IPA®
INFORMATION-TECHNOLOGY PROMOTION AGENCY, JAPAN

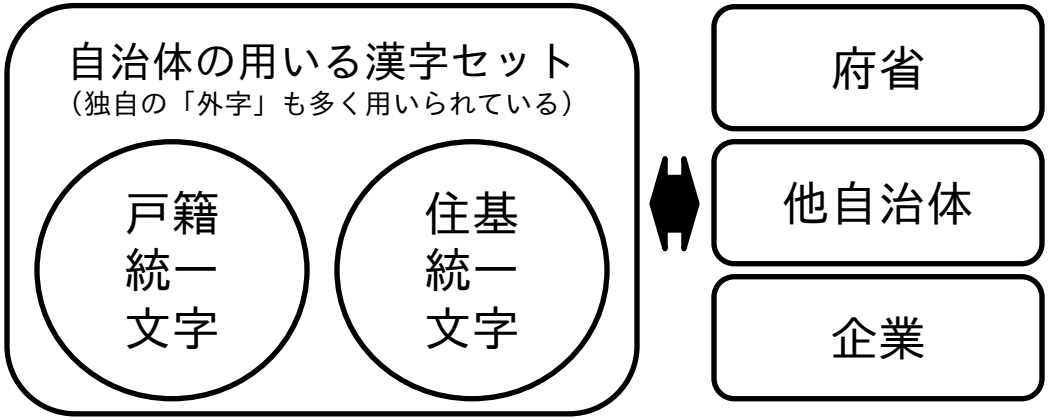
文字情報基盤構築事業の概要

2014 年1月

独立行政法人　情報処理推進機構

我が国で用いられる多様な文字セット

- 行政機関で用いる各種フォントセットは存在するが、利用制限が設定されていたり、広く活用することが困難。また、独自に作成した外字等、複数のコード体系が存在している。
  - 戸籍統一文字　56042字（内漢字55269字）
    - 戸籍業務に利用範囲を限定
  - 住基統一文字　21170字（内漢字19563字）
    - 本庁および支所、その他定める範囲
    - 市民病院や図書館では使用できない
    - 情報交換を目的とする



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

常用漢字　2136字

印刷用文字

表外漢字　1022字

人名用漢字（戸籍法施行令）

人名用漢字　2997字
常用漢字　2136字
漢字の表(一)　(常用漢字の異体字でないもの) 649字
漢字の表(二)　(常用漢字の異体字であるもの) 212字

情報交換用文字

ＪＩＳ　10038字
第一水準　2965字
第二水準　3388字
第三水準　1249字
第四水準　2436字

Unicode 3.1ではJIS X 0213が対応
日本語の文字コードで運用する必要性の高い文字。計11,233字

異体字の範囲と用途



多様性・アイデンティティ

交換性・意味

文字を巡る環境：文字情報利用への要望

- 住民
  - 個人のアイデンティティとして正確に表記すること
  - 行政コストを下げるべき
- 企業
  - 業務効率を落とさないこと
  - 行政機関の電子申請などの要求水準に対応できること
  - 現在一般に使われる以外の文字がむやみに広がるのは困る
- 行政機関
  - 住民を正確に確認できること（戸籍業務、証明業務等）
  - 行政効率を落とさないこと
    - 外字作成コストが減った代わりに、氏名の同定コストがかかったのでは意味がない
  - 誰もが扱える文字であること
  - すべての文字を使えるようにするのではなく、業務により文字の範囲を制限することも重要
- 情報転送を受ける行政機関
  - 正確な文字情報が伝わること

