

ITモダナイゼーションSummit 2018 ～どうする？最後のレガシー～

既存COBOL資産を有効活用した 事例紹介

～今後必要となる膨大な種類の漢字にどう対応するか～

2018/4/13

株式会社 日立製作所
公共システム事業部 全国公共システム第一本部

河合 孝志

Contents

1. システムを取り巻く環境における文字の課題
2. システム環境におけるCOBOLの現状
3. 政府が推進するSociety5.0、デジタル・ガバメント実行計画
4. IMI(共通語彙基盤と文字情報基盤)
5. COBOL資産を国際標準の文字情報基盤に対応させるための課題
6. 既存COBOL資産を有効活用した事例紹介
～国際標準の文字情報基盤に対応するために～
7. まとめ

1. システムを取り巻く環境における文字の課題

1.システムを取り巻く環境における文字の課題

システムを取り巻く環境

【オープン化】

- レガシーシステムの再構築
- 既存業務システムの段階的移行
- 既存業務システムとの並行稼働

【システム統合/マイグレーション】

- 複数業務システムの統合
- 既存業務システムの段階的移行
- 既存業務システムとの並行稼働

【クライアント環境の更新】

- 新OS (Windows 7,8.1,10)の段階的導入
- 新OS (Windows 7,8.1,10)への全面切替

【法・制度】

- 住民基本台帳法
- 出入国管理及び難民認定法
- マイナンバー制度
- 官民データ活用推進基本法
- Society5.0の実現

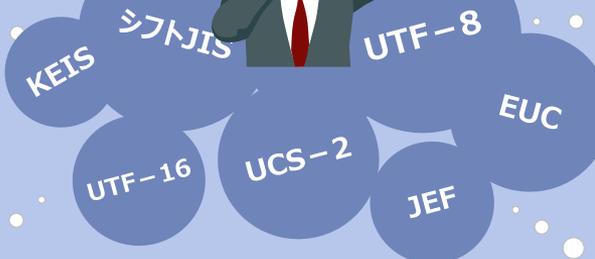
さまざまなプラットフォーム（OS）環境間でのシステム連携やデータ移行が発生

こんな不安はありませんか？

文字規格の混在で、文字化け・字形差が生じる

クライアントOSを更新すると“文字コード”が変わるらしい

文字の重複登録で、業務システム間で名寄せができない



法改正の時の文字対応って一体何をすればよいの

外字の移行・統合をどうしたらいいのかわからない

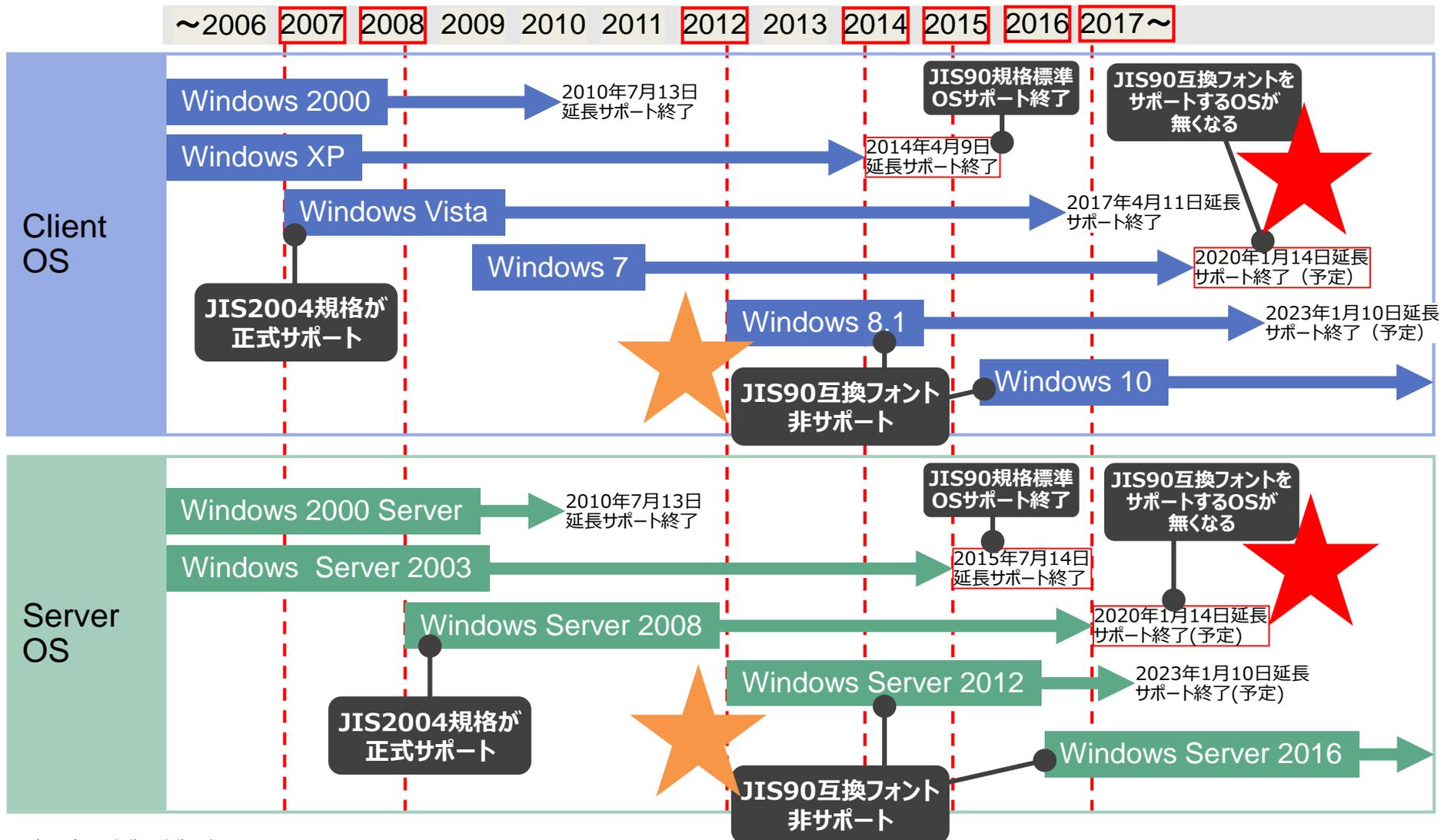
外字の作成・配布・管理を含めた運用に手間がかかる

めざすビジョン

お客様のシステム環境に合わせて、情報システムに必要とされる文字環境（外字を含む）の整備と既存資産を活かし、BPR^{*1}を実現するデータ流通基盤への対応

*1 「ビジネスプロセス・リエンジニアリング」(Business Process Re-engineering) の略。企業活動の目標（売上、収益率など）を達成するために、既存の業務内容や業務フロー、組織構造、ビジネスルールを全面的に見直し、再設計（リエンジニアリング）すること。

