

文字情報基盤 2013年度の事業報告

2014年2月

独立行政法人情報処理推進機構

- 文字情報基盤成果物バージョンアップ
 - MJ文字情報一覧表Ver.003.01, IPAmj明朝フォントVer.002.02を公開(2013年11月)
- IVD Moji_Joho(MJ)コレクションの登録申請
 - 文字情報基盤で整備した文字図形全てを、国際標準として定義された、一つの集合名で指せることを目指す。
 - 情報規格調査会SC2専門委員会と協力し、IVD登録を申請(2013年12月)
 - 現在Unicodeコンソーシアムが公開レビュー¹⁾中(2014年3月終了予定)
- 変体仮名の検討を開始

1)<http://www.unicode.org/ivd/pri/pri259/>で公開中

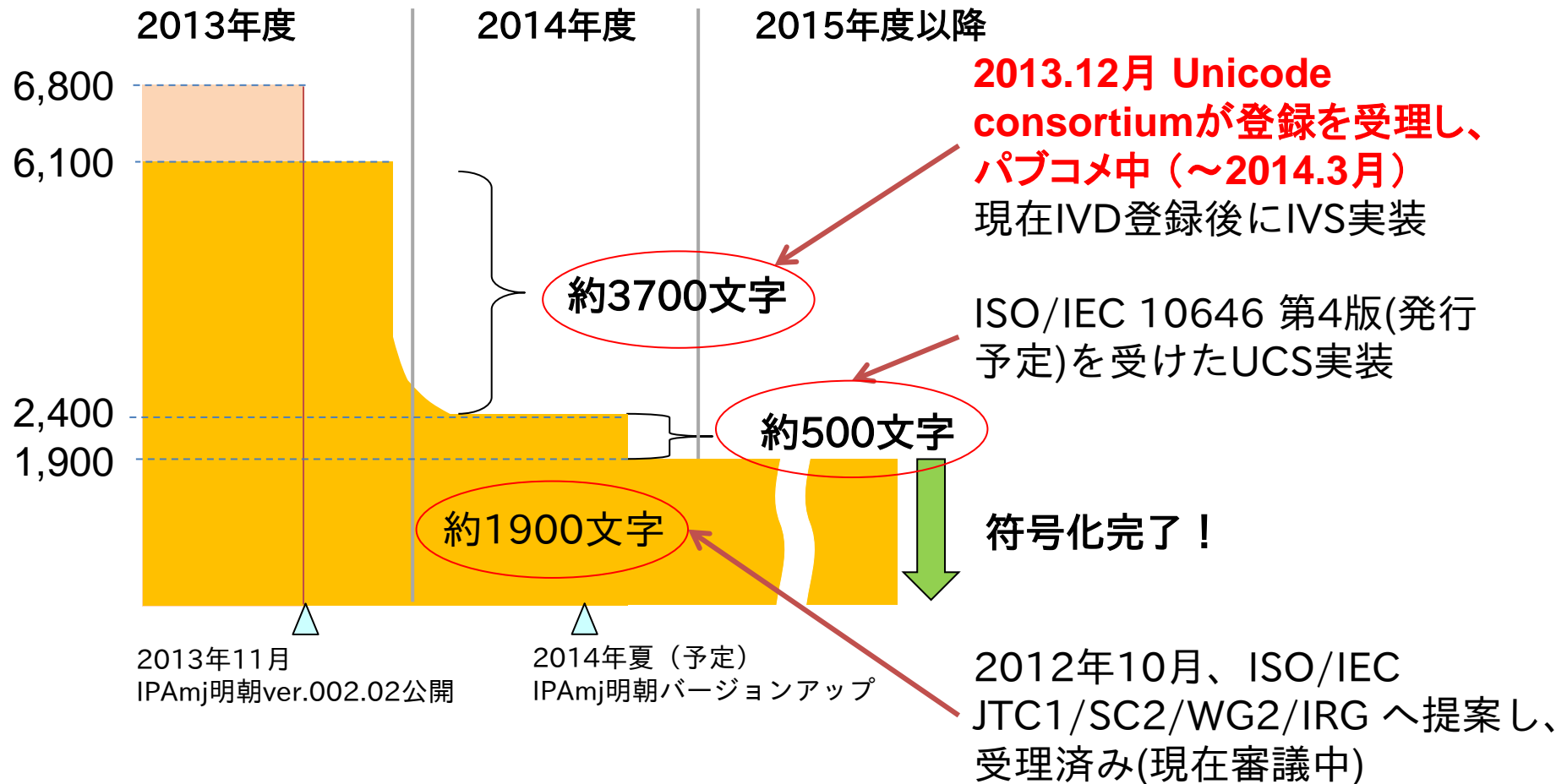
- Unicodeコンソーシアムの運用する文字図形のデータベース (IVD)に、新たな集合名として登録
- 文字情報基盤で整備した全文字図形を収録
 - 基底文字4779文字、合計10,720文字図形
 - 汎用電子成果物の多く (9,687文字図形) を継承
 - 今回の作業では、1 UCS符号位置に複数の図形を持つ文字図形のみ登録。次回、全文字図形の登録を行う計画。