これまでの経緯（～2010年度）

経済産業省 Ministry of Economy, Trade and Industry

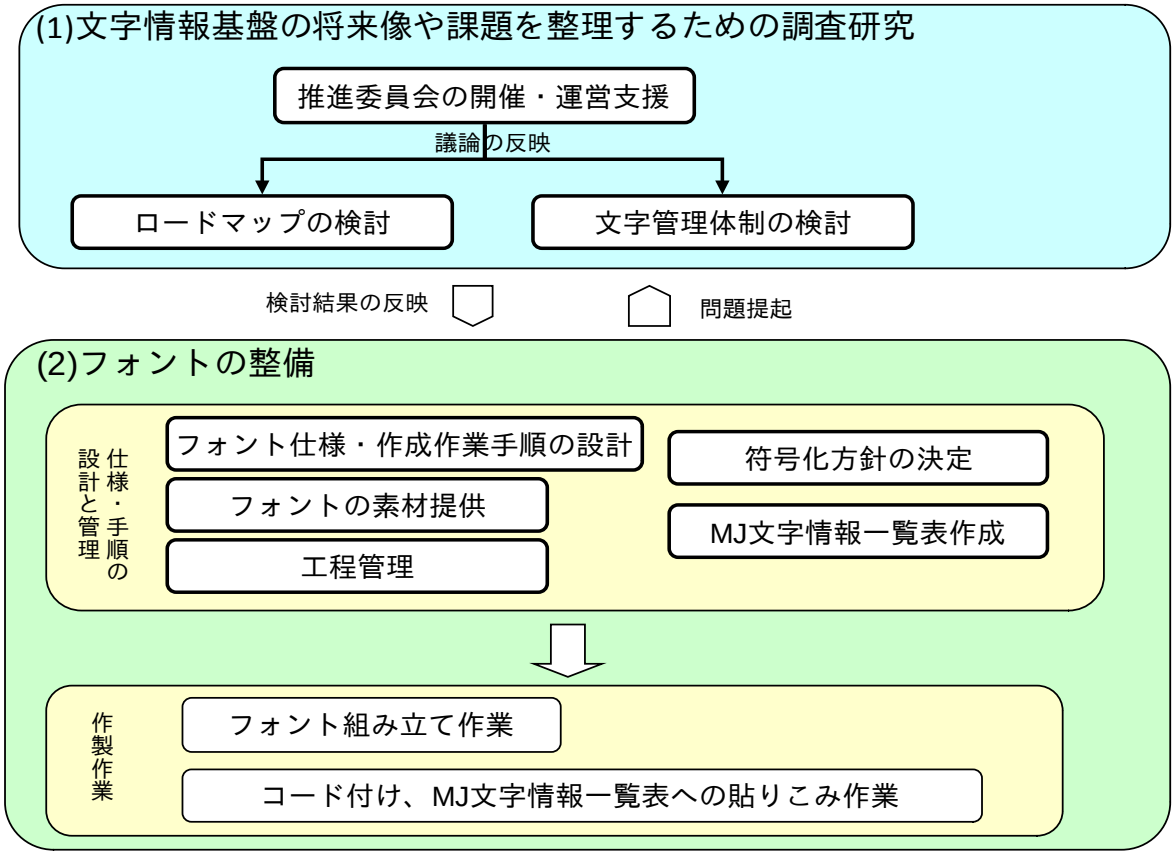
文字情報基盤

汎用電子情報交換環境整備プログラム

- 新たな情報通信技術戦略 工程表　平成22 年6 月22 日高度情報通信ネットワーク社会推進戦略本部
  - 文字も検討課題。
- 次世代電子行政サービス基盤等検討PT中間報告書平成21年12月21日
  - 「公共サービス連携基盤（仮称）」の中では、統一された文字コードで外字を扱えるようにするために、文字図形番号を用いる。また、文字の重複の排除や対象とする文字の範囲など、標準とする文字に関する基本的なルールを整理した上で、文字情報データベースを用意する。
- ＩＴ政策ロードマップ　平成２０年６月１１日ＩＴ戦略本部決定
  - 国・地方公共団体等の相互連携を推進するため、文字コードやデータ形式の標準化を推進するほか、諸外国における技術参照モデル（TR）の設定等も参考に、オープンな標準の採用を促進する。
- IT 政策パッケージ-2005　平成17 年2 月24 日IT 戦略本部決定
  - ③文字コード標準化の推進とデータベースの運用（経済産業省及び関係府省）
    - 住基ネット統一文字、戸籍統一文字の重複関係を整理、体系化したデータベースを2005 年末までに開発する。また、現在構築中の登記統一文字について、2006 年度末までに、このデータベースとの整合性を図ることとする。さらに、データベースの具体的な運用方法について関係府省で検討を行い、2005 年末までのできる限り早期に結論を得る。
- 電子政府構築計画2003 年　平成15年7月17 日各府省CIO連絡会議決定
  - （４）文字情報・コードの整備
    - 経済産業省及び関係府省は、行政情報化の共通基盤の一環として、将来的な国際標準との整合を視野に入れつつ、官民が汎用的に利用できる文字情報データベースの整備を進め、2005 年度末（平成17 年度末）までに構築する。
- e-Japan 戦 略 Ⅱ　平成１５年７月２日IT戦略本部決定
  - ⑨ 行政システム間をはじめ、民・官でのデジタル情報の自由な交換・共有のため、現在外字として利用されている文字について、2003 年度中に公開用文字情報データベースの試験的運用を開始し、2005 年度までに当該データベースを構築し、これを元に国際整合性も勘案した文字コード規格を整備する。
  - ③ アジアにおいて、コンテンツ流通のための権利処理・権利保護等の問題への対応、知的財産侵害品流通の防止、文字コードの標準化を進めると共に、放送・出版等で作成されたコンテンツや、美術館・博物館等の所蔵品のデジタル化及びアーカイブ化の推進、コンテンツの多言語化の推進により、アジア全域におけるコンテンツの積極利用を促進する。
- 行政情報化推進基本計画　平成６年12月25日閣議決定
  - 情報システムにおいて使用している外字について，ＪＩＳ第３水準及び第４水準の制定を待ってその解消を図るとともに，この場合においても残る可能性のある外字について，交換のルールを策定する。

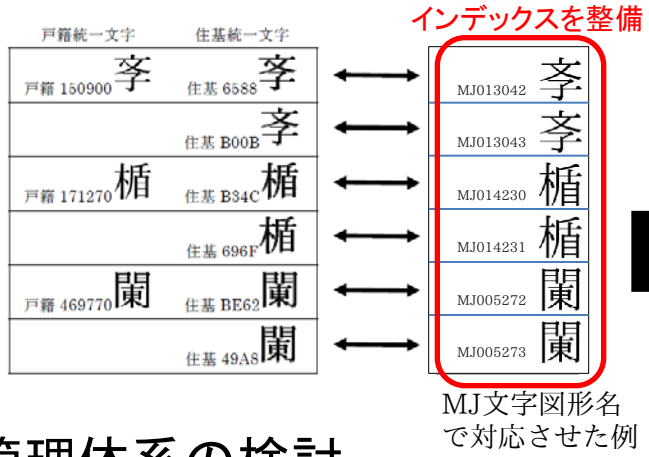
文字情報基盤構築事業（2010年度）

- 平成２２年度電子経済産業省推進費（文字情報基盤構築に関する研究開発事業）をIPAが受託
  - 課題の整理
  - フォントの整備



事業の概要

- 文字を巡る課題と解決案の整理（ロードマップの整備）
- 電子行政を推進するためのフォントと文字情報一覧表の整備
  - 行政情報の適正で効率的な管理や手続きを行うために統一された文字コードに基づくフォントの整備や利活用のための文字情報基盤を整備。行政機関のみならず、この基盤を民間との連携も視野に入れることで社会的コストを大幅に削減可能。



