーシング実証実験（南幌町／富士ゼロックスシステムサービス株式会社）	印刷を外部委託する際の情報連携を実験
5	住民情報系システム再構築における共通基盤導入評価（札幌市／ピースミール・テクノロジー株式会社）	文字情報基盤を活用した住民情報系システムの共通基盤構築の報告

(上記5については、IPAによる実証実験ではなく、札幌市で既に独自に行っていた事業についての評価を行った報告となっています。)

1,2は、異なるシステム間の連携のための文字コード変換に、MJ文字情報一覧表を活用したものです。1の石巻市では、戸籍システムと住基システムとの連携実験を行い、2の藤沢市では被災者支援システムと基幹系システムとの連携実験を行いました。文字コード変換の妥当性などについて、実際に自治体の現場の職員の方々に確認していただき、有効性を確認しました。

3の実験では、IPAmj明朝フォントとMJ文字情報一覧表を活用し、自動車税システムをはじめとする複数のシステムを用い、福岡県/北九州市/粕屋町での人名漢字情報を交換する連携実験を行い職員の方々にその有効性を確認していただきました。

4の実験は、IPAmj明朝フォントを活用することにより、自治体外部に印刷を委託する場合に、正確な人名漢字を伝達するための環境の構築を行ったものです。これを実際に宛名印刷に活用し、結果の有効性を、住民アンケート等を通じて確認しました。

5では、札幌市がその基幹システムの構築を、IPAmj明朝フォントを活用して再構築した経験を報告書として取りまとめたものです。札幌市では、システム更新にあたり、文字情報基盤を活用すること、大幅なコストダウンを目指しています。

- 文字コード体系、業者変更によるシステム移行（データコンバート）コスト、業者による囲い込み、不透明な費用積算が当面の課題である。（10万人以上30万人未満）
- 外字が国レベルで統一されることにより、ベンダの異なるシステム間の連携が円滑化するとと思われる。（30万人以上）
- 他団体間での文字を共通化することは難しい。（1万人未満）
- メーカー及びベンダがオープンな標準化を進められる環境が必要。また文字コード（外字）の標準化が実現すると、スムーズな共同化が図れるシステムが多いと感じている。（1万人未満）
- IPAの文字等外字に対する標準的な取組を行うことにより、外字にかかる連携費用等を削減することができる。（5万人以上10万人未満）
- 平成24年度は自治体実証実験によらず独自に普及活動を行っており、IVS実装検証をはじめ、クラウド型被災者支援システムにおけるベースフォントとしての活用や、文字情報基盤一覧表を活用による文字検索システムの実装やアンケート調査の宛名の文字にIPAMJの字形を実装するなど、実装においても積極的に活用を進めている。（30万人以上）
- 統一文字及び外字の統合について、システムやベンダ毎に違いがあることから検証に労力を使うこととなり、移行等に時間が必要となるため文字変換ツールの開発が急務である。これが実現すれば、かなりのコスト削減が見込まれる。また、システム連携ツールを作成することにより特別なインターフェイスが必要なくなるのではないだろうか。（3万人以上5万人未満） 100
- 文字情報基盤等の取組み・活用については、自治体内の固有事情（既存インフラやデータ等）への影響を最小限に留めつつ、情報技術や製品、国策等の変化に柔軟に対応するためには、自治体毎に様々な対策を講じることは避けられず、本団体においても、柔軟性を維持できる基盤整備に着手している。しかし、多くの自治体では、財政状況や技術力の問題、既存ベンダとの関係等により、TRM等の活用はおろか、最適なIT基盤の構築に向けた検討すら、着手できていない現状がある。（30万人以上）
- 今後、国及び地方自治体において最適な基盤を整備し、マイナンバー制度の有効活用を含めた住民サービス向上に資するためには、国及び自治体内で活用する情報技術（IT）や文字問題（外字等）を解決し、また、ベンダロックイン（特にSIベンダによるロックイン）を真に回避できる仕組みづくりが必要である。TRMをはじめ、文字情報等に関する経済産業省による規格化（JIS等）、総務省における自治体内での調達・運用指針の充実など、国及び地方自治体全体としての揺るぎないポリシーの整備が必須である。（30万人以上）

