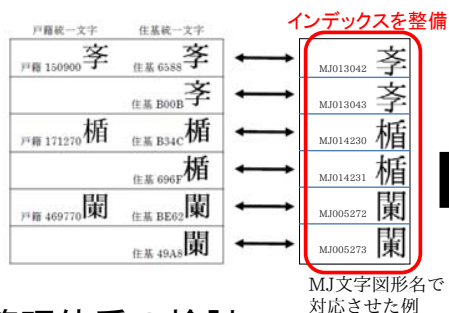


# 文字情報基盤構築事業の概要

2011年6月16日

## 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が正確にわかる

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

芦田さんを一括して検索

芦 芦 芦 芦

1 2 3 4

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

## 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。

2

## 文字を巡る環境

3

## 関係者から見た文字情報利用の要望

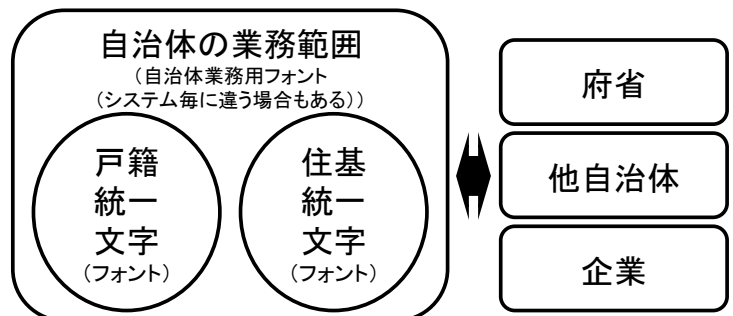
### ■ 多くの関係者の要望を整理すると以下の通りである。

- 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
- 文字ビジネス関連企業
  - ・ 民業圧迫になるのは困る

## 公的主要フォントと一般に使われる文字

### ■ 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である。そのため、行政機関内に複数のコード体系が存在している。

- 戸籍統一文字 **56040字**
  - ・ 戸籍業務に利用範囲を限定
- 住基統一文字 **21039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 **1945字** (新常用漢字試案 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 **2930字**

常用漢字 1945字  
漢字の表(一) (常用漢字の異体字でないもの) 776字  
漢字の表(二) (常用漢字の異体字であるもの) 209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

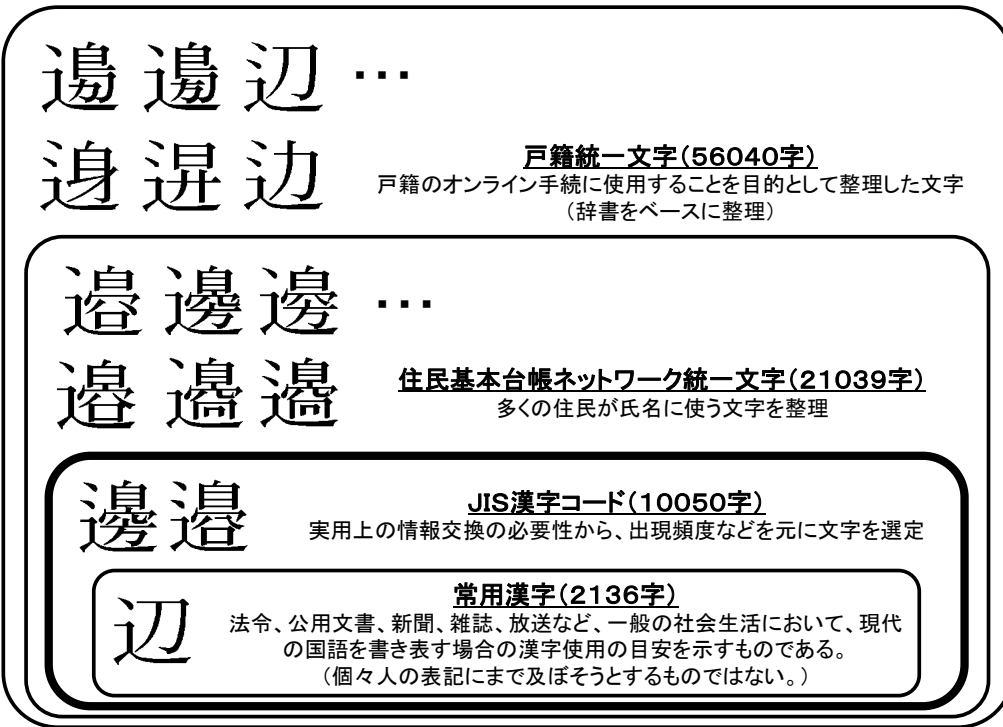
日本漢字能力検定

1級 **約6000字**  
準1級 約3000字  
2級 約1945字(他人名用漢字)

上記が読み書き可能

- 本事業で、4つの文字セットがシームレスに連携できるようになる。
- それぞれの意義・経緯を踏まえて利用方法を考える必要がある

多様性・アイデンティティ



交換性・意味

## 外字関係者数の推定と組み合わせ例

- 外字氏名の方がどのくらいいるかの正確な数字は計測されていない。
- 氏名表記情報が正確に公開されている衆議院議員名簿を元にするると、国民の5%が氏名内で外字を使用していると推定される。