府省、自治体内の他のシステムの文字体系との連携も可能

電子申請で、正しい氏名が送信できる

他システムに送信した文字が、相手のシステムでもどの文字か正確にわかる

システム更新時に、文字変換が容易になる

フォントをダウンロード可能にする事で、新規外字作成コストが削減される

- 文字管理体系の検討

82A6

芦

芦 芦 芦 芦

1 2 3 4

芦田さんを一括して検索

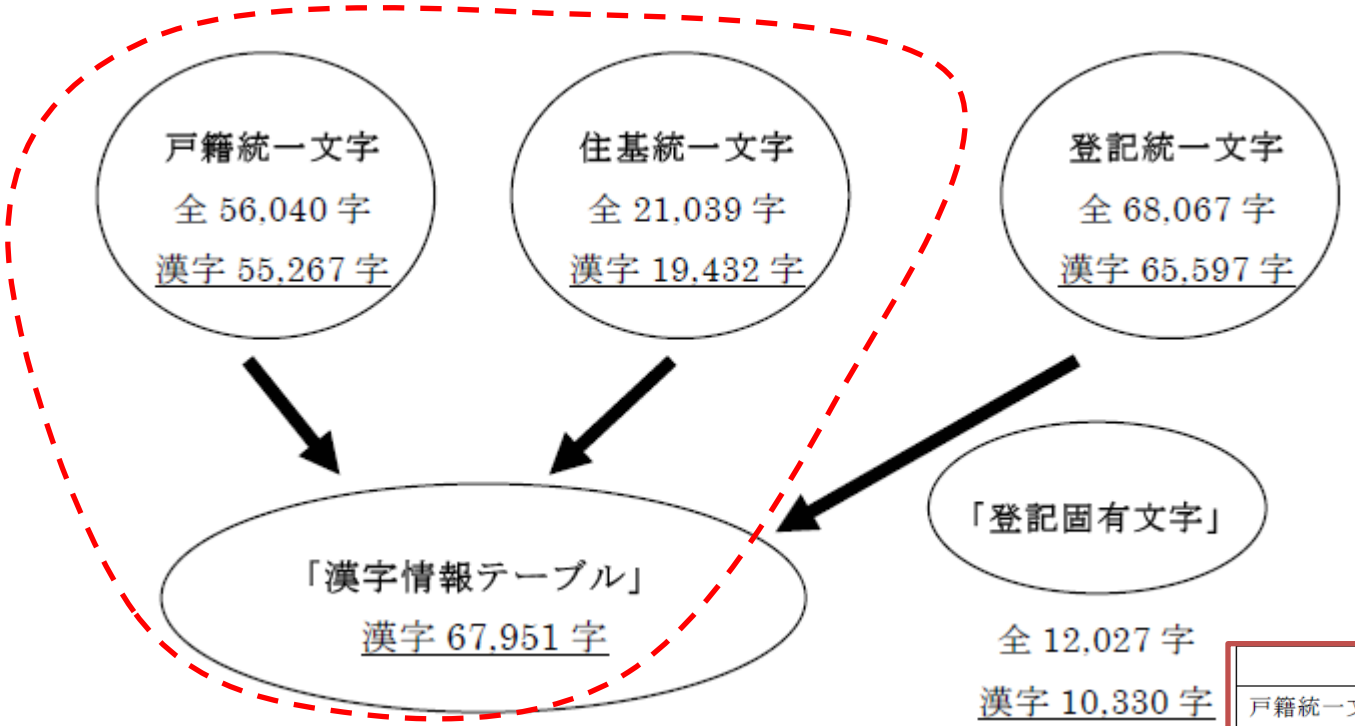
異体字による氏名が一括して検索可能になる

名寄せが簡単になる

文字の対象

■ 文字の対象は、汎用電子情報交換環境整備プログラムでの整理を元に、戸籍統一文字、住基統一文字の58,712文字とする。

- 登記統一文字だけに存在する文字は含んでいない。
- 変体仮名は現時点では含んでいない。



＋非漢字

パターン	文字数
戸籍統一文字・住基統一文字・登記統一文字に三者共通	15,986 字
戸籍統一文字・登記統一文字に二者共通	39,281 字
住基統一文字・登記統一文字に二者共通	1,092 字
住基統一文字のみ	2,354 字
登記統一文字のみ	9,238 字
計	67,951 字

本事業の趣旨と対象範囲について

■ 本事業では、行政機関として情報処理をするために必要となる文字情報基盤を整備。本事業の実施に当たっては、民間フォント・ビジネスとの棲み分けという観点から、国として取り組む必要性の高い国民生活に必要な最低限の文字について、基盤としてインデックスを整備するもの。

- 戸籍、住民票に必要な文字を対象
- 主要サイズの明朝体のみを整備
- 文字の美しさは公用文として必要十分なレベル
- インデックスで調べた後の各機関が外字を作成するのは大変であるので、そのフォントも提供