文字情報基盤データベースの概念設計

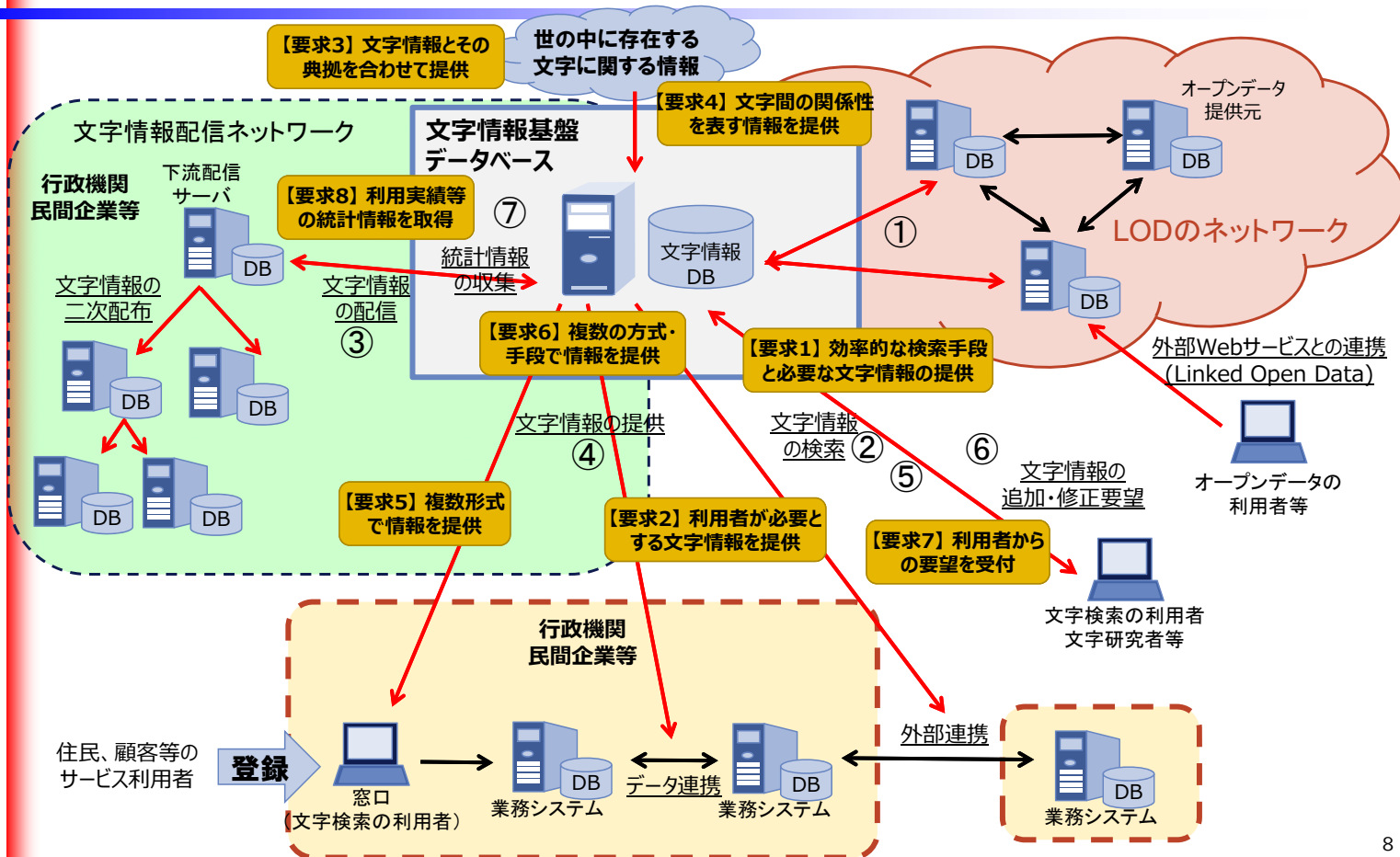
独立行政法人 情報処理推進機構

6

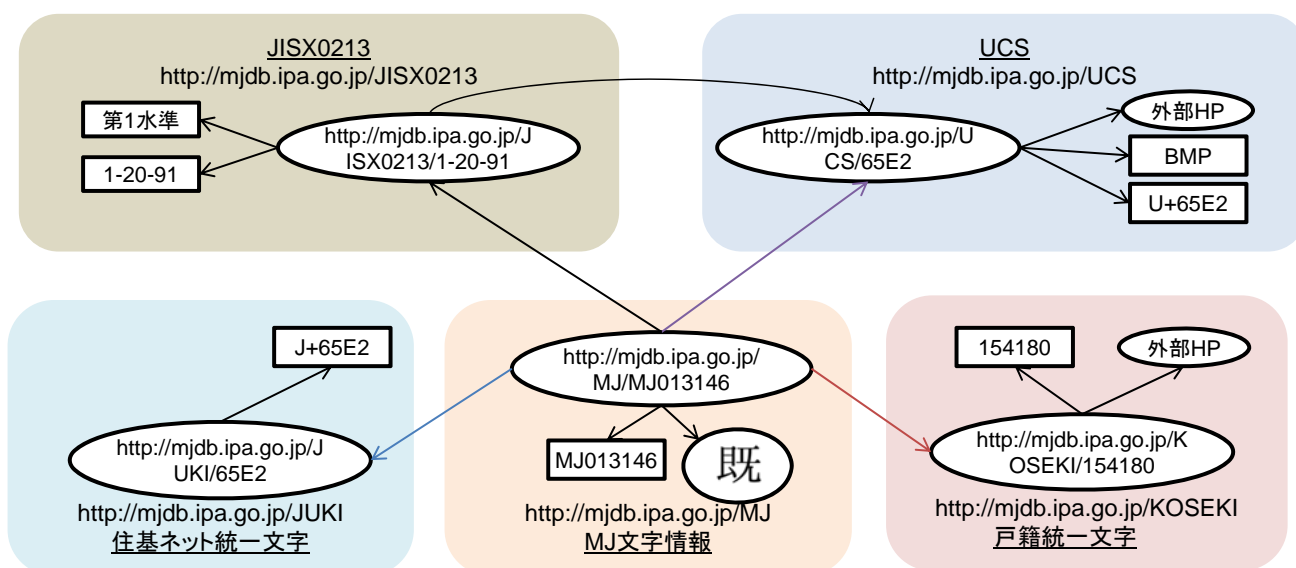
文字情報基盤データベース概念設計書 IPA[®]