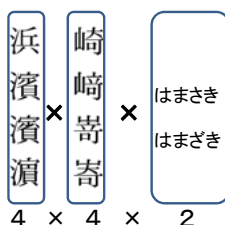


氏名	漢字	平仮名
井上 信治	井上 信治	いの上 しのぢ
石毛 謙子	石毛 謙子	いしげ けんこ
大石 謙徳	大石 謙徳	おおいし けんとく
菅 直人	菅 直人	すが なおひと
下妻 みつ	下妻 みつ	しもむら みつ
高木 義	高木 義	たかぎ ぎ
高松 和夫	高松 和夫	たかまつ わくお
高色 敏	高色 敏	たかいろ みのる
竹本 直一	竹本 直一	たけもと なおひと
田中 和徳	田中 和徳	たなか わくとく
田中 富紀子	田中 富紀子	たなか とみのりこ
徳田 義	徳田 義	とくた ぎ
中川 秀彦	中川 秀彦	なかつかわ ひでおひこ
飯島 吉郎	飯島 吉郎	いひじま よしろう
嶋山 河夫	嶋山 河夫	しまやま かわお
伊藤 耕輔	伊藤 耕輔	いとう こうすけ
町村 信孝	町村 信孝	まちむら のぶたか
松本 康	松本 康	まつもと やすひこ
村田 吉隆	村田 吉隆	むらた よしひろ
森山 幸	森山 幸	もりやま さち
矢嶋 幸二	矢嶋 幸二	やじま さちぢ
吉井 英徳	吉井 英徳	よしい ひでとく
和田 隆志	和田 隆志	わたり たかし

衆議院議員480人中  
正確な表記として画像の文字を公開している議員 24名(5%)  
そのうち名字に画像の文字を使用している議員 9名(2%)  
(参議院で同様に検証すると、氏名5%と名字3%)  
(2009年10月20日)

国民の5%が氏名の表記に関する課題を持っていると推定される  
また、2%の人が名字の表記に関する課題を持っていると推定される

注: 常用漢字表前書きにある「字体の違いではなくデザインの違い」と言う視点で各文字を見ると数は少なくなる



浜崎という名前は、「浜」で4通り、「崎」で4通りの文字があり、更に読み方が2種類あるので、32通りの組み合わせが考えられる。

同じ漢字氏名にも多くのパターンがあり、検索をすることが困難である

## インターネット申請の事例

- 実際の電子申請システムでは、利便性やシステムの接合性などを考慮して、文字の制限を実施している(外字がないことも一因)
  - e-Gov電子申請システム利用規約(使用可能な文字)
    - 第10条 本システムにおいて使用可能な文字は次の各号に掲げる文字とする。
    - 一 JIS X 0201として規格化されている英数字及び記号を含む1バイト文字。
    - 二 JIS X 0208として規格化されている2バイト文字。
    - 三 JIS第一水準漢字及びJIS第二水準漢字。
  - e-Taxで利用できる文字コードは、JIS X 0221をUTF8で符号化した以下の文字コードが使用できます。
    - 「基本ラテン(JIS X 0201と互換性のあるもの。ただし、文字タブ(0009)、改行(000A)及び復帰(000D)以外の制御文字(0000~001F、007Fを除く)」、「平仮名」、「片仮名」、「CJK統合漢字」、「CJK互換漢字」、「CJK用の記号及び分音記号」、「半角形・全角形(ただし、半角カナ(FF66~FF9F)を除く)」、「ラテン補助(ただし、制御文字(0080~009F)を除く)」、「矢印」、「一般句読点」、「罫線素片」、「幾何学模様」、「基本ギリシヤ」、「キリール」、「数字記号」、「数字の形」、「囲み英数字」、「囲みCJK文字/月」 「CJK互換文字」
    - なお、e-Taxで使用できない文字については、利用者の判断により、代替文字を選択していただくことになります。
  - 警視庁 警察官・職員採用 インターネット申込みFAQ
    - Q3-1 氏名が常用漢字ではないのですが、外字入力できますか？
    - A3-1 外字は入力できませんので、申込み時は正しい氏名に近い漢字またはひらがなで入力してください。第一次試験当日には受験票の氏名を赤字で修正したものを持参してください。