1.システムを取り巻く環境における文字の課題



(注1) 各OS名称は略称で表記しています。
 (注2) 本資料では、Microsoft ®,TM, ®マークは表示していません。
 (注3) Windows Server 2003は、Windows Server 2003 R2を含みます。
 (注4) Windows Server 2008は、Windows Server 2008 R2を含みます。
 (注5) Windows Server 2012は、Windows Server 2012 R2を含みます。

Windows Vista, Windows Server 2008以降の対応

**JIS X 0213 (JIS2004)
に対応した日本語入力、
日本語フォントを搭載**

変更点1

**JIS第1、第2水準漢字の
168文字の字形が変更となる。**

変更点2

**JIS第3、第4水準漢字の
文字が標準で入力される。**

**エンコード方式として、
UTF-16を採用**

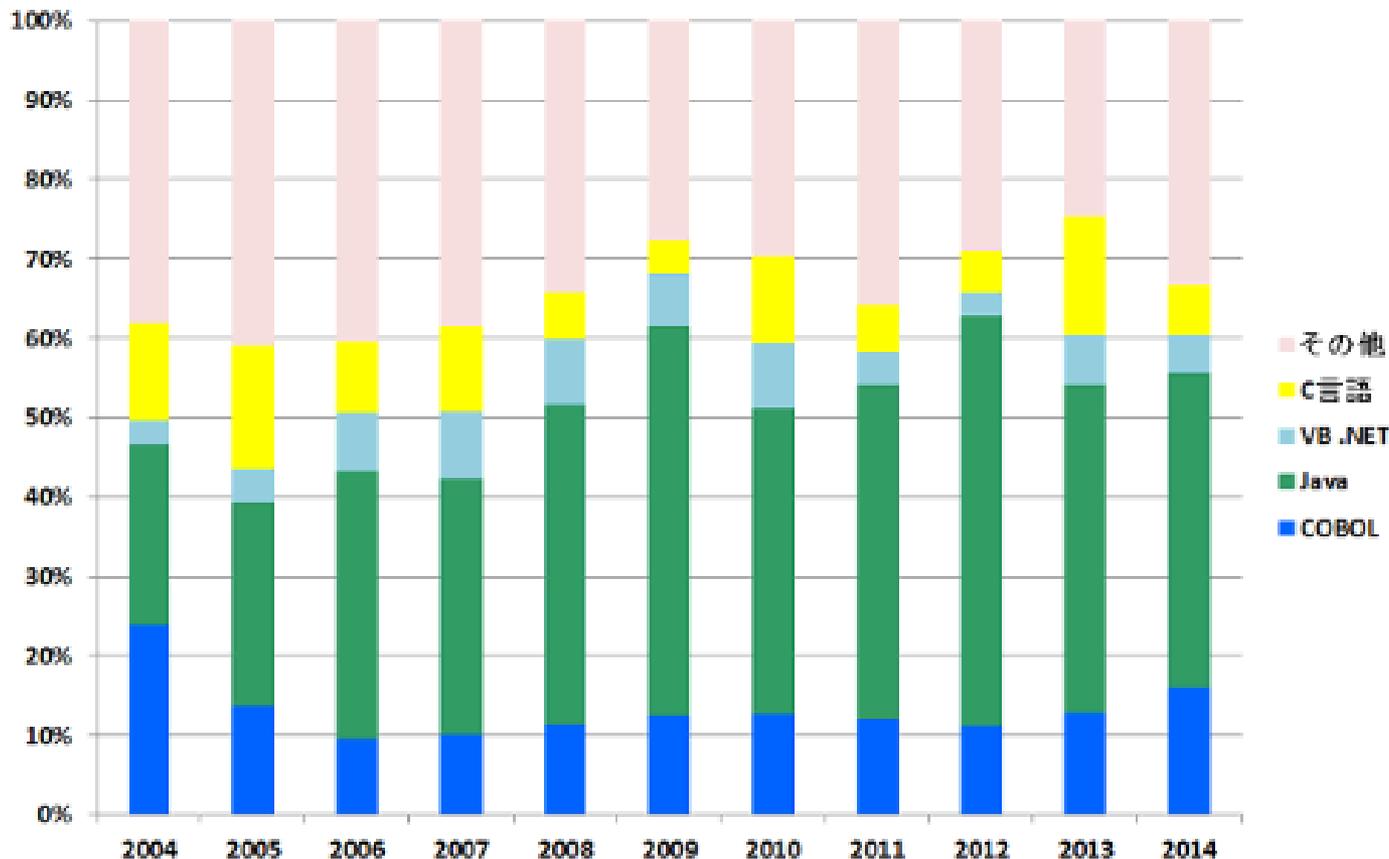
変更点3

**4バイト文字が
標準で入力される。**

2. システム環境におけるCOBOLの現状

■ COBOLの利用状況

開発に用いた言語の経年推移（国内）



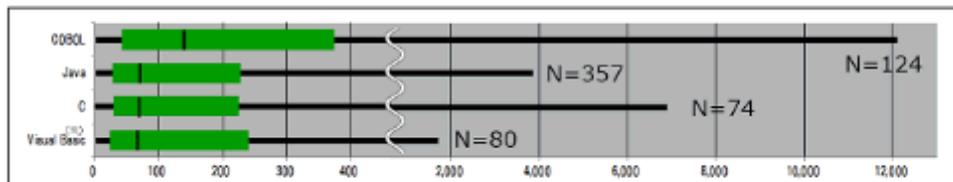
2004年以降に終了したプロジェクトの主開発言語の割合

出典：独立行政法人 情報処理推進機構(IPA) ソフトウェア開発データ白書2016-2017 Copyright 2016 IPA

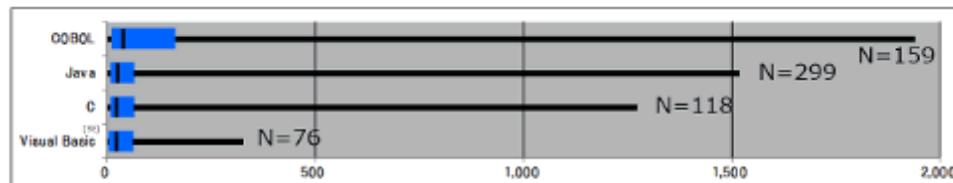
■新規・改良

新規および改良開発に用いた言語と開発規模（国内）

■ **新規開発**（プロジェクト数：635件） （単位：キロ ソースステップ数）



■ **改良開発**（プロジェクト数：652件）



※ ただし、Visual Basic 6.0以前。

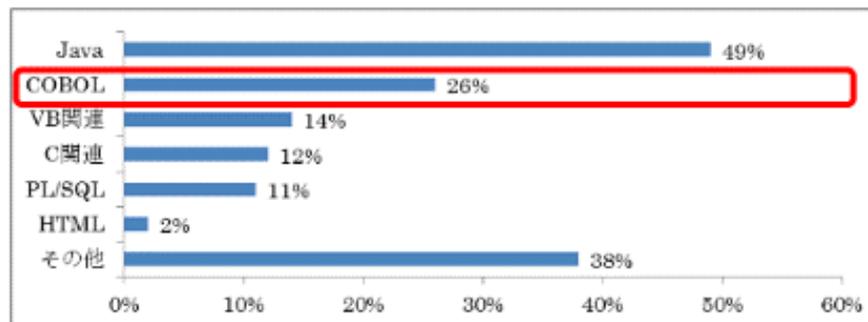
箱の4分の箱は、全体の25%から75%に相当するデータ範囲。箱の中の仕切りは中央値。

出典：独立行政法人 情報処理推進機構(IPA) ソフトウェア開発データ白書2016-2017 Copyright 2016 IPA

■保守

保守で使用している言語（国内）