- 文字情報基盤データベース
 - 文字情報基盤が提供する文字情報を充実させ、オープンデータとして提供することを目指す
- 概念設計の検討範囲
 - オープンデータの調査（技術動向、ライセンス）
 - 利用者ニーズとDBの要求要件
 - RDFやSPAQLでの文字情報の取扱い
 - データモデルの定義

概念設計書のURL: <http://mojikiban.ipa.go.jp/mojikibandb/>



RDFによる文字情報記述の検討



標準化の状況について

2013 年9月

独立行政法人情報処理推進機構

背景

- 文字情報基盤における国際標準化のゴール
 - 全てのMJ文字図形がISO/IEC 10646(UCS*)により一意に特定でき、情報交換可能とすること
- 新規符号化提案とIVD登録でゴールを目指す
 - 新規符号化提案（拡張F提案）

既存のUCS符号で扱うことができない文字については、ISO/IECへ新規符号化提案

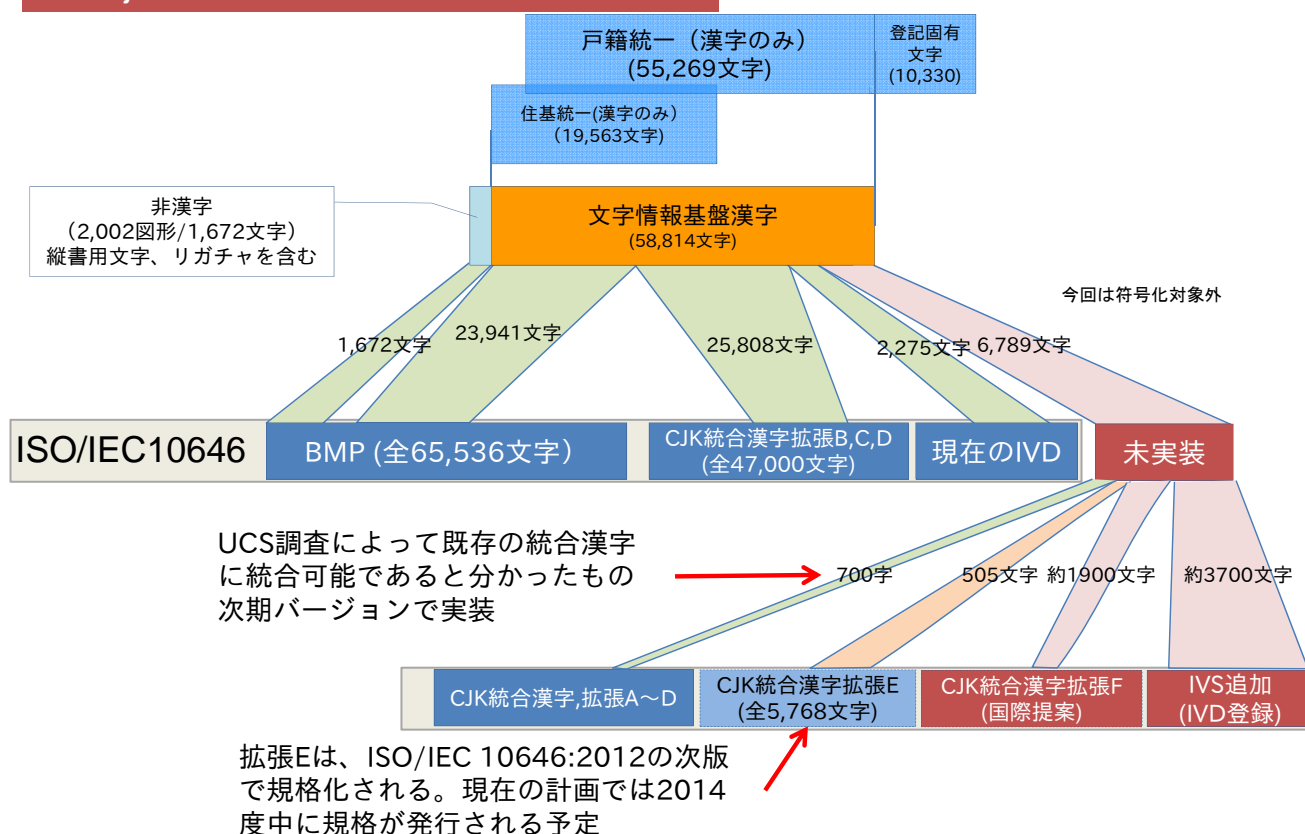
- IVD**登録

UCS符号に複数MJ文字図形が対応する場合(複数の字体が存在)字形選択子(VS)によって識別

*Universal Coded Character Set

**<http://www.unicode.org/ivd/>

IPAmj明朝フォントVer.002.01の符号化状況



12

拡張F提案の状況について

- 1,834文字を提案し受理された (2012年10月)
 - 電子行政での必要性を強調し、理解を得られた
 - 新規の提案に当たって提出が求められていた典拠資料として、電子政府用の文字情報データベースでも認められることとした。
- 拡張F分割 (2013年5月)
 - IRGは、拡張Fで提案された文字数が多い (8000文字) ため、先に作業するF1と、次に作業するF2に分割することに決めた。
 - 日本からは、総務省の「外字実態調査¹⁾」で実利用が確認された518文字をF1に入れるよう求め、承認された。

¹⁾ 市区町村の使用する外字の実態調査 (2012年3月)

- 文字情報基盤のIVDコレクション名
 - Moji-Johoコレクション(MJコレクション)と命名
- IVD登録に係る方針
 - 文字情報基盤が規範としている汎用電子の成果物から多くを継承
- 情報規格調査会SC2専門委員会への依頼
 - SC2専門委員会がIVD登録主体となる
 - 汎用電子成果物の継承に係る合意形成

14

IVD登録状況と今後の見通し

時期	内容
2013年7月 (完了)	第1期IVD登録候補の確定、SC2専門委員会への依頼と登録案の提出
2013年7月～9月 (完了)	SC2専門委員会による登録案のレビュー
2013年秋	Unicodeコンソーシアムによる公開レビューの開始
2014年上期	MJコレクション第1期の登録完了
2014年夏以降	MJコレクションに対応したIPAmj明朝フォントの公開