※CJK統合漢字(20902字):日本中国・台湾韓国の各漢字の間で統合化操作が行われたもの

8

## これまでの経緯

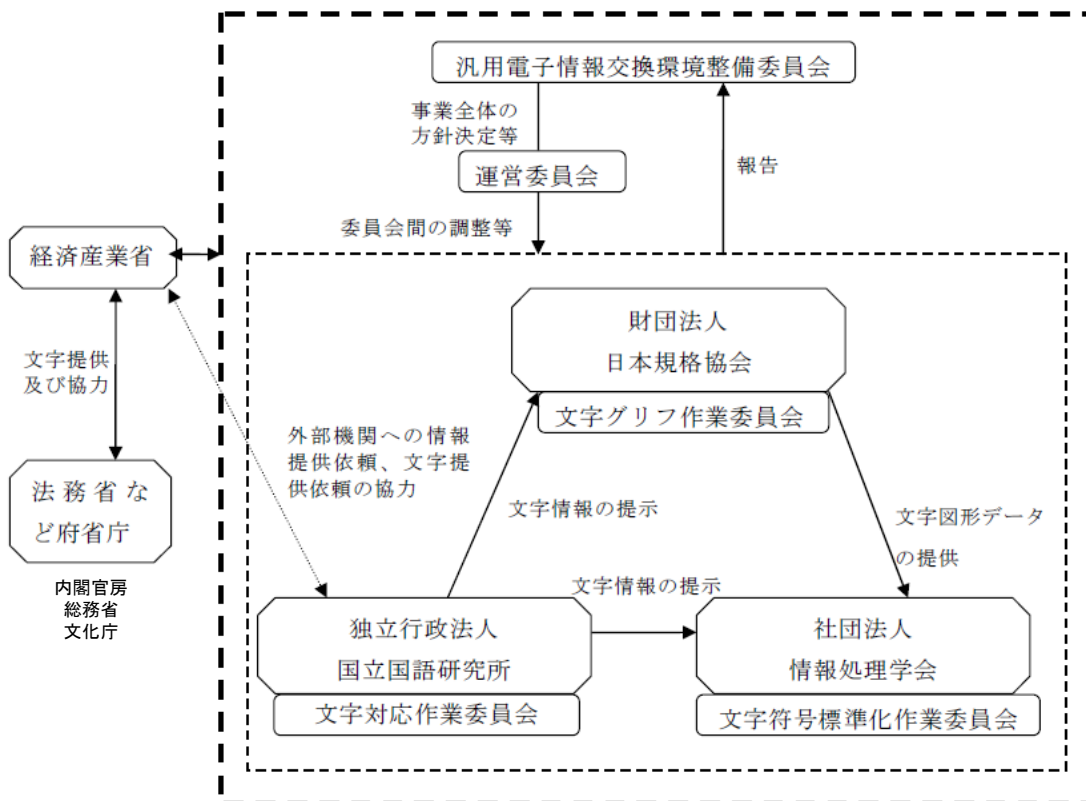
9

■ これまで多くの計画や戦略で目標が定められてきたが、文字の多さ、同定の困難さ、権利関係の複雑さ等から実現は大幅に遅れた。

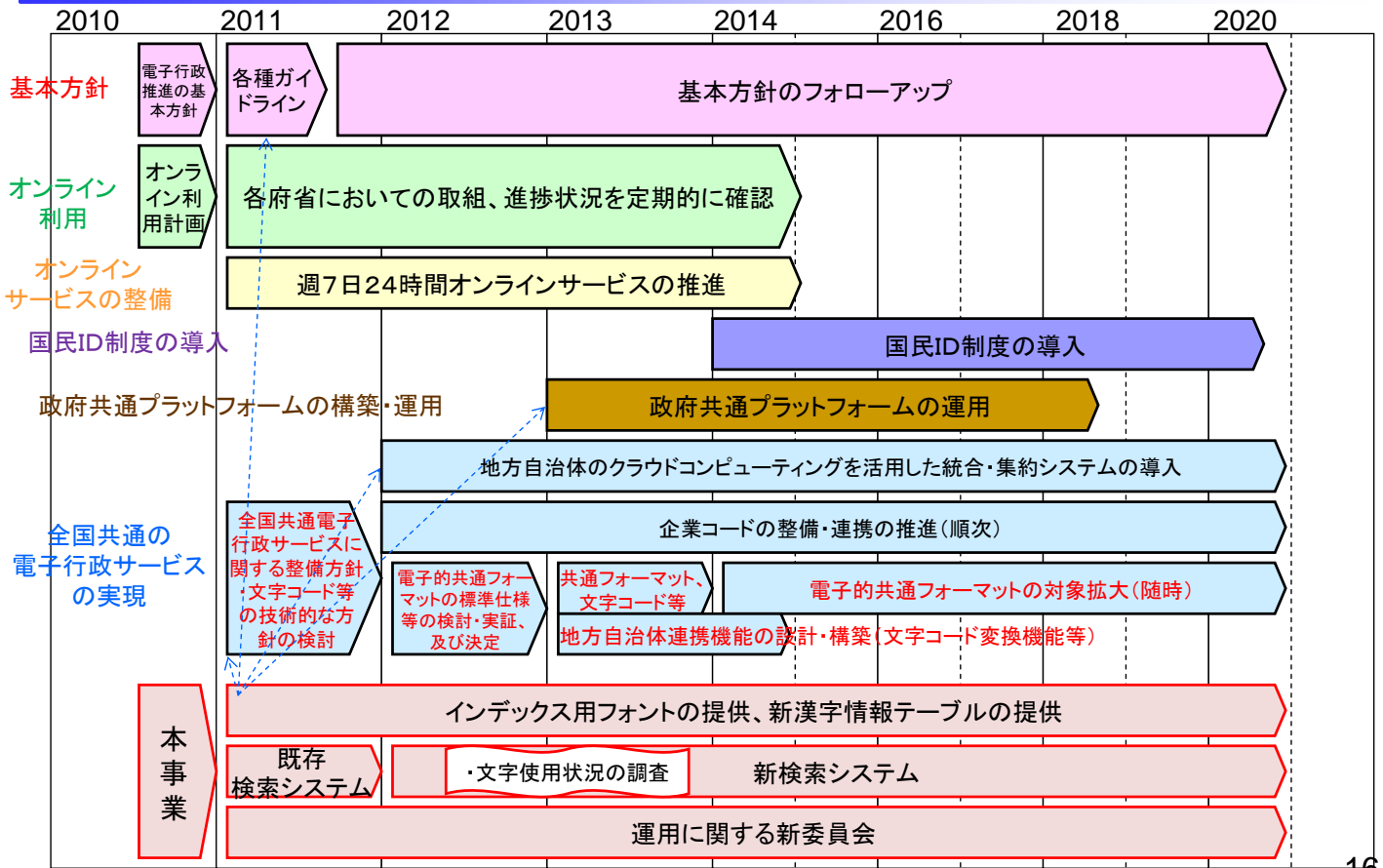
- 新たな情報通信技術戦略 工程表 平成22年6月22日 高度情報通信ネットワーク社会推進戦略本部
  - ・ 文字も検討課題。
- 次世代電子行政サービス基盤等検討PT中間報告書 平成21年12月21日
  - ・ 「公共サービス連携基盤(仮称)」の中では、統一された文字コードで外字を扱えるようにするために、文字図形番号を用いる。また、文字の重複の排除や対象とする文字の範囲など、標準とする文字に関する基本的なルールを整理した上で、文字情報データベースを用意する。
- IT政策ロードマップ 平成20年6月11日 IT戦略本部決定
  - ・ 国・地方公共団体等の相互連携を推進するため、文字コー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(TRM)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水準の制定を待つ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。