回答プロジェクト数：528件



出典：経済産業省 情報処理振興課/日本情報システム・ユーザ協会(JUAS)
ユーザ企業 ソフトウェアメトリックス調査2014

■ システム再構築におけるCOBOL資産の取り扱いは・・・

① COBOLからの脱却

- 動的プログラミング言語（JavaScript、Ruby、Python、等）
- 静的型付き言語（Java、C、Swift、C++、等）

② オンライン処理はCOBOL脱却、バッチ処理はCOBOL継続

- オンライン処理は、業務端末要件も加味して、JavaScript、Ruby、Python、Java、C、Swift、C++、等でスクラッチ開発。
- バッチ処理は、大量データ処理を書きやすい（レコード入出力指向）、計算機は二進数でも人に合わせて十進数が中心（基本的に演算は十進数）のCOBOLを継続利用。

③ COBOLの継続利用（保守性が高い）

- 静的言語（アプリケーションが稼働時に使うメモリ量を想定可能）
- 一つの文に多くの処理を詰め込まない（1ステップずつ処理が進み、理解しやすい）
- 長期にわたる互換性
（国際規格が言語仕様を定義、互換性を保ちながら言語仕様が進化）

ターゲット

■ システム再構築におけるCOBOL資産の取り扱いは・・・

③ COBOLの継続利用（保守性が高い）

ターゲット

- 静的言語（アプリケーションが稼働時に使うメモリ量を想定可能）
- 一つの文に多くの処理を詰め込まない（1ステップずつ処理が進み、理解しやすい）
- 長期にわたる互換性
（国際規格が言語仕様を定義、互換性を保ちながら言語仕様が進化）

■ 外的環境変化(P:政治面、E:経済面、S:社会面、T:技術面)へ対応するには・・・

【オープン化】

- レガシーシステムの再構築
- 既存業務システムの段階的移行
- 既存業務システムとの並行稼働

【クライアント環境の更新】

- 新OS (Windows 7,8.1,10)の段階的導入
- 新OS (Windows 7,8.1,10)への全面切替

【システム統合／マイグレーション】

- 複数業務システムの統合
- 既存業務システムの段階的移行
- 既存業務システムとの並行稼働

【法・制度】

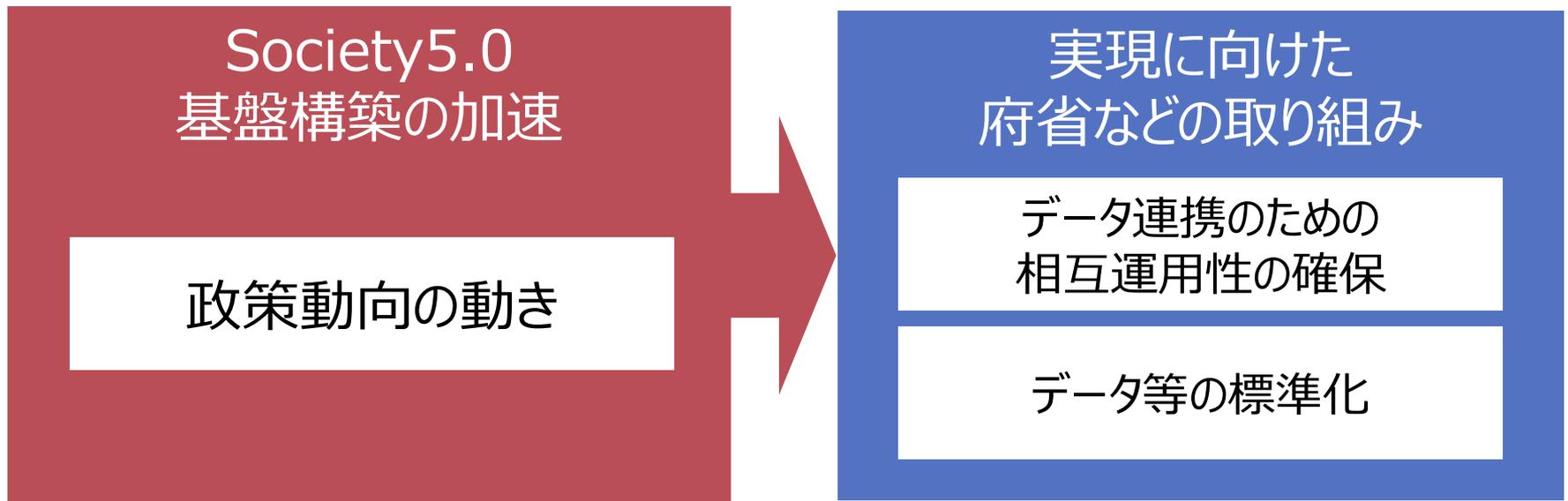
- 住民基本台帳法
- 出入国管理及び難民認定法
- マイナンバー制度
- 官民データ活用推進基本法
- Society5.0の実現