■ 近年の経済社会における活動では、電子的手段による情報交換が行われる機会が増加傾向にあり、個人から企業・行政機関に、企業等から行政機関等に情報転送が行われる場面も多い。このため、行政機関のみによる独自の文字体系を持つだけでは経済社会全体としての効率について十分な改善が見込めないことから、本事業の成果物については、個人・企業で活用をさまたげない。

■ 基準となる文字セットと検索環境がインデックスとして整備されることにより、国内の情報流通を促進し、それに伴いフォント・ビジネス及び周辺ビジネスの活性化が期待できる。

文字情報基盤推進委員会(2010年度)

経済産業省 Ministry of Economy, Trade and Industry

文字情報基盤の整備にあたり、国内の関係機関が参加する委員会を構成

委員長　林　史典　聖徳大学・聖徳大学短期大学部人文学部長
大学院言語文科研究科長人文学部日本文化学科長教授

委員参加組織

財団法人日本規格協会
社団法人情報サービス産業協会
社団法人コンピュータソフトウェア協会
社団法人電子情報技術産業協会
財団法人日本情報処理開発協会
大学共同利用機関法人人間文化研究機構国立国語研究所
社団法人日本印刷産業連合会
社団法人日本印刷技術協会
社団法人情報処理学会情報規格調査会SC2
CIO補佐官技術タスクフォース
電子政府推進員
川口市
内閣官房
総務省
法務省
文化庁
経済産業省

オブザーバ参加組織

国会図書館
警察庁
法務省
国税庁
厚生労働省
独立行政法人国立公文書館
日本年金機構
社団法人日本情報システム・ユーザー協会

導入に向けて　：　ポリシーの策定

経済産業省 Ministry of Economy, Trade and Industry

■ 全ての業務で全ての文字を使うと住民や行政の利便性を損なうこともあります。そのため業務目的に応じて適切なポリシーを整備する必要があります。

例
- 戸籍業務
  - 戸籍統一文字
  - 誤字、俗字は法務省の方針に従う
- 住民票業務
  - 住基統一文字＋戸籍統一文字
- 卒業証書等の重要名証明書類等
- その他業務（庁内で閉じている業務）
  - カタカナでの手続も含めて考える
- その他業務（オンライン申請業務等外部連携がある業務）
  - 民間ではJIS第4水準までしか使っていないことを考慮
  - 現時点ではJIS第2水準までしか対応できていない企業も多いことに留意
- その他業務（法定受託業務など他機関との連携がある業務）
  - 外部連携先のポリシーを確認する



■ ポリシー間の連携を考える
- 戸籍業務では正確な記述をして、その他の業務ではIVS代表文字を利用するなどのルールを検討

■ 同定基準
- 基本は常用漢字の同定基準。

委員会（2011年度～2012年度）

文字情報基盤推進委員会

委員長：　須藤 修　東京大学大学院 情報学環　教授

委員

- 文化審議会国語分科会会長 / 聖徳大学
- ISO/IEC JTC1/SC2国際議長 /
  長岡技術科学大学
- 川口市
- 文化庁
- 内閣官房 情報通信技術(IT)担当室
- 内閣官房 社会保障改革担当室
- 総務省 行政管理局
- 総務省 自治行政局
- 法務省 民事局
- 経済産業省　CIO補佐官
- 文字情報検討WG主査
- 運用検討WG主査
- 技術検討WG主査