## 汎用電子情報交換環境整備プログラム(H14-20)

■ 5府省連携で本格的にコードの整理を実施した。



# 新たな情報通信技術戦略 工程表(文字関連主要事項)



16

## 文字情報基盤推進委員会

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  
(2010年9月—2011年3月)

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

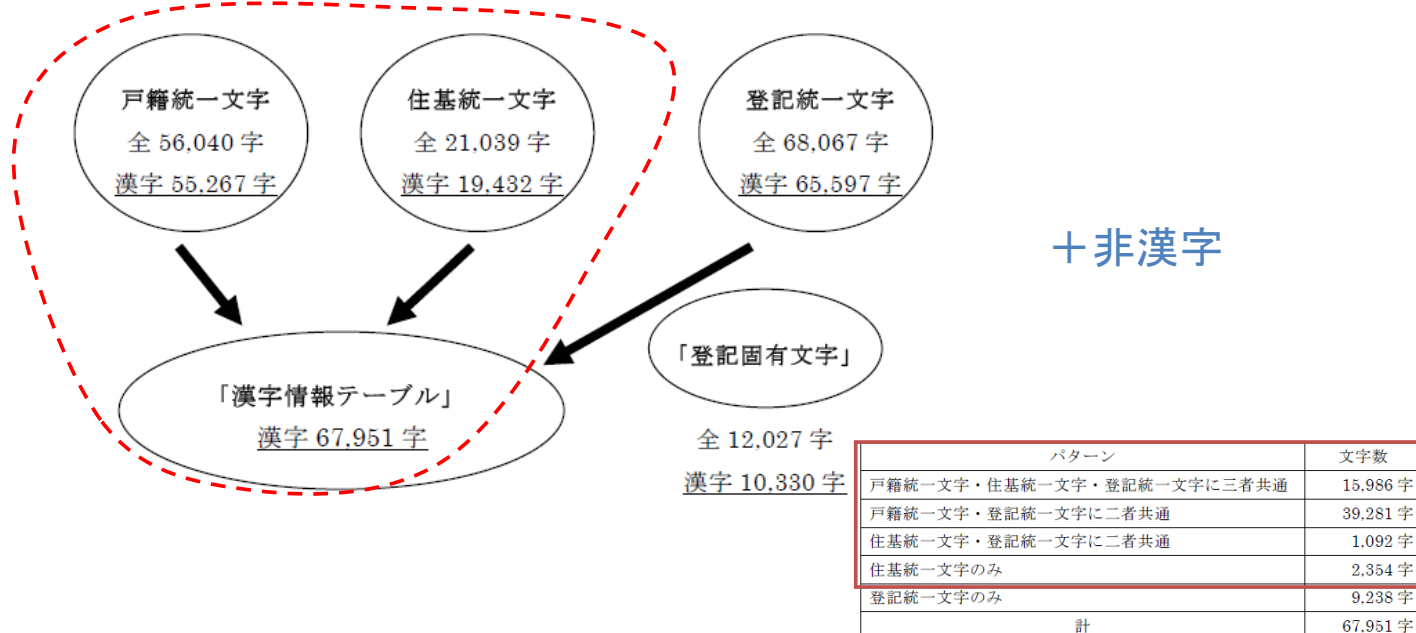
国会図書館  
警察庁  
法務省  
国税庁  
厚生労働省  
独立行政法人国立公文書館  
日本年金機構  
社団法人日本情報システム・ユーザー協会

## 事業の成果物等

14

## 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。(今後の検討事項)



15



- 明朝体のみ整備 (OpenTypeフォーマット)
- メッシュサイズは、1024 × 1024。
- 8ポイントから24ポイントまで公的文書の印字に使用可能な品質。
  - 範囲外の大きさのフォントも、印字可能