- 拡張F審議（2013年11月 ISO/IEC JTC1 SC2 WG2 IRG会議 於東京）
 - 前回のIRG会議（2013年5月）で拡張FをF1,F2の二つに分割して作業することに合意したが、うち、先に作業を進めるF1については、2014年度中に内容をほぼ確定させることを確認。
- 2014年度作業
 - IRGの拡張F規格化作業を引き続き支援
 - UCSの部分集合を定義する規格(Annex A)に、文字情報基盤の整備した文字の集合を規定する項目を入れるべく、準備を進める。

今後のIPAmj明朝フォント実装計画

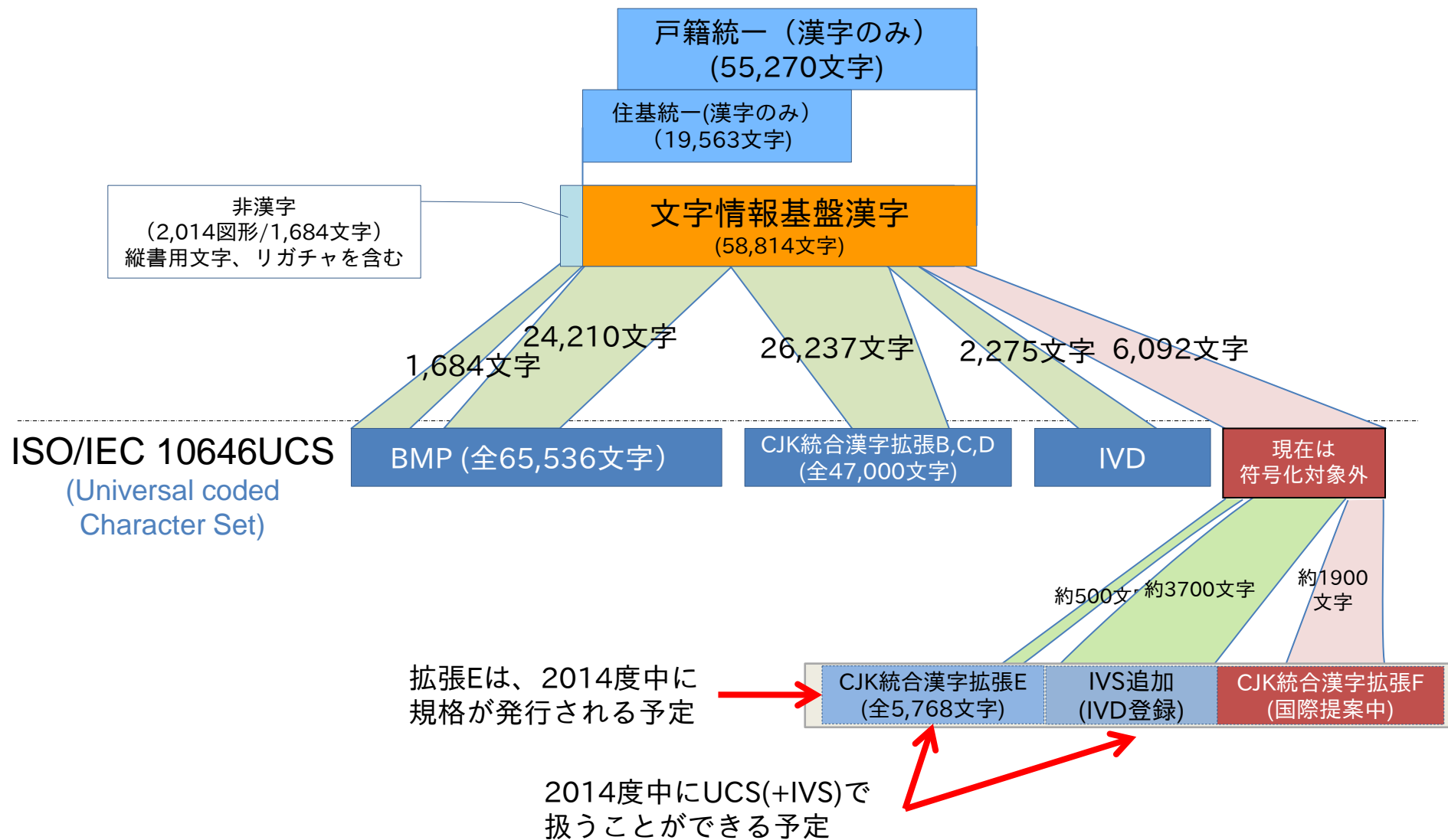


全約6万文字のうち、符号化実装が未了の文字



現在のUCS符号化状況

IPAmj明朝フォントVer.002.02の符号化状況