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で、大幅なコストダウンを目指しています。

世界最先端IT 国家創造宣言（平成25年6月14日閣議決定）

経済産業省 Ministry of Economy, Trade and Industry

Ⅰ. 基本理念

Ⅱ. 目指すべき社会・姿

Ⅲ.目指すべき社会・姿を実現するための取組

1．革新的な新産業・新サービスの創出と全産業の成長を促進する社会の実現

2．健康で安心して快適に生活できる、世界一安全で災害に強い社会

3．公共サービスがワンストップで誰でもどこでもいつでも受けられる社会の実現

（1）利便性の高い電子行政サービスの提供

（2）国・地方を通じた行政情報システムの改革

（3）政府における IT ガバナンスの強化

Ⅳ. 利活用の裾野拡大を推進するための基盤の強化

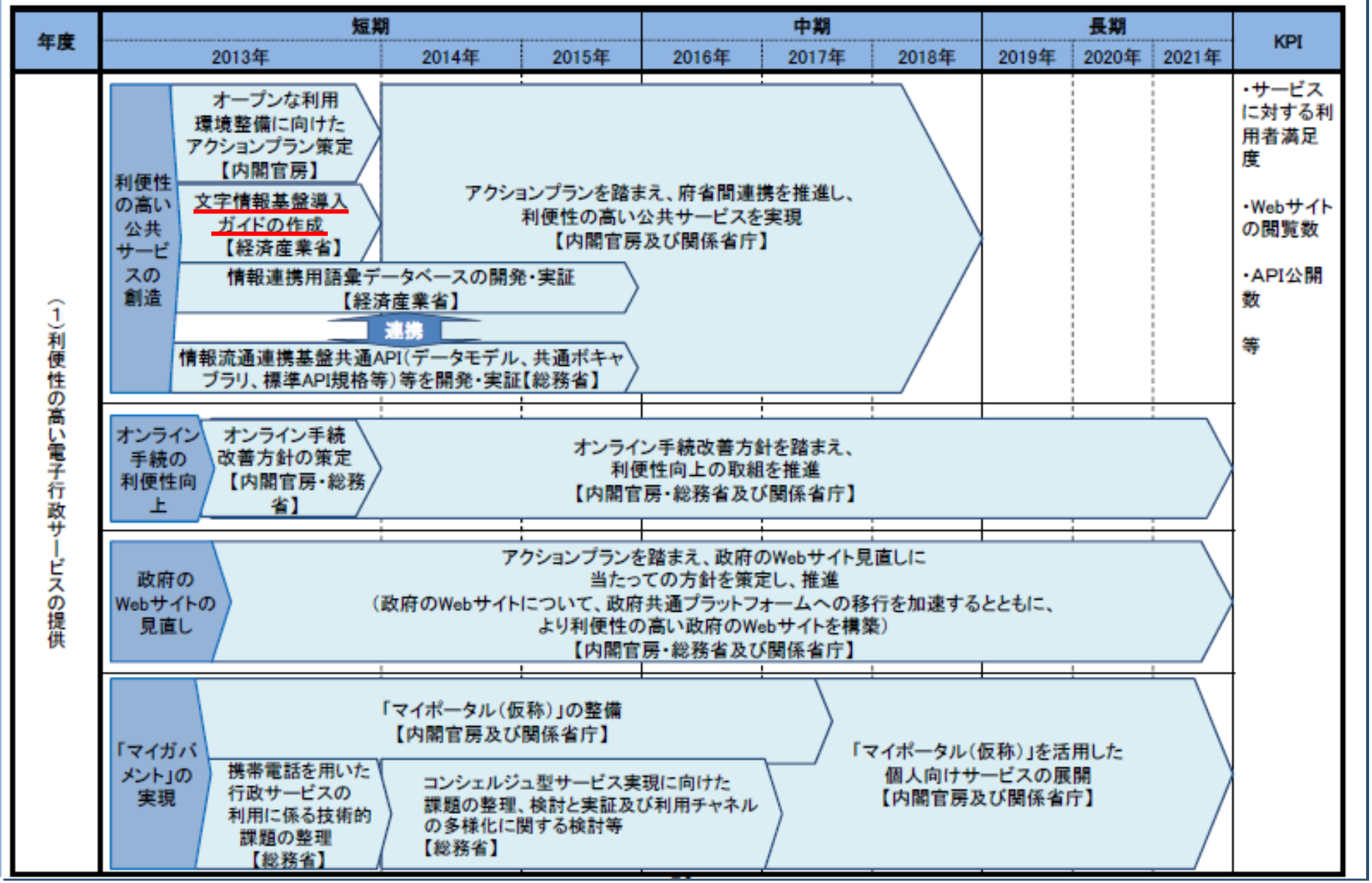
Ⅴ. 戦略の推進体制・推進方策

従来政府が担っていたサービスの提供機能を民間にも開放し、官民の協働によって、より利便性の高い公共サービスを創造する。国民がステークホルダーとして積極的に参加できるよう、このためのクラウドを活用したオープンな利用環境を、**データ・フォーマット、用語、コード、文字等の標準化・共通化**、また、アプリケーション・インターフェイス（API）の公開等を行いつつ整備する。**特に文字に関しては、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。**

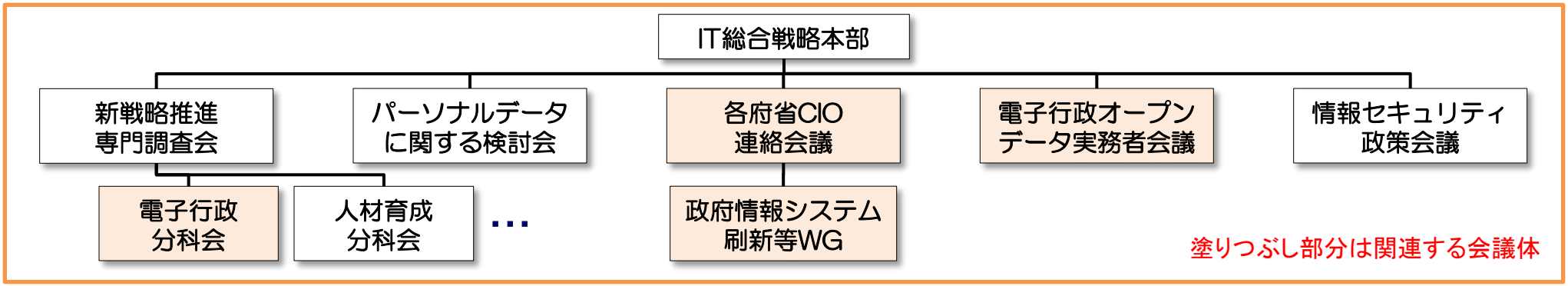
世界最先端IT 国家創造宣言工程表（電子政府）

経済産業省 Ministry of Economy, Trade and Industry

実施スケジュール （3. 公共サービスがワンストップで誰でもどこでもいつでも受けられる社会の実現）



検討体制の構築（2013年度～）



協力依頼　　報告

情報共有基盤推進委員会

委員長　須藤修（東京大学情報学環長）
有識者　武田英明（共通語彙基盤WG座長）、林史典（文字情報基盤WG座長）
関係団体　JISA（ソフト業界）、JEITA（ハード業界）
オブザーバ　内閣官房（IT総合戦略室、社会保障改革担当室）、総務省（行政管理局）