11 美しい文字で楽しいソフトウェア開発

24 美しい文字で楽しいソフトウェア

# 72 美しい文

(上記は、イメージの貼り付けであり、実際の印字はアウトラインがきれいに印字される)

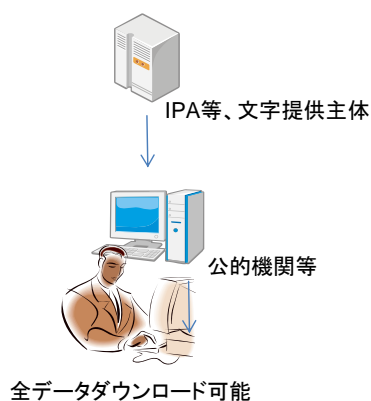
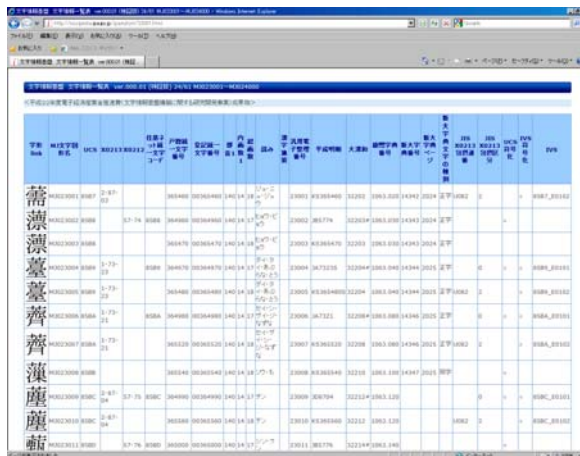
## 文字情報一覧表の整備

- 文字のコード間の関連や文字の持つ情報を一覧で整備。

項目	内容
MJ文字図形名	文字情報基盤の文字図形名。
汎用電子整理番号	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。
平成明朝	汎用電子情報交換環境整備プログラムで制作された平成明朝体グリフのグリフ名。
部首	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	内画数欄には部首内画数を記した。
総画数	総画数欄には文字の総画数を記した。
読み	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。
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)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。
戸籍統一文字番号	6桁の10進数。
住基ネット統一文字コード	4桁の16進数。
登記統一文字番号	8桁の10進数。
X0213	X0213欄には、JIS X 0213:2004の面区点コードを記した。
JIS 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。
JIS 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。

項目	内容
X0212	X0212欄には、JIS X 0212:1990の区点コードを記した。
UCS	UCS欄には、ISO/IEC 10646:2003(Amd.5までの)16進のUCSコードを記した。
UCS符号化	IPAmj明朝フォントへの実装において上記UCSを割り当てた文字にOを記した。
IVS	Unicode ConsortiumのIVD Version 2010-11-14のHanyo-Denshiコレクションに登録されているIVSを記した。
IVS符号化	IPAmj明朝フォントへの実装において上記IVSを割り当てた文字にOを記した。
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。
康熙字典番号	康熙字典番号欄には、『康熙字典』(中華書局発行第7版、1989年)の4桁のページ数と2桁の内文字番号(ページ数 ページ内文字番号)と末尾にOを記した。(参照: <a href="http://www.unicode.org/reports/tr38/#klIRGKangXi">http://www.unicode.org/reports/tr38/#klIRGKangXi</a> )
新大字典番号	新大字典番号欄には、『新大字典』(講談社、1993年3月11日発行第一版)の見出し文字通し番号を記した。
新大字典ページ	新大字典において、当該文字が掲載されているページ
新大字典文字の種別	新大字典に掲載されている正字、俗字等の文字の種別(特に明示のないものは正字としている)

- 提供は無償を想定している。メンテナンス情報などを伝えるため、登録制とする。

文字ID	文字	Unicode	変種	変種ID	変種Unicode	変種名	変種Unicode	変種名	変種Unicode	変種名	変種Unicode	変種名
漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字
漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字	漢字



検索システム整備は別途実施予定

## 今後の検討項目

- 類似文字を容易に扱うことが可能となる文字管理体系の検討を引き続き実施。
  - 渡「辺」、渡「邊」、渡「邊」などの類似の文字を一括して扱える仕組みを検討。
- 文字管理体系は、基本的に国際標準に基づくものとする。国際標準に登録されていない文字に関しては、画像データで提供する。
  - 類似文字の整理方法
  - デザイン差情報の付加
  - IVS等 (Ideographic Variation Sequence) への対応

芦 芦 芦 芦  
1 2 3 4



ベースキャラクターと異体字群で構成される

芦田さんは芦屋のお嬢様だ

82A6

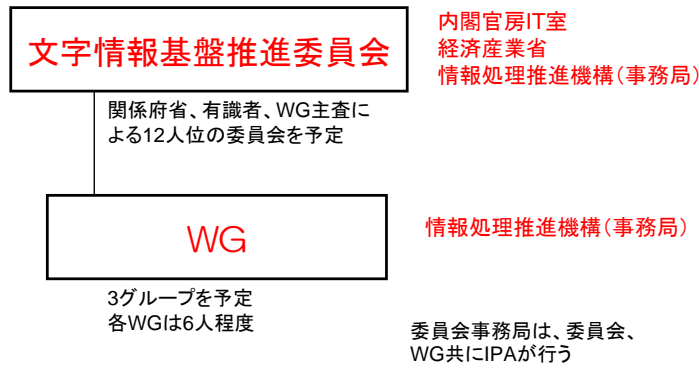
82A6 E0134

82A6で検索することで、一括検索が可能。更に、IVSに対応しているソフトでは異体文字で表示し、未対応ソフトで親文字で表示するなどの工夫も可能

資料: Unicode Technical Standard #37

IVS等の整理時に、文字の統合も図っていくべきであるが、今回の技術的な解決策とは別に議論が必要と考えられる。また、芦 芦 は、表外漢字字体表でデザイン差と記述されている。