注目

3. 政府が推進するSociety5.0、 デジタル・ガバメント実行計画

未来投資戦略などの政策にて、Society5.0の実現に向けて
Society5.0基盤構築の加速をうたっている。

Society5.0実現に向けた政府動向と府省の取り組みについて紹介する。



■ 未来投資戦略2017

Society5.0の実現に向け、先端技術をあらゆる産業、社会生活に導入していくことが必要だと、未来投資戦略などで示している。

第4次産業革命（IoT^{*1}、ビッグデータ、人工知能（AI）、ロボット）の先端技術をあらゆる産業や社会生活で導入



Society 5.0 の実現

狩猟社会 ▶ 農耕社会 ▶ 工業社会 ▶ 情報社会 ▶ Society 5.0

- ・革新的技術を活かして一人一人のニーズに合わせたサービス提供による社会課題の解決
- ・成長のフロンティア（新たな需要の創出と生産性革命）

 **Society5.0の実現に向けて、施策がより具体化**

*1 「Internet of Things」の略。センサーやデバイスといった「モノ」がインターネットを通じてクラウドやサーバーに接続され、情報交換することにより相互に制御する仕組み。

未来投資会議

2017年11月17日

- 「第4次産業革命」で世界最先端をめざすために必要な事項を整理
- 5つのイニシアティブにて、**Society5.0基盤構築**の加速を宣言

新しい経済政策パッケージ

2017年12月8日

- 「新しい経済政策パッケージ」へSociety 5.0のインフラ整備として横断的な**データ連携基盤構築**を推進
- 国内の整備にとどまらず、欧米との連携を年頭に、日米欧10億人の**データ共通市場**の整備をめざす
- **2020年までに全国に展開できる基盤を整備。**
2019年G20でSociety5.0を世界に発信。

総合科学技術・イノベーション会議(CSTI^{*1})

2017年12月25日

- Society5.0の実現に向け「人工知能」と「データ連携基盤」を両輪として推進
- 「**データ連携基盤**」の推進は、CSTIが司令塔となり、SIP^{*2}／PRISM^{*3}を中核に、**分野毎、分野間のデータ連携基盤を整備**する

*1 Council for Science, Technology and Innovation

*2 内閣府総合科学技術・イノベーション会議が司令塔機能を発揮し、府省の枠や旧来の分野を超えたマネジメントにより、科学技術イノベーション実現のために創設した国家プロジェクト。
Strategic Innovation Promotion Programの略称

*3 政府研究開発投資目標(対GDP比1%)の達成に向け、2018年度に創設する官民研究開発投資拡大プログラム。
Public/Private R&D Investment Strategic Expansion Program

3. 政府が推進するSociety5.0、デジタル・ガバメント実行計画

3

データ連携基盤サブWG(第1回)
(2018.1.23)資料3から抜粋

データ連携基盤の全体イメージ

○Society5.0は全ての産業の生産性を横断的かつ飛躍的に向上。これを実現するためには、現在分散しているデータを連携してビッグデータ化して扱うことが急務。こうした「ビッグデータ」にアクセスするためのデータ連携基盤(例:日本版NIEM[®])を整備。
 ○農業、ものづくり、自動運転、インフラ、防災・減災、スマートシティ等の主要分野でデータ連携基盤・基盤技術を構築すべく、SIP/PRISMで各省庁施策を糾合。2020年までに全国に展開できる基盤を整備。2019年G20でSociety5.0を世界に発信。
 ※National Information Exchange Model



※既にCSTIを司令塔として各省連携(SIP)で進めている農業、自動運転、インフラ、防災・減災分野のデータベース整備を他の分野へ拡大。SIP、PRISMの枠組を活用。

Society5.0 重要課題WGでの整備方針

データ流通の拡充、相互運用性の実現

- Society5.0アーキテクチャの整備
- 日本発の国際標準化

■ デジタル・ガバメント実行計画について

(平成30年1月16日開催「eガバメント閣僚会議(第4回)」資料1-1「デジタル・ガバメント実行計画について」より抜粋)

これまでの電子行政分野の取組の成果

(1) 政府情報システム改革

※青字はここ半年間の新たな進捗

- ① システム数や運用コストの削減など、**政府情報システム改革に大きな成果**
 - システム数: 2018年度までにシステム数半減という目標に対し、**約56%の削減見込(1450→634)**
 - 運用コスト: 2021年度までに3割削減という目標に対し、**約29%の削減見込(削減総額約1,118億円)** (平成29年3月時点から14億円上積み)

(2) IT化・業務改革 (BPR) の推進

- ① 人事・給与、旅費・会計等の府省共通プロジェクトについて、次官級の推進会議を設けて改善計画を策定。**人事・給与システムについては平成28年度末までに全府省等の移行が完了**
- ② ハローワーク、公的年金、国税、登記・法人設立等の**大規模システムについて業務改革まで踏み込んだ刷新を実施し、3割を超える運用コストの削減を実現**

(3) 各府省ITガバナンスの強化

- ① 各府省にサイバーセキュリティ・情報化審議官(副CIO)等を設置(平成28年4月〜)。副CIOに対する研修を実施し、実例による学びを提供。また、NISCと連携してセキュリティ・IT人材の育成を推進 ^{*1} (平成29年度も月1回ペースで研修を実施)

(4) 国・地方を通じたシステム改革の推進

- ① **農地情報公開システムの全国一元化**など、地方自治体も含めたシステム統合の取組にも着手
- ② 政府CIO等が自治体を訪問し、地方自治体における**自治体クラウドの導入を推進** (クラウド導入回体: 948回体 ^{*2} (平成29年4月1日現在、前年より106回体増加))

(5) データ標準の確立

- ① 「文字情報基盤整備事業」で推進していた漢字6万文字の国際規格化 (ISO/IEC 10646) が完了 (平成29年12月)
- ② 行政運営の基本データや各種コード、API等に係る各種ガイドを公開

(6) 行政手続等・行政保有データの棚卸 (全数調査) (平成29年6月〜)

- ① 本年6月以降、各府省が所管する行政手続等や行政保有データについて、各府省の協力のもと、IT室が中心となり、実態を把握するための棚卸 (全数調査) を実施。

『取組の中で得られたノウハウ』(これまで焦点が大幅に不足していた部分)

○事実の正しい認識

(必要な細かさまで実態や問題を徹底的に把握→前提の誤り、業務のばらつきなど問題点の気づき→解決策の発見)

○幹部の意識改革とトップマネジメント

○先行事例から得られたノウハウの共有と横展開

*1 「chief information officer」の略。最高情報責任者を意味する。

*2 「National center of Incident readiness and Strategy for Cybersecurity」の略。内閣サイバーセキュリティセンターを意味する。