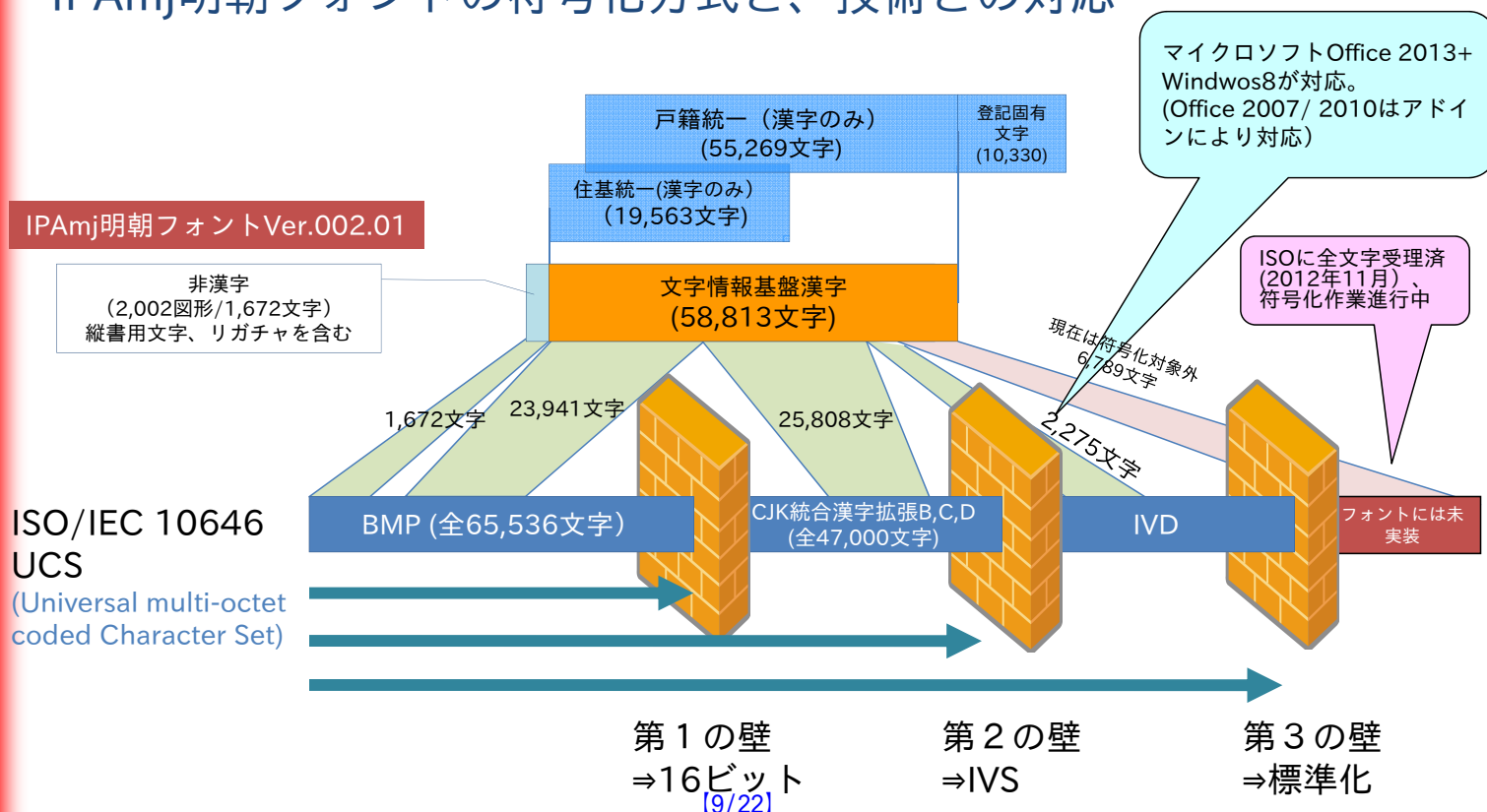
「文字情報基盤導入ガイド」の方針案

2013 年9月

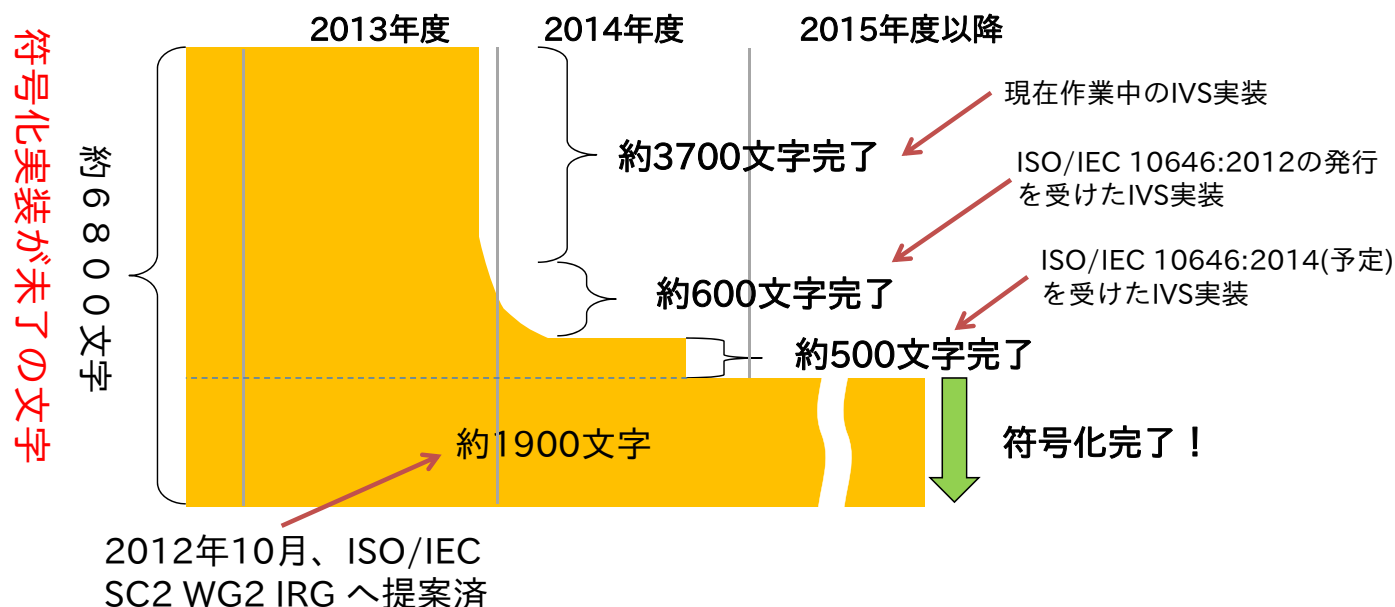
独立行政法人 情報処理推進機構

技術的背景(1)

IPAmj明朝フォントの符号化方式と、技術との対応



IPAmj明朝フォント実装計画



導入ガイドの目的

- 「閣議決定」の方針に従い、今後整備する情報システムに文字情報基盤を活用するにあたっての、典型的な具体的導入手順、運用法等を明らかにする。
- 今後、文字活用に係る調達仕様等を策定するにあたって、参考となる資料（技術モデル）とする。
- 今後、文字に係るシステムを運用するにあたって、参考となる資料（運用モデル）とする。