- 政策的な検討は、「新たな情報通信技術戦略 工程表」の枠組みで、内閣官房IT室を中心に、関係府省が連携して推進をしていく。また、文字データの実務的運用主体は情報処理推進機構 (IPA) とする。



[これまでの会議資料・議事録](http://ossipedia.ipa.go.jp/article/31/)  
<http://ossipedia.ipa.go.jp/article/31/>

## 検証について

## 検証期間の必要性について

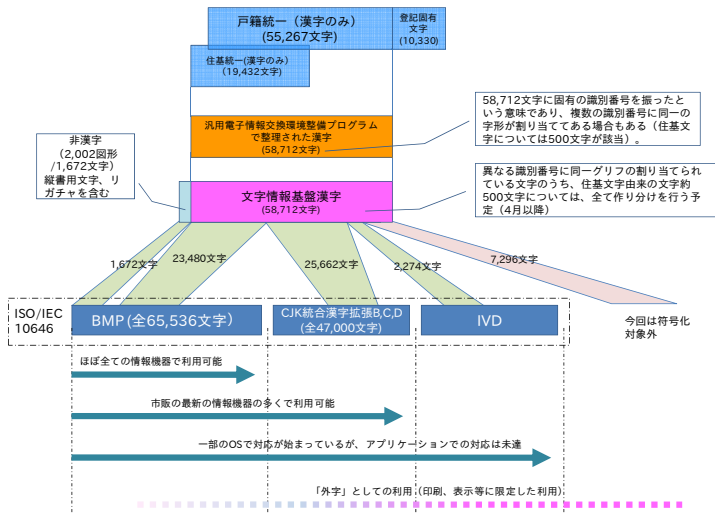
- 行政活動の基盤になるので実装の技術的な検証や運用ルールを導入する方が安全。
- また、運用上必要な500文字を追加開発中。

インデックスとなるフォントで対応させた例

戸籍統一文字	住基統一文字	インデックスを整備 MJ文字図形名
戸籍 150900 孝	住基 6588 孝	MJ013042 孝
	住基 B00B 孝	MJ013043 孝
戸籍 171270 楯	住基 B34C 楯	MJ014230 楯
	住基 696F 楯	MJ014231 楯
戸籍 469770 闌	住基 BE62 闌	MJ005272 闌
	住基 49A8 闌	MJ005273 闌

筆押さえ有り  
筆押さえ無し  
タッチ有り  
タッチ無し  
ハネ  
トメ

辞書がベース  
現場がベース



## 検証版の条件

- 追加開発分を除きダウンロード可能
- システムへの組み込みテストも可能
- 同定作業も可能
- プリンタの検証も可能
- ただし、検証後の本公開で、若干のデータ修正があるので、その場合には上記作業にて戻りが発生する可能性があります。