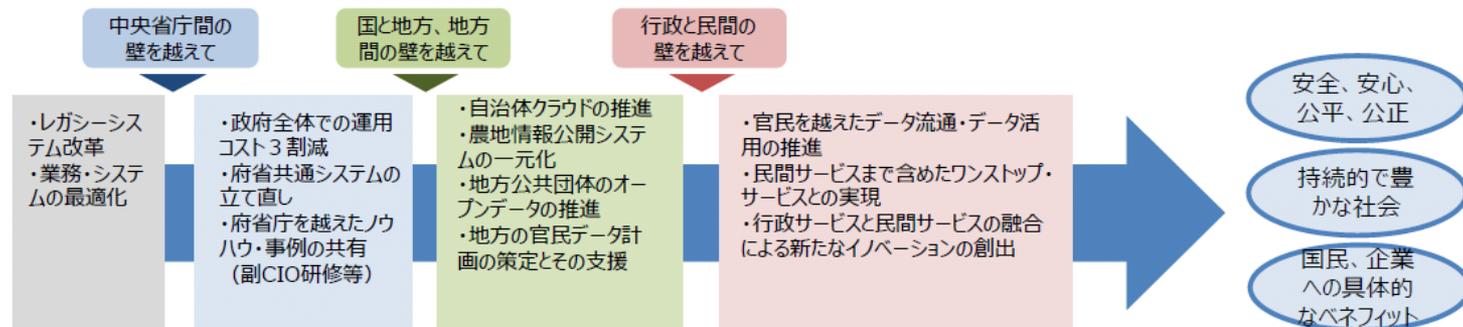
■ デジタル・ガバメント実行計画について (平成30年1月16日 資料1-1より抜粋)

「電子政府」から「デジタル・ガバメント」へ

- これまでの電子行政の取組では、行政内部事務の効率化や、各手続のインタフェースのオンライン化等に取り組んできた（＝「電子政府」）。

- 平成25年の政府CIO設置以降、府省庁の壁を越えた取組（運用コスト3割削減、人事給与システムの本格稼働、副CIO制度の導入・副CIO研修の実施等）を推進。
- 加えて、地方公共団体まで含めた取組（自治体クラウド、農地台帳の一元化、マイナンバーの導入等）を既に実施しつつある。これによって、民間への展開も含め、着実に成果を積み重ねてきているところ。

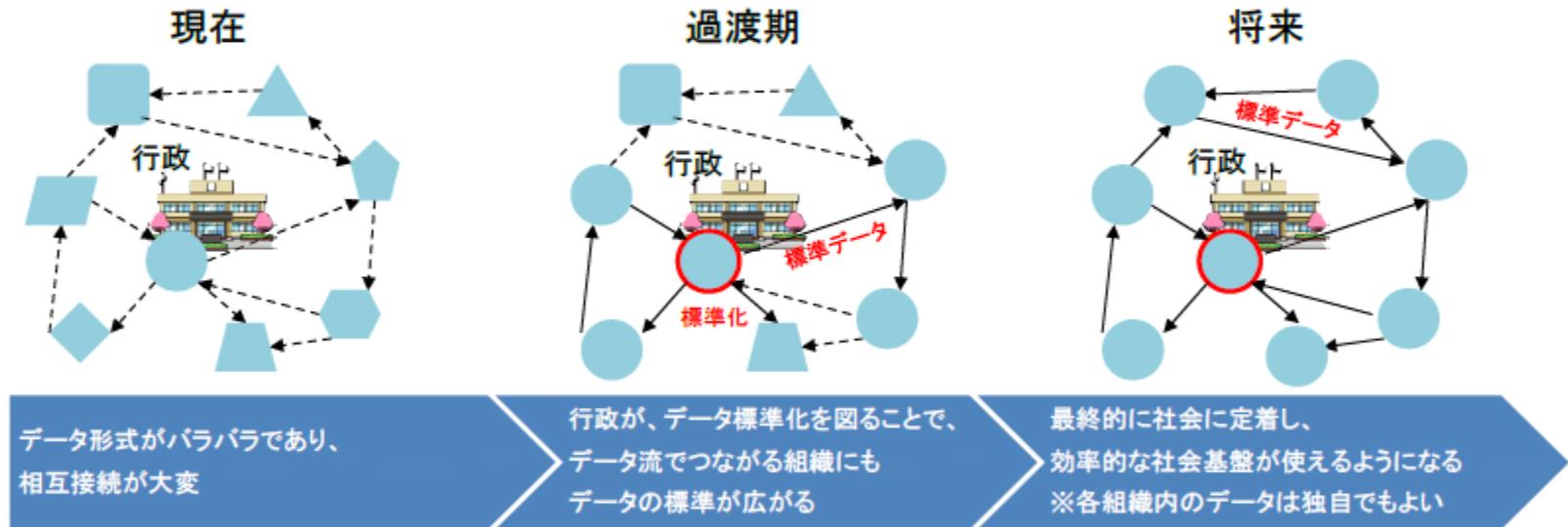
- こうした取組は国際的にも先進事例。今後は、これを更に拡大し、政府・地方・民間全てを通じたデータの連係、サービスの融合を実現し、世界に先駆けた、日本型の「デジタル・ガバメント」の実現を目指す。



■ データ標準の社会全体への展開について

行政におけるデータ標準の重要性

- データ駆動型の社会を作るには、データのインターオペラビリティの確保が重要であり、その核になるのがデータの標準化。



- 申請・届出文書、公開情報、オープンデータの標準化を通じて、データの標準を広げていく。

■ デジタル・ガバメント実行計画について (平成30年1月16日 資料1-1より抜粋)

デジタル・ガバメント実行計画 (案) 具体的な取組 (主なもの)

横断的施策による「行政サービス改革」の推進

【デジタルファースト】

(1) 行政サービスの100%デジタル化

- 各種手続の**オンライン原則の徹底**
- ✓ 手続毎に**業務改革 (BPR)、システム改革**を実施の上、**行政サービスのデジタル化を徹底**する
- ✓ 押印や対面等の本人確認等手法の在り方を再整理
- ✓ 民-民手続についてもオンライン化に向けた見直しを実施

【ワンズオンリー】

- **行政手続における添付書類の撤廃**
- ✓ マイナンバー制度等を活用し、**既に行政が保有している情報は、添付書類の提出を一括して撤廃**
- ✓ 以下の事項の検討と合わせ、添付書類を一括して撤廃するための**法案を可能な限り速やかに国会に提出**
 - ・ 登記事項証明書の提出不要化
 - ・ 住民票の写し・戸籍謄抄本等の提出不要化

【コネクテッド・ワンストップ】

- ✓ 主要ライフイベントである以下の3分野を先行分野とし、民間サービスとの連携も含めたワンストップ化を推進
 - ・ 引越しワンストップサービス
 - ・ 介護ワンストップサービス
 - ・ 死亡・相続ワンストップサービス

【オープンデータの推進】

(2) 行政保有データの100%オープン化

- **オープンデータ・バイ・デザイン**の推進
- ✓ **オープンデータを前提とした業務・システムの設計・運用**の推進
- **ニーズの把握と迅速な公開**
- ✓ **民間事業者等との直接対話**を通じた民間ニーズの把握とこれに対応したオープン化の加速
- ✓ **推奨データセット**に基づくデータ公開の推進