- 国際標準と矛盾しない技術的指針を示す
- 経済合理性を考慮した技術的指針を示す
 - 既存システムとの共存
 - 低い導入コスト
- 技術の進展、標準化の進展に沿った工程案を示す
- 文字そのものの運用ポリシーはスコープ外とする
 - 別途法律等で定められている方針に従う

20

文字情報交換の原則（案）

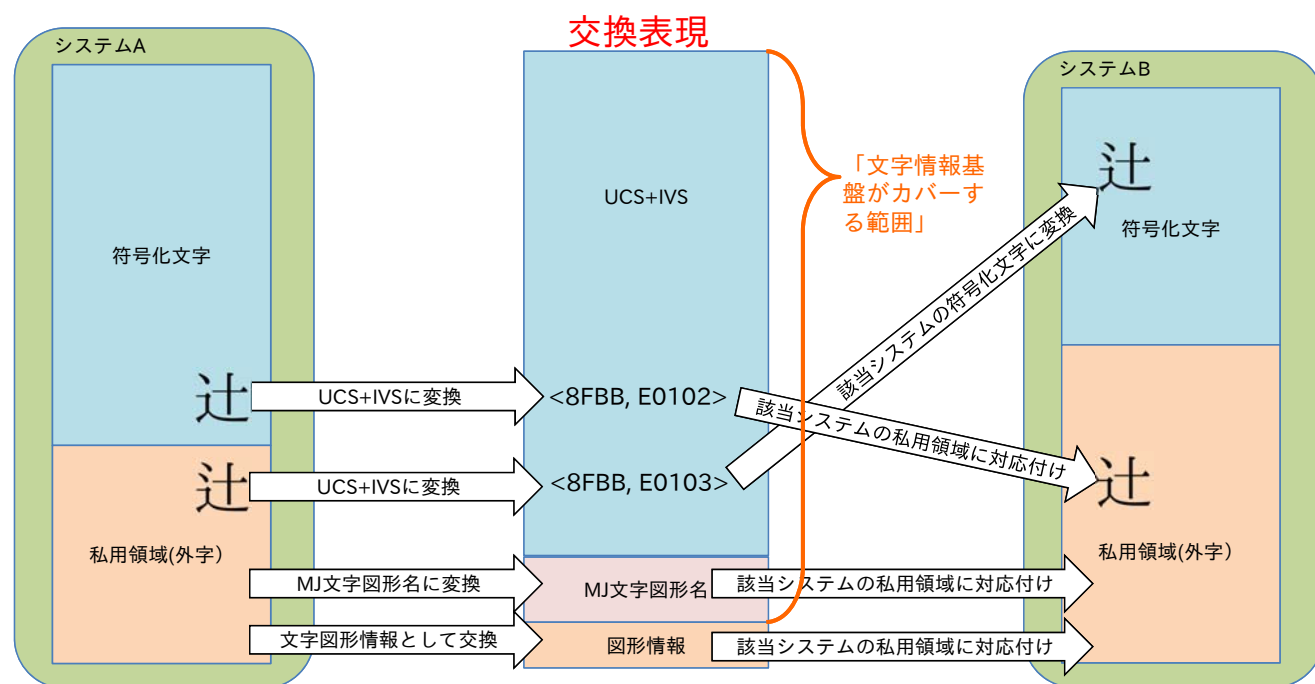
- 組織内部
 - 短期的：現状の文字体系を変えない
 - 中・長期的：IVSを使用可能なシステムの導入を図る
- 政府の組織間、自治体間
 - 短期的：IVS+MJ文字図形名+図形情報を交換できる統一的な「交換表現」の導入
 - 中・長期的：IVS+図形情報を交換できる統一的な「交換表現」の導入

- 組織外との情報交換（表示の場合）
 - 短期的：JIS X 0208+図形^(*)
 - 中期的：JIS X 0213（UCSによる符号化）+図形^(*)
（Webによる情報交換等の場合、サービスとして、MJ文字図形名をメタデータとして添付することを検討）
 - 長期的：IVS+図形^(*)（戸籍が手書きで記載されている等の場合に限られる）
- 組織外との情報交換（入力の場合）
 - 短期的：JIS X 0213に限定し、別途手動による入力手段を用意
 - 中・長期的：IVSを受け入れ、別途手動による入力手段を用意