## 文字情報基盤のご説明

2011年6月16日

独立行政法人 情報処理推進機構

24

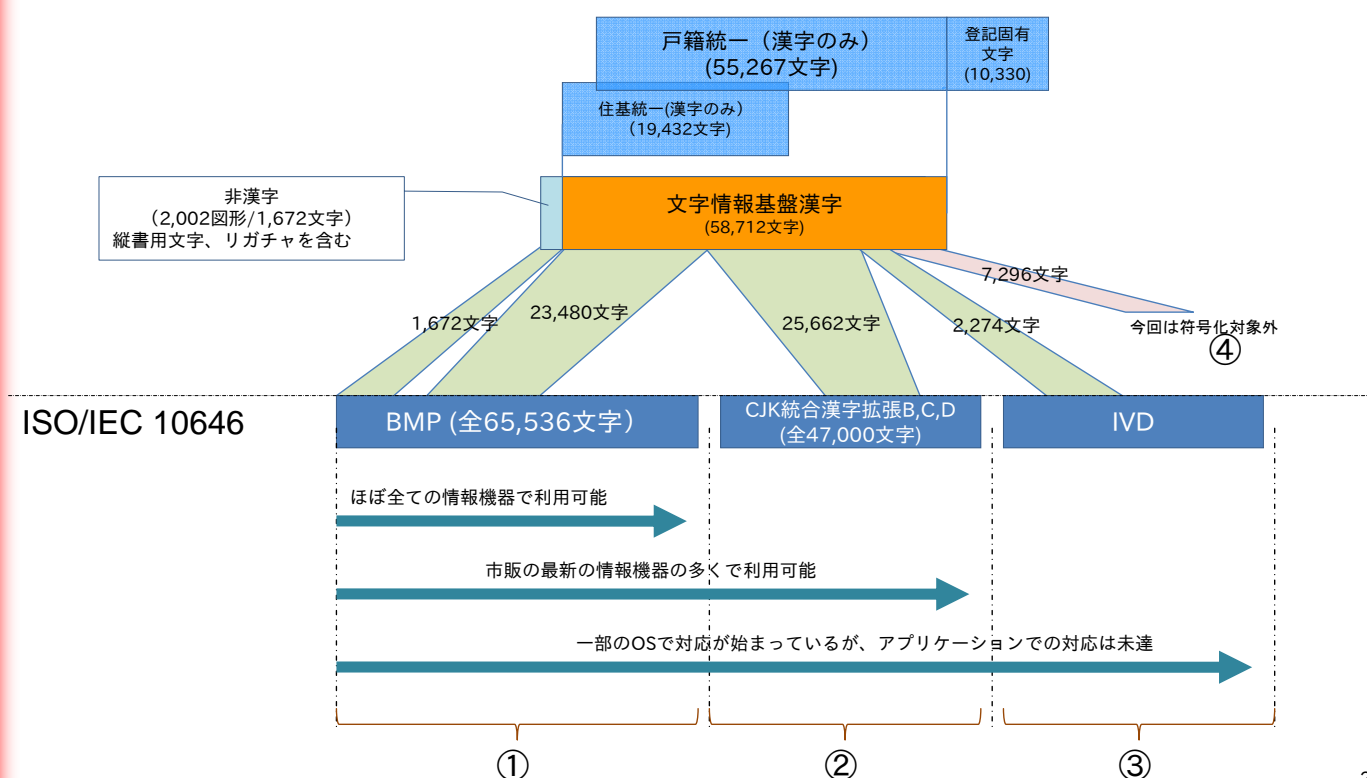
## 成果物の概要

- 報告書
  - ロードマップや利用ガイド（案）等を収録
- IPAmj明朝フォント
  - 戸籍統一文字(55,267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(19,432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7千文字についても継続的に符号化作業を続行）。
- 文字情報基盤文字情報一覧表
  - 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

## 想定される利用局面

- 自治体内での複数システムの連携
  - 戸籍システムと住記システムの連携
  - 住記システムとその他住民サービスシステムの連携
  - 住民サービスのオンラインサービス化
  - 住民サービスのクラウドサービス化
- 複数自治体間の情報交換と連携
  - 転居手続き等のワンストップサービス化、クラウドサービス化

## IPAmj明朝フォントの符号化状況



- 文字フォントのデモ
- 文字情報一覧表のデモ

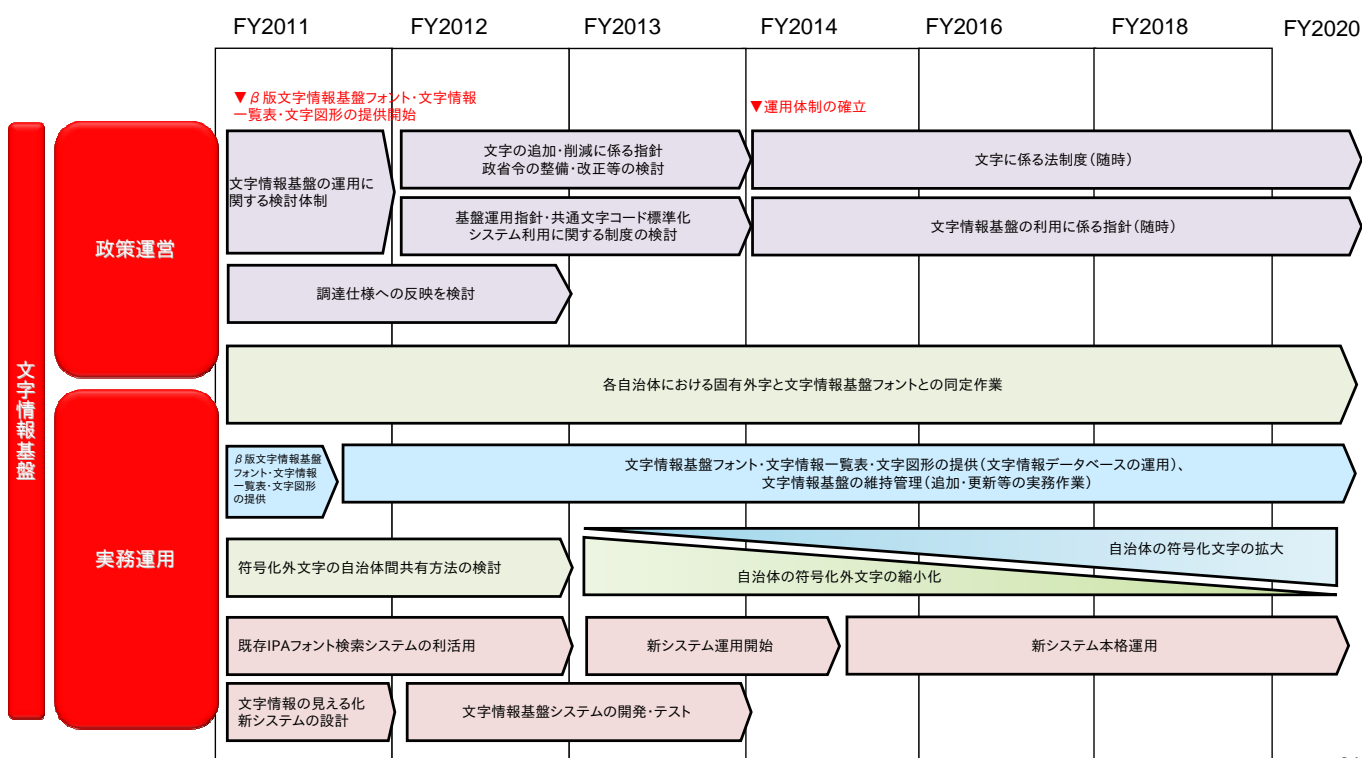
## 今後の実務課題(1)