【行政データ標準の確立】

(3) デジタル改革の基盤整備

- **行政データ連携標準**の策定
- ✓ 日付・住所等のコアとなる行政データ形式について、平成29年度末までにデータ連携の標準を策定
- **語彙・コード・文字等の標準化**
- ✓ **施設・設備・調達等の社会基盤となる分野**について、語彙・コード等の体系を行政データ標準リスト (仮称) として整理

【法人デジタルプラットフォームの構築】

- ✓ **複数手続を一つのIDで申請できる認証システム**の整備や**法人インフォメーションの活用**等を通じ、データが官民で有効活用される基盤を構築

【政府情報システム改革の着実な推進】

- ✓ これまでの取組により、約**1,118億円**の運用コストの削減を見込んでいる状況。政府情報システム改革を引き続き推進し、システム数の半減、**運用コストの3割削減を確実に達成**

各府省計画の策定と個別分野のサービス改革

【各府省中長期計画の策定】

- ✓ 各府省のITガバナンスを強化し、上記の各種取組を推進するため、**各府省におけるデジタル改革の中長期計画を平成30年上半期を目途に策定**

【個別分野におけるサービス改革】

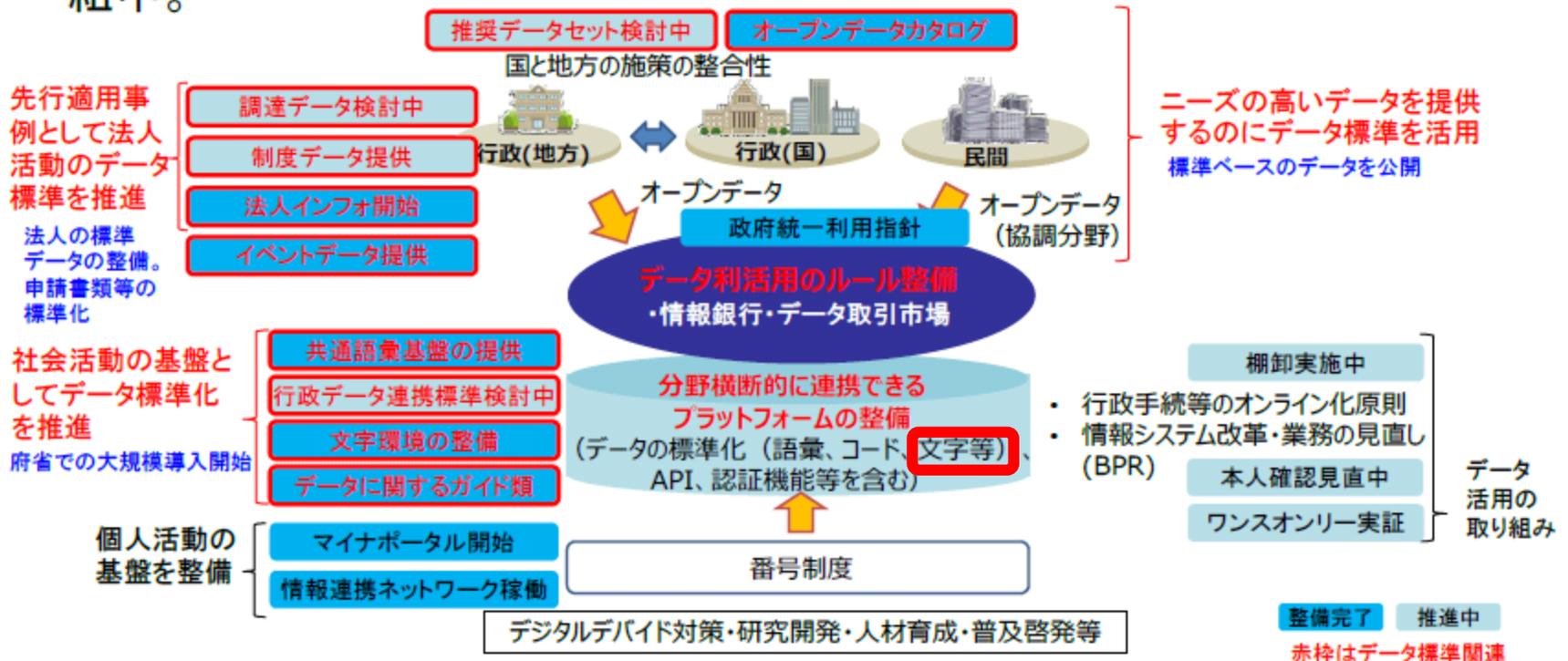
- ・ 金融機関×行政機関の情報連携(預貯金等の照会)
- ・ 電子調達サービスの利便性向上
- ・ 社会保険・労働保険関係の電子申請の利用促進
- ・ 遺失物法関係サービスの利便性向上
- ・ 法人設立手続のオンライン・ワンストップ化、法人登記情報連携の推進
- ・ 公的年金関連サービスのデジタル化
- ・ 自動車安全運転センターによる各種証明書発行サービスの利便性向上
- ・ 在留資格に関する手続のオンライン化
- ・ ハローワークサービスの充実
- ・ 無線局開設手続等に係る行政サービスの更なるデジタル化
- ・ 旅券発給申請方法等のデジタル技術の活用による多様化
- ・ 特許情報提供サービスの迅速化
- ・ 住民税の特別徴収税額通知の電子化等
- ・ 相続税申告のオンライン化
- ・ 自動車保有関係手続のワンストップサービスの充実

「行政手続等の棚卸」等を踏まえ、以下の分野で先行的にサービス改革を推進

■ データ標準の社会全体への展開について

官民データ活用推進基本法との関連

- オールデジタルに向け、分野横断的に横断できるプラットフォームを整備するとともに、ショーケースとして法人分野のデータ標準化に集中的に取り組中。

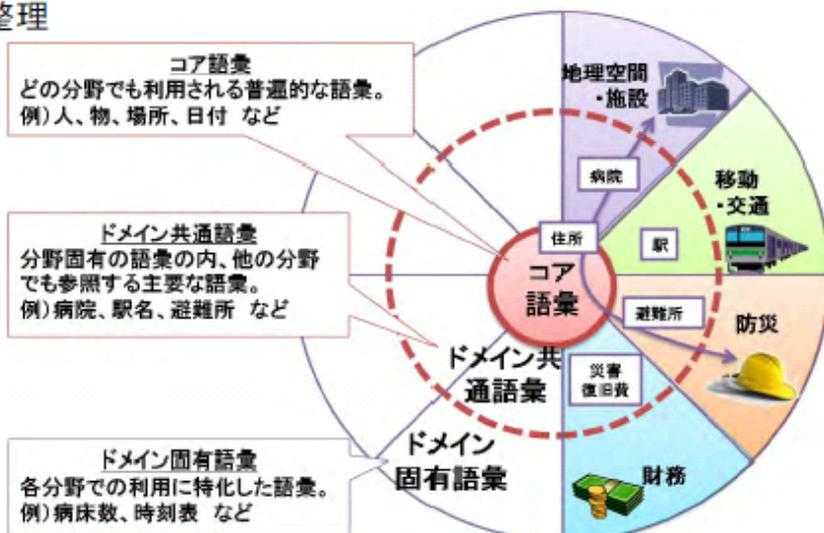


4. IMI(共通語彙基盤と文字情報基盤)

共通語彙基盤 (IMI: Infrastructure for Multi-layer Interoperability)