(*) 国際標準に則った図形交換方式を採用する

22

「交換表現」のイメージ



◇ 符号化文字の集合としてはJIS X 0208, JIS X 0213, UCSなどが想定される。規格により例示字形が異なる場合があることに留意する。

◇ 既存システムの私用領域(外字) には、UCS、IVS、MJ文字図形に対応するものがあることに留意する。

◇ 複数のシステム間で文字情報を交換する際には、原則として文字情報基盤文字情報一覧表のMJ文字図形名もしくは対応するUCS符号位置及びIVSを用いる。

◇ MJ文字図形名を図形情報の固有名として用い、外部の文字図形データベースへの参照データとして用いてもよい。

◇ UCSから該当システムの符号化文字に変換する際、縮退対応（複数のUCSを単一の該当システム符号位置に対応させ、再度該当システムからUCSへの対応付けを行うと、双方向での情報の保全是保証されていないこと）に留意する。

「文字セットを定義する規格」の必要性について

2013 年9月

独立行政法人 情報処理推進機構

背景

コンピュータで使う文字を指定するためには

1. コーディングスキーム（ISO/IEC 10646 UCS等）
2. エンコーディングスキーム（UTF-8等）
3. 文字セット

の3つを指定する必要があるが、行政で使用する文字図形（グリフ）のセットを指定するにあたって3)に適切な規格が無く、そのため政府・自治体等での情報システム調達等で混乱を招いており、また、相互運用性に問題が生じている。

- 行政場面に応じたいいくつかの文字図形の集合
 - 常用漢字
 - 文字情報基盤文字
(住民基本台帳ネットワーク統一文字と戸籍文字との和集合)
 - その他（要検討）

を規定し、電子行政に係るシステム調達等において、文字に係る要件を客観的に正確に指定できるようにする。

26

期待される効果

- 情報システムの調達にあたり、文字に係る要件を、特定の実装を指定することなく、規格によって客観的に指定できる。
- 規格に沿ったフォントを自由に製品化できるため、フォントや関連情報製品/サービスに係る市場の活性化が期待される。

- 電子政府で用いる（べき／ことのできる）グリフ（文字図形）の集合を規定する。
- 新たな符号化は行わず、既存の規格：
 - 文字符号についてはISO/IEC 10646
 - IVSについては、IVD^(*1)を参照する。

(^{*1})<http://www.unicode.org/ivd/>

文字情報基盤成果物の バージョンアップについて

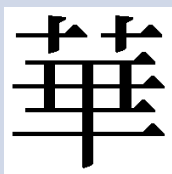

2013 年9月

独立行政法人情報処理推進機構

バージョンアップの概要

- MJ文字情報一覧表
 - 行政実務で用いられる情報を拡充、国際標準化状況を反映した表記に変更、使い勝手を向上
- IPAmj明朝フォントのUCS実装追加
 - 698文字図形に新たにUCS符号を実装
- 公開予定時期
 - 2013年10月上旬～中旬
- 次期バージョン名
 - IPAmj明朝フォントVer.002.02
 - MJ文字情報一覧表Ver.002.03

- 平成24年4月に追加された戸籍統一文字に対応し、IPA明朝デザインポリシーに基づき新たな文字図形（MJ068053）を作成

戸籍統一文字	MJ文字図形
	
552710	MJ068053

32

MJ文字情報一覧表の更新案(1/2)

- 行政実務で用いる情報を拡充、国際標準ISO/IEC 10646との対応情報をより分かりやすく変更（詳細は項目一覧案を参照）
 - 入管正字コード*の欄を新設
 - 戸籍実務で用いられる「新潮日本語漢字辞典」の検字番号欄を新設
- 「対応するUCS」、「UCS実装」の表記内容の変更、および「対応する互換漢字」欄の新設
 - 互換漢字に該当するMJ文字図形名については、「対応する互換漢字」欄にそのUCSコードを表記するとともに、「対応するUCS」欄に正規化後のUCSコードを表記
 - IPAmj明朝フォントには「UCS実装」欄のUCSコードを実装する。





*平成23年法務省告示第582号第二項第四号で規定されたコード等

- 「対応するUCS」情報の更新
2012年度～2013年度に実施した調査結果に基づき、「対応するUCS」欄の値を以下のように更新。
 - 新規にUCS値を記載するもの・・・353文字
 - 対応するUCS値の記載を変更するもの・・・118文字
(内 101文字は互換漢字の欄へ移行)
 - 対応するUCS値を削除するもの・・・12文字
(ISOで審議中(9文字は拡張F,3文字は拡張E)の文字へ同定)
- 「UCS実装」情報の更新
 - 698文字についてUCS符号位置を追加
 - 符号位置変更 2文字
 - 符号位置削除 2文字(新規に国際提案中)

34

符号化実装の削除を検討している文字

符号位置削除

MJ文字図形	現状
 MJ001865	 U+3C08 [4]
 MJ058862	 U+285C7 [0]

↑
規格表の当該符号位置に
示されている字形

↑
自治体における
外字作成実績数

- これらはISO 10646の「拡張F」へ新規の文字として提案中。
- 次期バージョン一覧表の「対応するUCS符号位置」欄からは削除し、フォントについては、「拡張F」が確定後に実装変更を行う。
- フォント実装変更に当たっては後方互換性を配慮した検討を行う。