今後の見通し

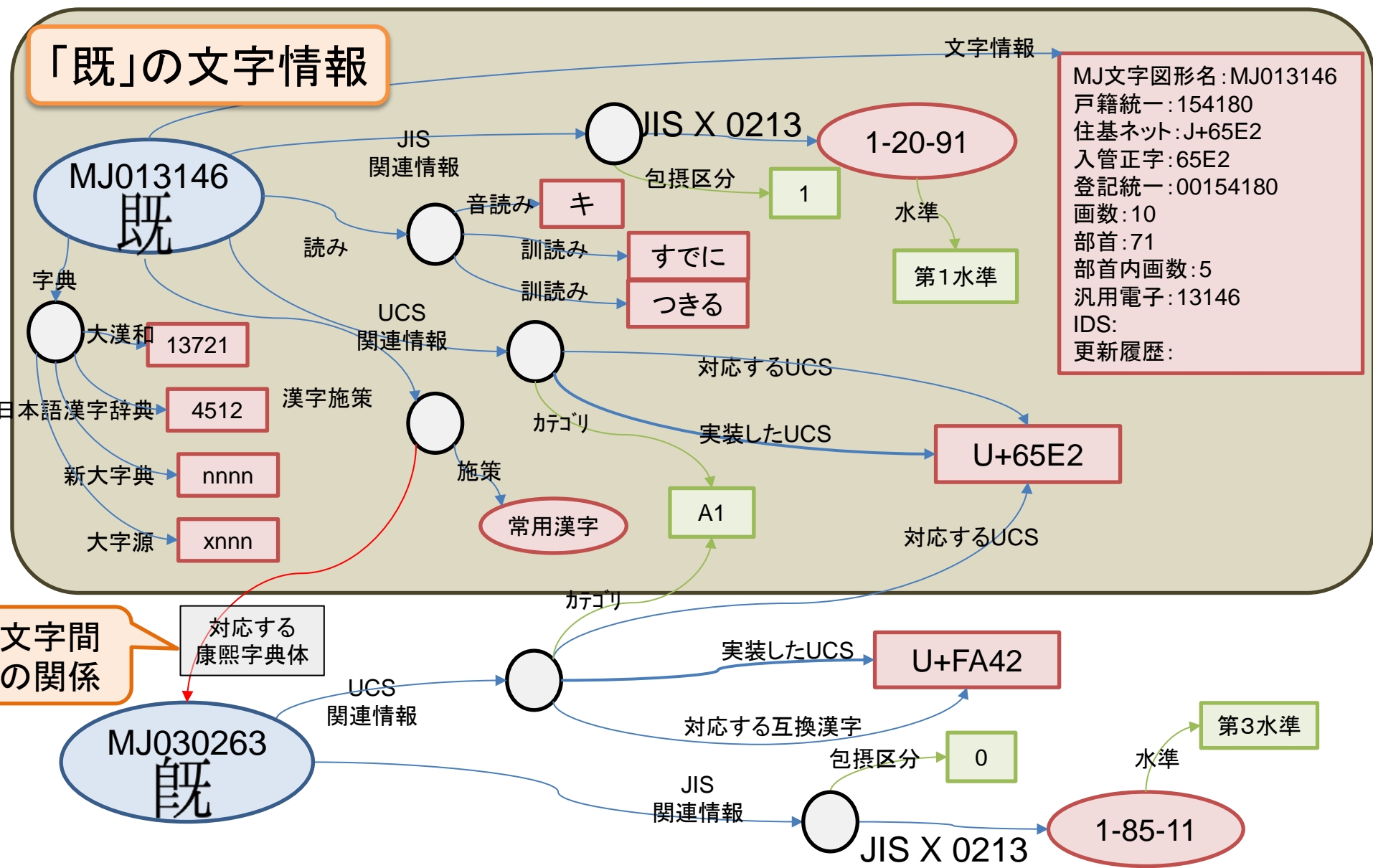
時期	内容
2013年11月 (完了)	MJ文字情報一覧表Ver.003.01, IPAmj明朝フォントVer.002.02を公開
2013年12月 (完了)	MJコレクションのIVD登録をUnicodeコンソーシアムに申請、Unicodeコンソーシアムが公開レビューを開始
2014年3月	Unicodeコンソーシアムの公開レビュー終了、MJコレクション第1期の登録完了
2014年6～8月夏	MJコレクションに対応したMJ文字情報一覧表, IPAmj明朝フォントの公開
2014年度末	ISO/IEC 10646第4版(拡張E) に対応したMJ文字情報一覧表, IPAmj明朝フォントにバージョンアップ
2014年度中	第2期MJコレクション登録の開始