分野横断でのデータ交換を目的としたフレームワーク(世界最先端IT国家創造宣言で推進)

- ・デジタル・ガバメントの基盤として取り込まれるが、将来は社会全体で活用することを前提に設計を実施
- ・米国政府の推進するNIEMと同様に、社会の中核になるコア語彙と分野別の専門分野(ドメイン)語彙を体系的に整理



IMIの特徴

- ・分野横断 (社会基盤のコアな情報を重点推進)
- ・グローバル連携 (EU、米国との情報交換)
- ・IoTへの配慮 (将来的な連携を視野に入れて設計)
- ・オープンデータでの活用 (社会全体のデータ利活用を促進)
- ・検索性向上への配慮 (検索サービス標準の参照)
- ・既存システムへの配慮 (既存データを活かしデータ連携時に活用)

国内の実装状況

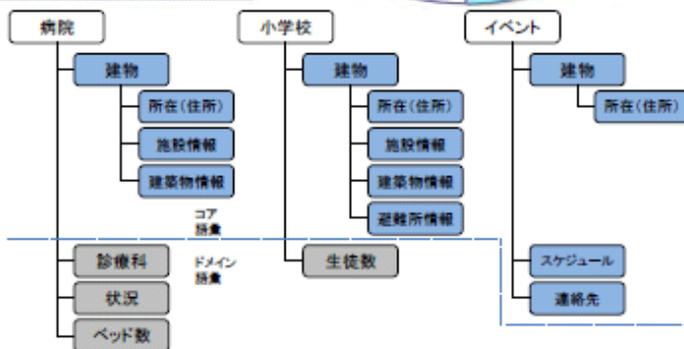
- ・政府機関、自治体、民間で導入が開始
 - ✓ 国が保有する4百万法人の情報を提供する「法人インフォメーション」でIMIを全面採用
 - ✓ 埼玉県では県下市町村を含んだ広域の情報提供に導入
 - ✓ 多くのオープンデータ関連プロジェクトで活用

海外との連携状況

- ・EUのデータ標準プロジェクト総会に毎年参加。(講演等)
- ・米国も2年に一度情報交換。
- ・行政コアデータとオープンデータを同時に進める方式は日本独自の方式で注目されている

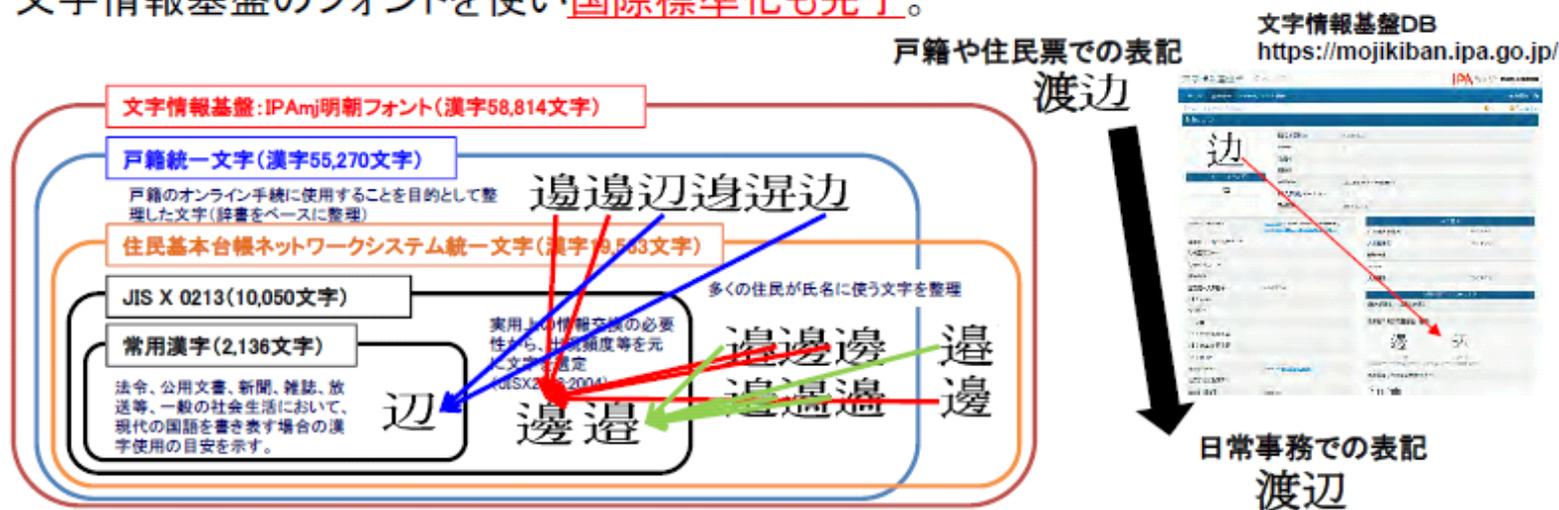
体制

- ・創造宣言の中で取り込まれてきたが、官民データ活用推進基本法が整備されたこともあり、重点化。
- ・IT総合戦略室が経済産業省、情報処理推進機構と推進。



文字情報基盤等

- 戸籍システムや住民基本台帳ネットワークの文字を全て網羅し、全ての国民の氏名をコンピュータで扱うことが可能になるよう、無償のフォントを提供。
- 日常生活での利便性を確保するため、これまで外字で表していた戸籍の文字等をJIS第4水準に対応させる縮退マップを提供。(法人番号のシステムでは法人名を縮退マップでJIS第4水準に縮退)
- 行政機関の全ての業務で活用可能であり、外字作成・管理コストの削減、システム間の相互運用性の確保が可能。国や他自治体へのデータ連携等も容易。
- 文字情報基盤のフォントを使い国際標準化も完了。