符号位置変更

MJ文字図形	現状	変更先
 MJ000379	 U+35B4	 U+26BB6
 MJ006392	 U+4E54	 U+215E4
 MJ007306	 U+5188	 U+7F53

- ISOの統合規則に照らし、誤った割り当てであった。

- U+4E54の実装変更はしない
- U+215E4に対応した新たなMJ文字図形名を追加

- 入国管理局正字において、当該字形をU+7F53に対応付けられた
- MJ007306の対応をU+7F53に変更し、U+5188に対応付ける新たなMJ文字図形名を追加

規格表の当該符号位置に示されている字形

自治体における外字作成実績数

36

IPAmj明朝フォントについて(1/2)

- UCS実装について
 - MJ文字情報一覧表の「UCS実装」情報の更新に伴う、UCS実装の追加・修正（修正については、前項の考え方の通り）
- その他の修正・追加
 - 縦組み表示改善のためのフォントメトリクス調整
 - Ideographic Description Characters (U+2FF0 – U+2FFB) の追加実装

- UCS実装の優先順位^(*)の追加(赤字部分を追加)
 1. 常用漢字表に掲載されている字形
 2. 人名用漢字別表（戸籍法施行規則別表第二漢字の表）に掲載されている字形
 3. 住民基本台帳ネットワークシステム統一文字コードと対応するUCSが一致する字形
 4. JIS X 0213:2004に掲載されている例示字形
 5. UCSのJ欄（日本提案の字体）に掲載されている例示字形
 6. 法務省告示第582号別表1に掲載されている字形
 7. 康熙字典体
 8. 「市区町村が使用する外字の実態調査」の出現頻度

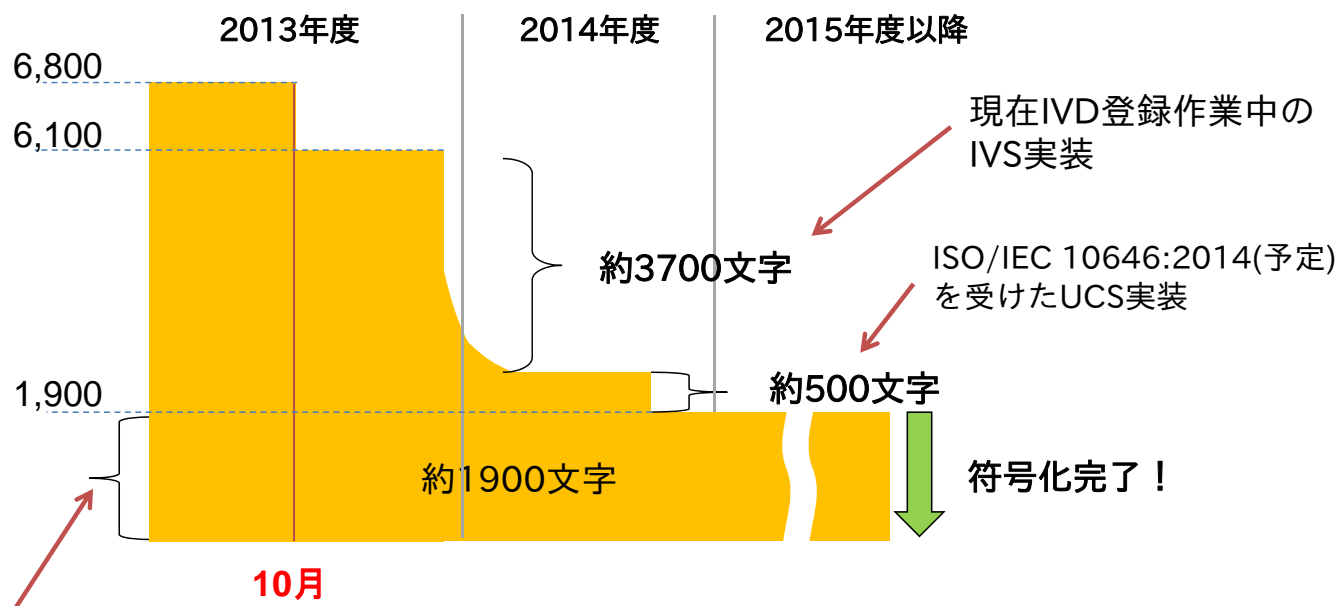
^(*)同一のUCSコードを持つ図形が複数ある場合において、フォント実装において対応付ける図形(デフォルトグリフ)の優先順位。

38

参考： Ideographic Description Characters(IDS)

2FF0	2FF1	2FF2	2FF3	2FF4	2FF5
2FF6	2FF7	2FF8	2FF9	2FFA	2FFB

符号化実装が未了の文字



2012年10月、ISO/IEC SC2 WG2 IRG へ提案済