- 変体仮名標準化への現場からの強い要望
- 字形検討のための素材調達（2013年春）
- ヒアリング
 - 学界、印刷・出版業界等
- 標準化検討開始
 - 1 字母- 1 id原則の検討
 - VS(variation selector) 方式の可能性
 - パブリックコメント募集

- 文字のメタ情報を幅広く搭載する
 - 文字コード (JIS, ISO/IEC, その他)
 - 漢字施策との関係
 - 文字の検索を支援する情報
 - 読み、画数、有名辞書の検字番号、IDS(*)情報など
- 文字と文字を多様な観点からリンク付け
 - 字形の類似性
 - 字の起源、意味などの共通性・類似性
 - 大きさの異なる文字集合間での縮退対応への活用 (MJ→JIS X 0213, JIS X 0213 → JIS X 0208等)
- 政府オープンデータ戦略の趣旨に沿い、電子的に再利用しやすい形式で情報を公開する。

(*) IDS: Ideographic Description Sequence :文字を「部品」にばらして表現する方法

RDFによる文字情報記述のイメージ





委員会での検討のポイント

- ✓ 文字間の関係情報について
 - 常用漢字字体と許容字体／康熙字典体との関係
 - 正字と異体字の関係
 - その他文字を関係つける根拠となる情報の検討
- ✓ オープンデータ化をふまえた、データ形式の検討
 - 内部表現：RDF
 - 外部表現：検索API等

一部公開を開始：
文字のメタ情報、対応関係
情報などは利用可能に。

地方自治体向け文字情報基盤対応製品・サービス

製品・サービス	内容	発表時期	発売元
外字統合管理サービス	文字検索・外字作成ソリューション	2013.4月	富士ゼロックスシステムサービス
漢字かなめ	漢字統合管理システム	2014.1月	日立公共システム

IVS対応ソフトウェア(2014年2月現在)

製品	内容	発売元
一太郎2014徹	日本語ワードプロセッサ	ジャストシステム
Microsoft Office 2013	オフィス系ソフトウェア(ワードプロセッサ、表計算ソフト、プレゼンテーションソフト)	マイクロソフト
iWork	オフィス系ソフトウェア(ワードプロセッサ、表計算ソフト、プレゼンテーションソフト)	アップル
Acrobat	PDF編集ソフト	アドビシステムズ
InDesign	印刷・DTPソフト	アドビシステムズ
EdianWing	印刷・DTPソフト	キャノンITソリューションズ