- 各委員会、WGの設置主体は経済産業省、事務局はIPA
- 各WGの下には、具体的な検討を行なうSWGを設置。

共通語彙基盤WG

＜実施内容＞
- 情報連携用語彙基盤DBの構築
- APIカタログの整備
- 情報連携用ツールの整備

【主査】NII　武田先生
【関係省庁】（課長級）
IT室、社保室、内閣府防災、総務省、国交省、国会図書館

文字情報基盤WG

＜実施内容＞
- 導入ガイドの作成
- 文字情報DBの構築
- フォント整備
- 標準化の実施

＜メンバー＞
右表のとおり

	氏名	位置付け	所属
座長	林 史典	有識者	聖徳大学教授
委員	三上 喜貴	有識者	長岡技術科学大学教授(副学長)
委員	氏原基余司	有識者	文化庁
委員	大山 水帆	自治体	川口市
委員	荻野　敦	関係団体	LASDEC
委員	【課長級】	関係省庁	総務省自治行政局　住民制度課
委員	【課長級】	関係省庁	総務省総合通信基盤局 高度通信網振興課
委員	【課長級】	関係省庁	法務省 民事局
オブザーバ	【課長級】	関係省庁(利用)	内閣官房IT室
オブザーバ	【課長級】	関係省庁(利用)	内閣官房 社保室
オブザーバ	【課長級】	関係省庁(利用)	総務省行政管理局 情報システム管理室
委員	-	SWG座長(文字情報)	
委員	-	SWG座長(運用)	
委員	-	SWG座長(技術)	

公開中の成果物

- IPAmj明朝フォント（http://mojikiban.ipa.go.jp/）
  - 戸籍統一文字の漢字(55,270文字)
  - 住民基本台帳ネットワークシステム統一文字の漢字(19,563文字)

  を包含する約6万文字を収録したフォント
  - 国際標準に則って符号化（符号化未完の約6千文字についても継続的に符号化作業を続行中）。
- 文字情報基盤 文字情報一覧表（MJ文字情報一覧表）
  - 文字図形、各種文字コード、読み等の各種情報を一覧できる
  - 文字の「見える化」を図る
  - 文字図形データをSVGフォーマットで提供
- 報告書等
  - 平成２２年度調査事業の成果報告書
    （ロードマップや利用ガイド（案）等を収録）
  - 平成24年度実証実験成果報告書
    （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）
  - 関連委員会議事概要

公開物のライセンス

- IPAmj明朝フォント
  - IPAフォントライセンス
    - 無償利用、複製可能。変更可能。ただし、変更を加えたフォントを配布した場合、その利用者が当該フォントをオリジナルへ戻せる手段を設ける必要
- 文字情報基盤 文字情報一覧表（MJ文字情報一覧表）
  - クリエイティブ・コモンズ・ライセンス 2.1「表示、継承」
- SVGフォーマットによる文字図形データ
  - クリエイティブ・コモンズ（表示・継承）

文字情報一覧表の項目一覧（1/3）

- 文字コードの対応関係その他のメタデータを収録 -

項目	内容
MJ文字図形名	MJ文字情報一覧表内で一意に定めた文字図形の名称
戸籍統一文字番号	6桁の10進数の戸籍統一文字番号を記した。
住基ネット統一文字コード	「J+」に続けて4桁の16進数の住民基本台帳ネットワークシステム統一文字コードを記した。
入管正字コード	平成23年法務省告示第582号第二項第四号 正字に示された「工業標準化法（昭和二十四年法律第百八十五号）に基づく日本工業規格（以下「日本工業規格」という。）X〇二二一号（以下「国際符号化文字集合」という。）に規定する漢字（国際符号化文字集合附属書ＪＡ日本文字部分レパートリに該当するものに限る。）及び別表第一に定める漢字をいう」の16進数を記した。
入管外字コード	平成23年法務省告示第582号第二項第7号に示された「国際符号化文字集合に規定する私用文字として別表第一に定める文字の符号化表現」の16進数を記した。
漢字施策	「常用漢字」は、常用漢字表（平成22年11月30日内閣告示）に示された通用字体（常用漢字表で括弧が添えられていない漢字）に示された漢字である。「人名用漢字」は、戸籍法施行規則別表第二の「漢字の表一・二」に示された漢字である。
X0213	X0213欄には、JIS X 0213:2004の面区点位置を記した。
X0213 包摂連番	包摂連番欄には、該当する文字にJIS X 0213:2004の包摂規準を適用させると符号化できる場合、包摂規準連番を記した。
X0213 包摂区分	包摂区分欄には、「0」「2」の2種を記した。「0」は、該当する文字がJIS X 0213:2004規格票の例示する字体に相当する字 形であることを示す。「2」は、該当する文字がJIS X 0213:2004の包摂規準を適用できる字体に相当する字形であることを示す。
X0212	X0212欄には、JIS X 0212:1990の区点位置を記した。
対応するUCS	対応するUCS欄には、「U+」に続けてISO/IEC 10646:2012の16進数のUCSコードを記した。ただし、UCSコード情報の確度や対応状況が異なるため、「UCS対応カテゴリー」欄に確度情報を記した。