- 文字図形の追加
  - 住基ネット統一文字では字形を区別しているが、IPAmj明朝フォント検証版では同一の字形となっている文字（約500文字）の作り分け
- ライセンスの検討
  - 文字情報基盤を運用する現場の要望
  - 市場の要望
- 運用ガイドの整備
  - 利用現場向け
    - ユースケースの明確化
    - 文字同定の指針
  - 開発現場向け
    - 文字情報基盤活用のための技術指針

## 今後の実務課題(2)

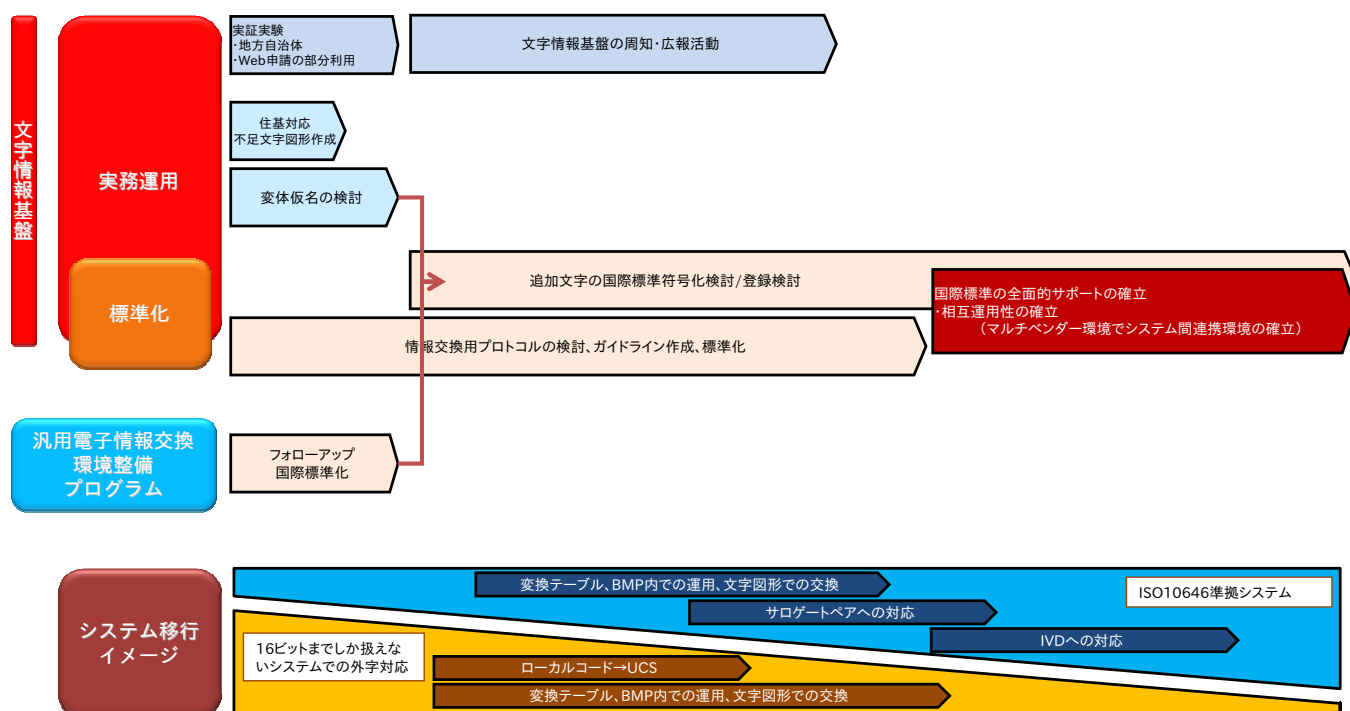
- 符号化作業
  - UCS符号化、IVD登録
- 文字情報一覧表の整備
  - 文字を利用する現場に資する
  - システムを開発する現場に資する
- 文字情報データベースの設計、構築、運用
- 実証実験
  - 一般利用者へのデモ
  - 自治体現場への導入実験

## 文字情報基盤のロードマップ(1)





## 文字情報基盤のロードマップ(2)



32

## 文字情報基盤推進委員会

- 文字情報基盤運用に係る基本方針を検討
- 専門的検討を行うWGを設置
  - 文字同定WG
    - 文字情報一覧表の検討
    - 文字符号化の検討
    - 文字同定ハンドブック等の検討
  - 運用検討WG
    - 技術ガイドラインの検討
    - 文字情報データベースの検討
  - 技術検討WG
    - 情報交換プロトコル等の検討
    - Webを用いたデモ/実証事件の検討

33

ご質問