INFORMATION-TECHNOLOGY PROMOTION AGENCY, JAPAN

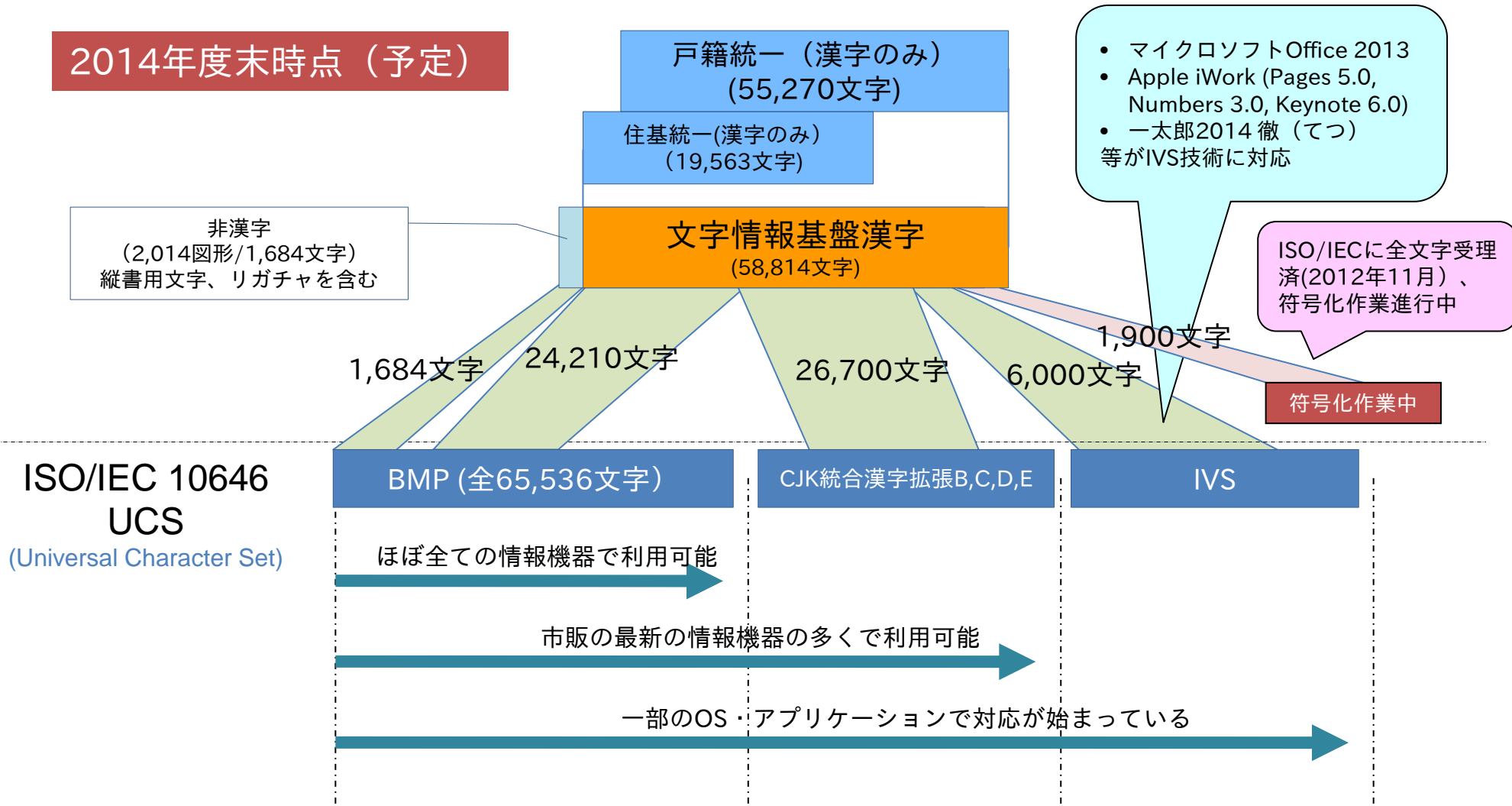
「文字情報基盤導入ガイド」(案)
「文字情報基盤導入テクニカルスタディ」(案)
の概要