文字情報一覧表の項目一覧（2/3）

- 文字コードの対応関係その他のメタデータを収録 -

項目	内容
対応する互換漢字	互換漢字に該当するMJ文字図形名には、対応する互換漢字欄に、「U+」に続けてISO/IEC 10646:2012の16進数のUCSコードを記した。
実装したUCS	IPAmj明朝フォントVer.002.02の実装において当該図形に対して割り当てるUCSコードを記した。 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）は、下記の優先順位により決定した。 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
UCS対応カテゴリー	「対応するUCS」欄の確度のカテゴリー分けを下記(A1,A2,A3,A4,E,F)のように行った。 【カテゴリーA】 平成23年度から平成25年度までの文字情報基盤整備事業で確認済みのもの。 • A1) ISO/IEC10646 CJK統合漢字にJソースがあるもの（JIS X 0213、JIS X 0212に例示されている文字を含む）。 • A2) 住基ネット統一文字で住基ネット統一文字コードが3400～9FA5のもの。 • A3) Unicode ConsortiumのIVD Version 2010-11-14のHanyo-Denshiコレクションとして登録されているもの。 • A4) 文字情報基盤整備事業において個別に確認したもの。 【カテゴリーE】 ISO/IEC 10646:2014(予定)の発行により、対応するUCSコードの確定が見込まれるもの。 【カテゴリーF】 現在ISOに提案中であり、対応するUCSコードが存在しないもの。