2014年2月

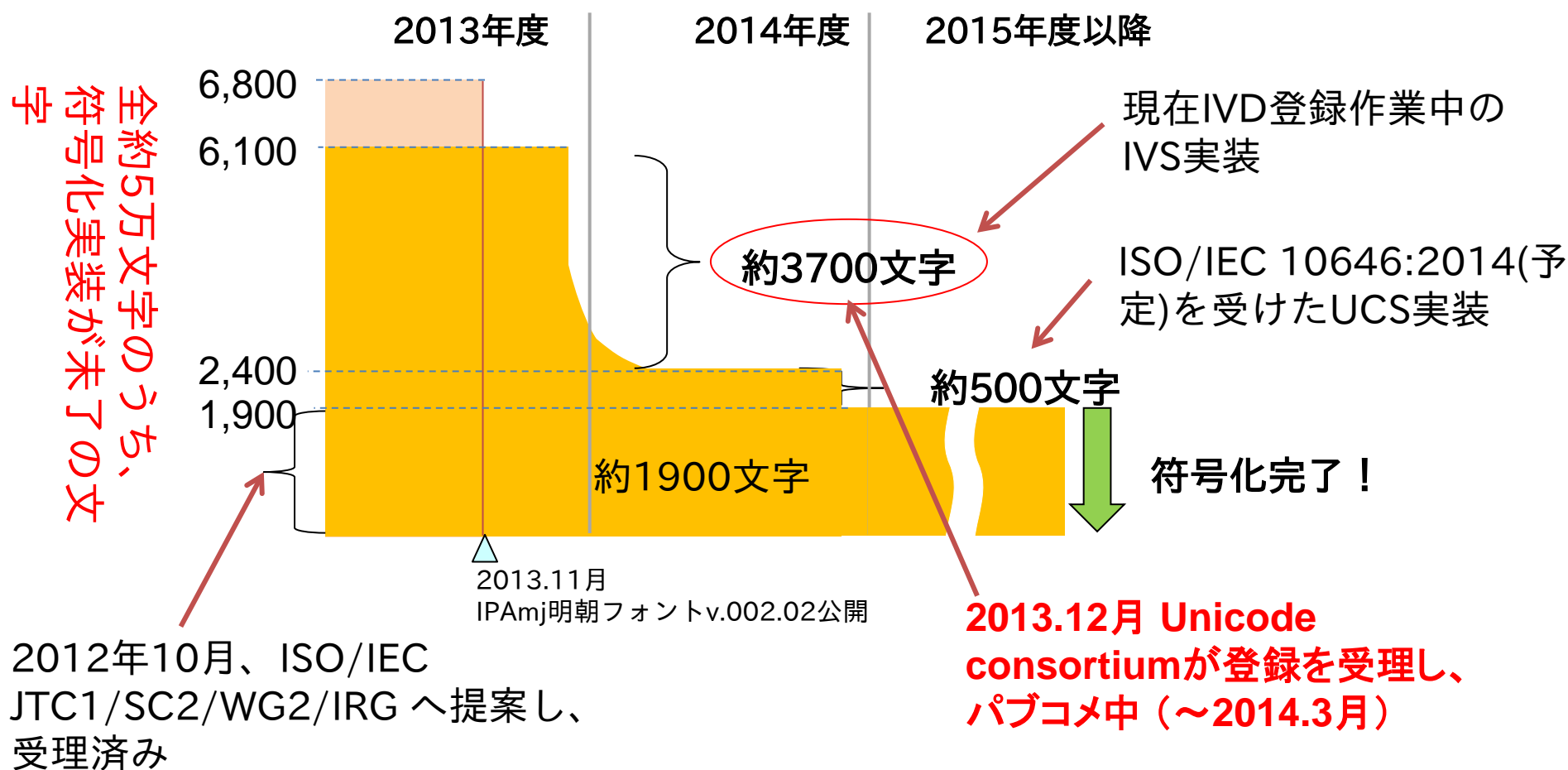
独立行政法人 情報処理推進機構

技術的背景(1)

IPAmj明朝フォントの符号化状況(2014年度末時点予定) と、技術との対応



IPAmj明朝フォント実装状況/計画



「導入ガイド」の目的と方針

- 「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基本的方針を記述
- 国際標準と矛盾しない技術的基本指針を示す
- 経済合理性を考慮する
 - 既存システムとの共存
 - 低い導入コスト
- 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

- 組織内部
 - 短期的：現状の文字体系を変えない
 - 中・長期的：UCS+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：
 - 相互に合意した組織間での使用領域（外字）の利用
 - MJ文字図形名＋図形情報（イメージ）を交換する「交換表現」の活用
 - 中・長期的：UCS+IVSによる情報交換

「テクニカルスタディ」の目的と方針 IPA[®]

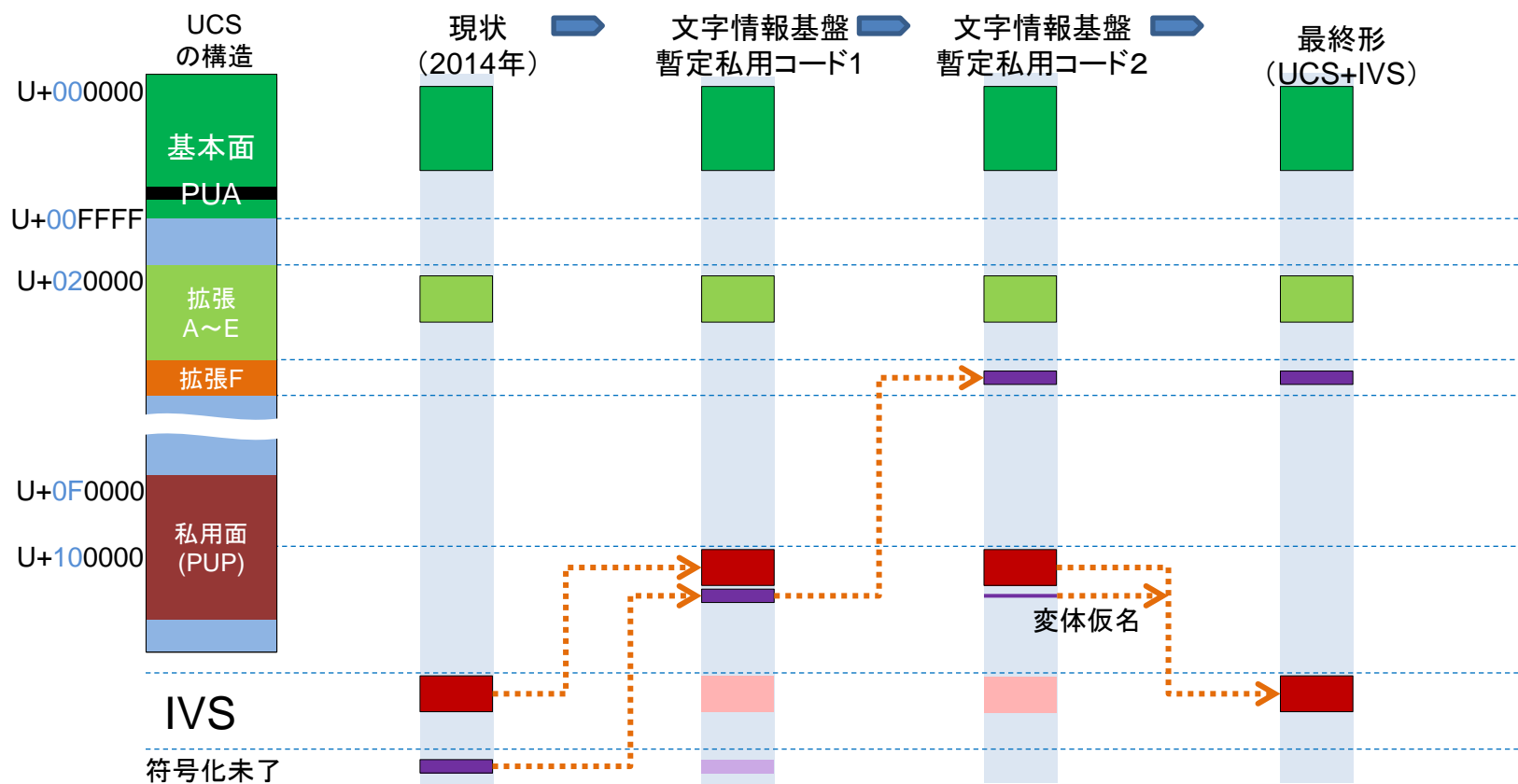
- 符号化、製品の対応状況が未整備な過渡的期間における、文字情報基盤の具体的な導入方法を例示する。
 - 文字情報基盤暫定私用コード1,2の策定(次頁参照)
- 文字活用に係る調達仕様等を策定するにあたって、参考となる技術的資料とする。
- 将来へ向けた移行の手順を例示する。

移行の手順(例)

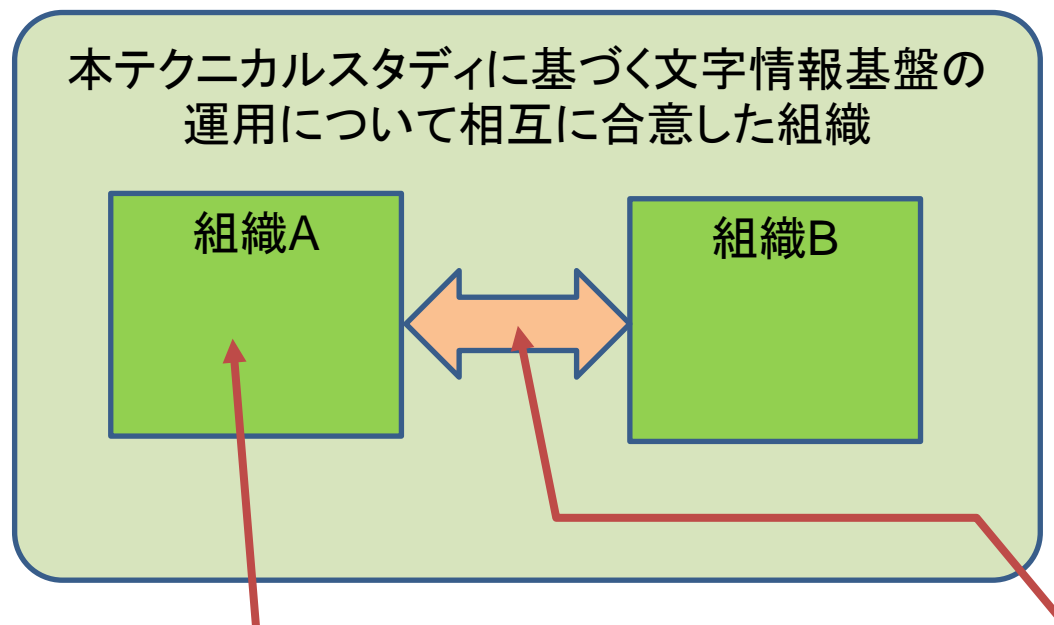
2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |

ISO/IEC作業期間

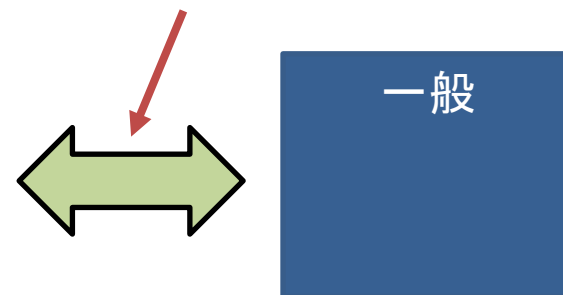
非IVS機器併存期間



「ISO/IEC作業期間」中における 情報交換の原則



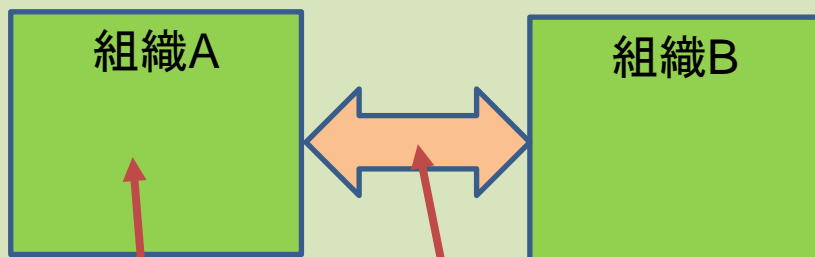
JIS X 0213:2012による交換を原則とする。
Web等での表示の際に、必要に応じ、JIS X 0213:2012範囲外の文字を図形イメージによって表示することは有り得る。



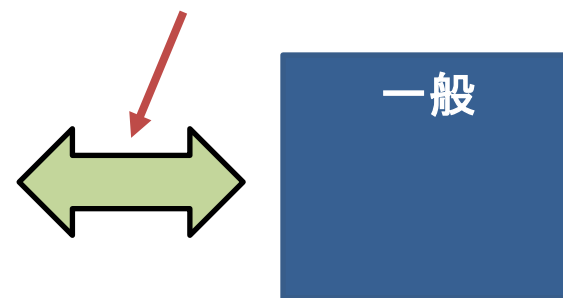
- 組織内におけるコード体系は自由。
- 文字情報基盤の整備した文字図形への同定作業を行うことが推奨される。
- 「ISO/IEC作業期間」中は「文字情報基盤暫定私有コード1」の運用を可とする。
- 「ISO/IEC作業期間」中は、文字情報基盤で整備した範囲の漢字を文字情報基盤暫定私有コード1で交換することを可とする。
- 文字情報基暫定私有コード1の使用は、2019年末を持って終了し、以後、全ての漢字はUCS+IVSにより交換する。

「ISO/IEC作業期間」終了後における 情報交換の原則

本テクニカルスタディに基づく文字情報基盤の
運用について相互に合意した組織



JIS X 0213:2012による交換を原則とする。
Web等での表示の際に、必要に応じ、JIS
X 0213:2012範囲外の文字を図形イメー
ジによって表示することは有り得る。



- 組織内におけるコード体系は自由。
- 「非IVS機器併存期間」においては、「文字情報基盤暫定私有コード2」の運用を可とする。ただし、組織内においても、文書ファイルの保存等はUCS+IVSを用いることが推奨される。
- 文字情報基盤で整備した範囲の漢字をUCS+IVSによって情報交換することを可とする。（文字情報基盤暫定私有コード2は原則使用しない）。

- 符号化提案中の文字図形リスト
- IVSを使用する文字図形リスト
- 「交換表現」のスキーマリファレンス
- 変体仮名の文字図形
- 文字の対応関係のデータベース化
 - 文字セットの縮退対応表への活用度

- 提案中の文字（約1,900文字）の規格化作業フォローアップ
- Unicodeコンソーシアムの字形データベース（IVD）へ登録した“Moji_Johoコレクション”の拡充
- UCSの部分集合定義の新規提案
 - 文字情報基盤で整備した文字図形集合の規定
 - 常用漢字等行政で必要とされる文字図形集合の規定
- 変体仮名の符号化提案