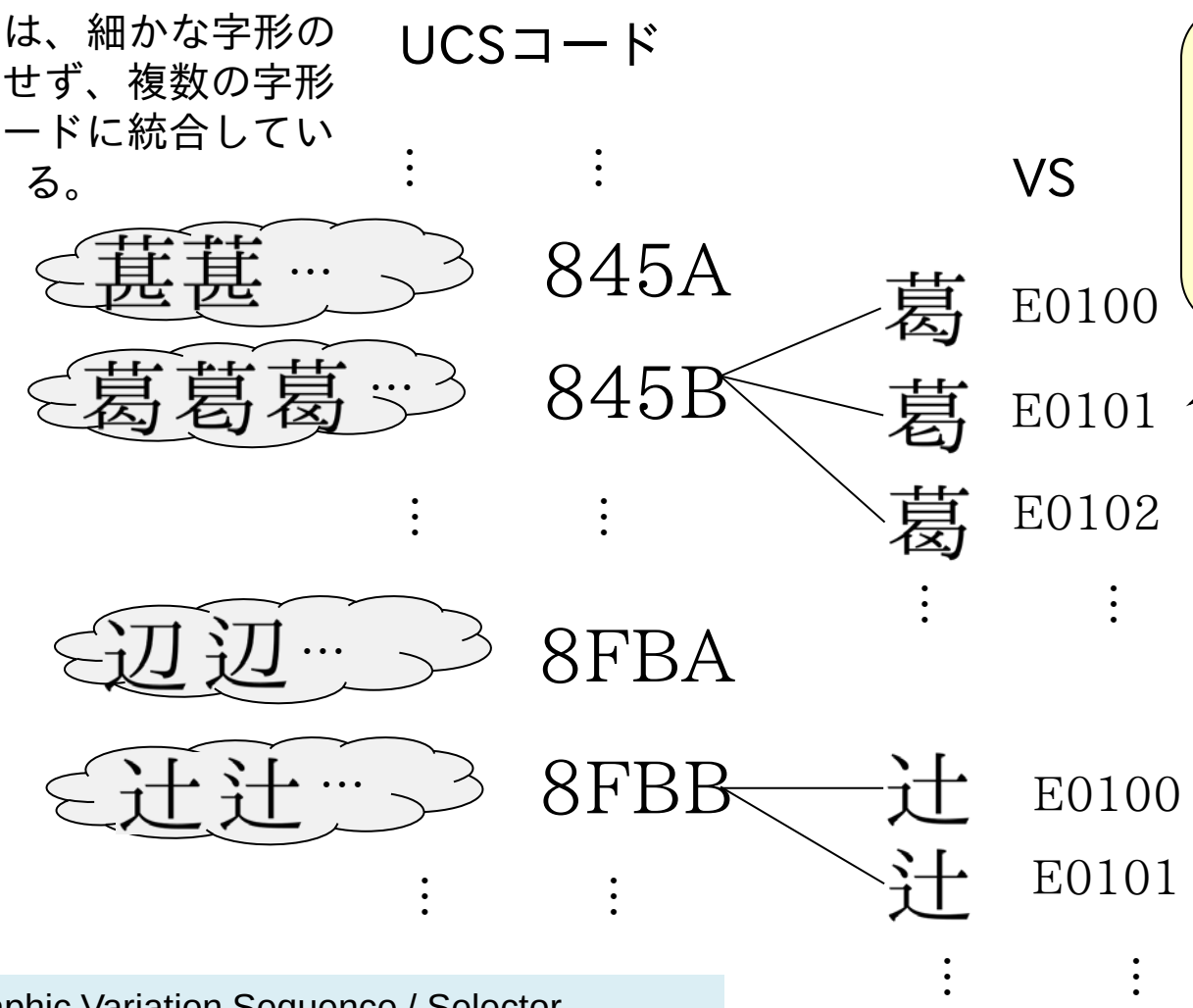
文字情報一覧表の項目一覧（3/3）

- 文字コードの対応関係その他のメタデータを収録 -

項目	内容
実装したIVS	Unicode ConsortiumのIVD Version 2010-11-14のHanyo-Denshiコレクションにより実装したIVSを記した。
MJ文字図形バージョン	IPAmj明朝フォントVer.002.02およびSVG図形ファイルに収録した文字図形（MJ文字図形）のバージョンを記した。
登記統一文字番号(参考)	8桁の10進数の登記統一文字番号を記した。
部首(参考)	部首欄には部首番号を記した。『康熙字典』の部首の通し番号により、1から214までである。
内画数(参考)	内画数欄には部首内画数（部首を除いた残りの部分の画数）を記した。
総画数(参考)	総画数欄には文字の総画数を記した。
読み(参考)	読み欄には文字の読みを記した。原則として、音読みは片仮名、訓読みは平仮名である。
大漢和	大漢和欄には、諸橋轍次『大漢和辞典』（修訂第二版第六刷、大修館書店、2001年）及び鎌田正・米山寅太郎『大漢和辞典補巻』（初版、大修館書店、2000年）の親字番号を記した。なお、『大漢和辞典』及び『大漢和辞典補巻』の親文字と漢字字体に差異があるものの同定が可能なものには、親字番号の末尾に＃を付けて示した。
日本語漢字辞典	日本語漢字辞典欄には、新潮社編「新潮日本語漢字辞典」（第四刷、新潮社、2008年）の検字番号を記した。
汎用電子整理番号(参考)	汎用電子情報交換環境整備プログラムの漢字情報テーブルにおけるレコードの通し番号を記した。
平成明朝(参考)	平成明朝体グリフのグリフ名。平成明朝体グリフは汎用電子情報交換環境整備プログラムの文字グリフ作業委員会にて制作されたものである。
更新履歴	本一覧表において、過去に変更された情報の履歴を記し、変更のあった事項毎にMJ文字情報一覧表バージョン番号を行頭に付した。

(参考) IVS / IVD の考え方

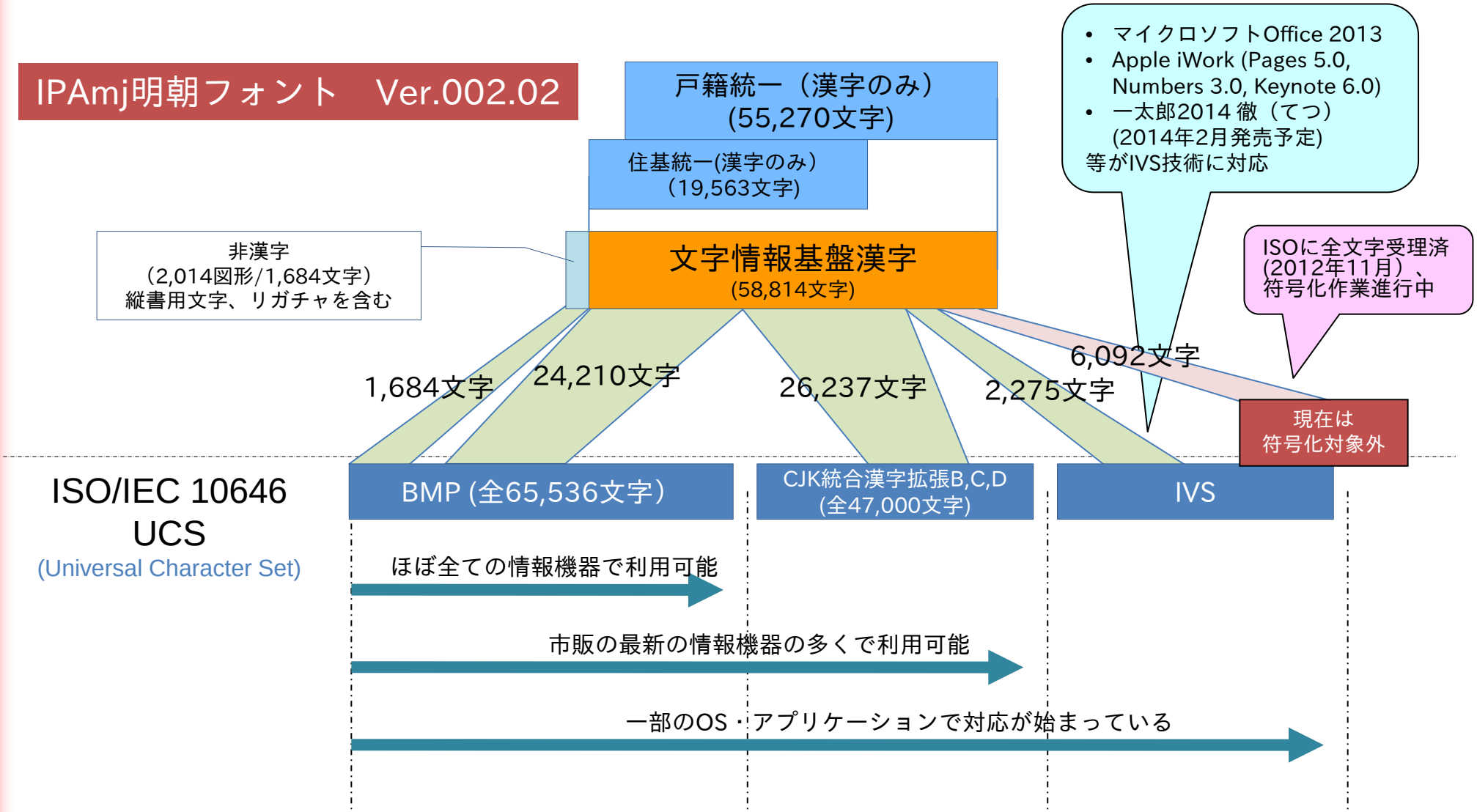
UCSコードは、細かな字形の差異は区別せず、複数の字形を一つのコードに統合している。



UCSコードと字形を指定するVS（字形選択子）を並べて記述することで、字形を指定することができる。
※VSの値は実際の規格と異なります。

IVS: Ideographic Variation Sequence / Selector
IVD: Ideographic Variation Database

IPAmj明朝フォントの符号化と活用技術



IPAmj明朝フォント実装計画 (2013年12月現在)