※JIS X 0213(JIS4水準)は、最新の情報機器において正確な表記が可能

- 実装を進めるため、代替文字、よみがな、英字の扱いを検討開始するとともに、対象も氏名だけでなく、法人名、地名等に拡大。

■デジタル・ガバメント実行計画について

(平成30年1月16日開催「eガバメント閣僚会議(第4回)」 「デジタル・ガバメント実行計画」より抜粋)

1) 行政データ標準の確立

イ. 文字情報基盤の活用等による文字環境の整備 (◎内閣官房、◎経済産業省、法務省、財務省、全府省)

文字情報基盤については、2017年(平成29年)に国際標準(ISO/IEC 10646第5版)が完了したところである。本基盤の活用を推進するとともに、更なる環境の整備を実施し、文字情報に関する相互運用性を確保する。

各府省は、今後情報システムを整備する際、一般的業務に係る行政システムにおいては使用する文字の範囲は **JIS X 0213 を原則** とし、UCS⁵に従った表現(符号化及び記述法)でシステムの設計・構築を行う。加えて、現在個別に外字を使用している情報システムは更改時にその必要性を見直す。この際、**戸籍氏名文字を表すことが必要な場合には、情報システムの現状を把握しつつ、文字情報基盤を通じたデータ連携を可能とする。**

内閣官房は関係府省の協力の下、漢字、代替文字、フリガナ及びローマ字等を含む文字情報の現状や導入方法に関するガイドを2017年度(平成29年度)末までに整備する。

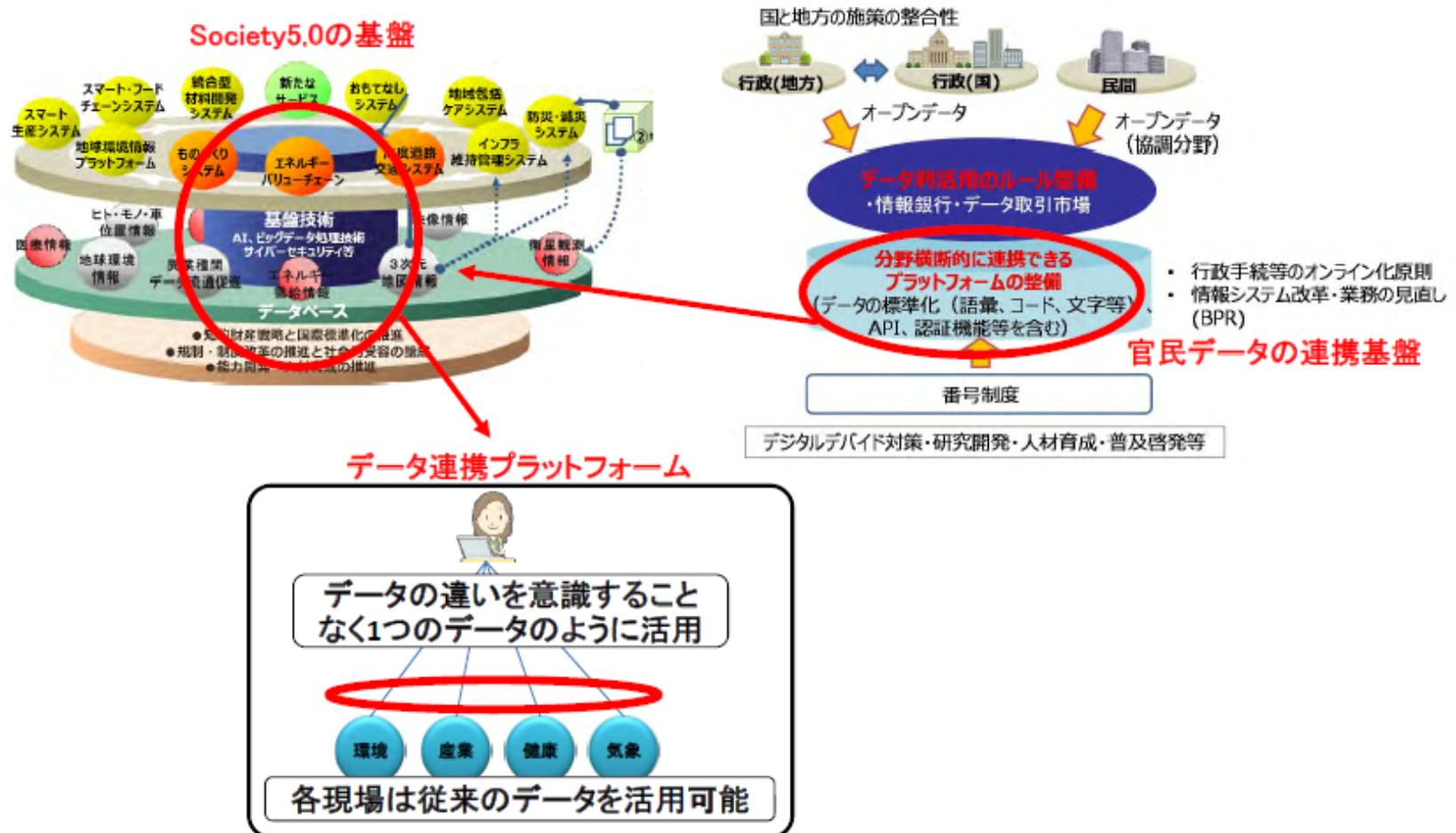
法人名に関しては、内閣官房は関係府省と協力し、法人名の英語表記の普及を図る。法務省は登記申請書への法人名のフリガナの記載を2018年度早期に開始する。

KPI: 文字情報の導入方法に関するガイドの策定

KPI: 地方公共団体の調達における文字情報基盤参照割合

Society5.0との連携

■ Society5.0の各分野をつなぐ基盤として官民データの連携基盤を活用。



5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■文字情報基盤事業とは・・・

独立行政法人 情報処理推進機構（IPA）が
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）を受託

●行政で使われる文字を整理

戸籍統一文字
55,270の漢字

住民基本台帳ネットワークシステム統一文字
19,563の漢字



同定・整理して58,842字を国際規格と対応付け or 新規提案

- JIS漢字，常用漢字，表外漢字，人名用漢字まで含めて整理
- 文字図形，各種文字コード，読み等の情報を一覧化
- フォントを整備

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■文字情報基盤事業とは・・・

- 同定・整理した全ての文字を国際標準で使えるようにする
 - ISO/IEC 10646 (Unicodeの国際標準版) に登録済みの文字と対応付け
 - ISO/IEC 10646 に登録されていなかった文字を提案し標準化
 - UnicodeコンソーシアムにIVD^(*)提案し、規格上はひとつの文字に包摂（統合）された複数の文字を、IVS^(*)を使って区別できるようにする

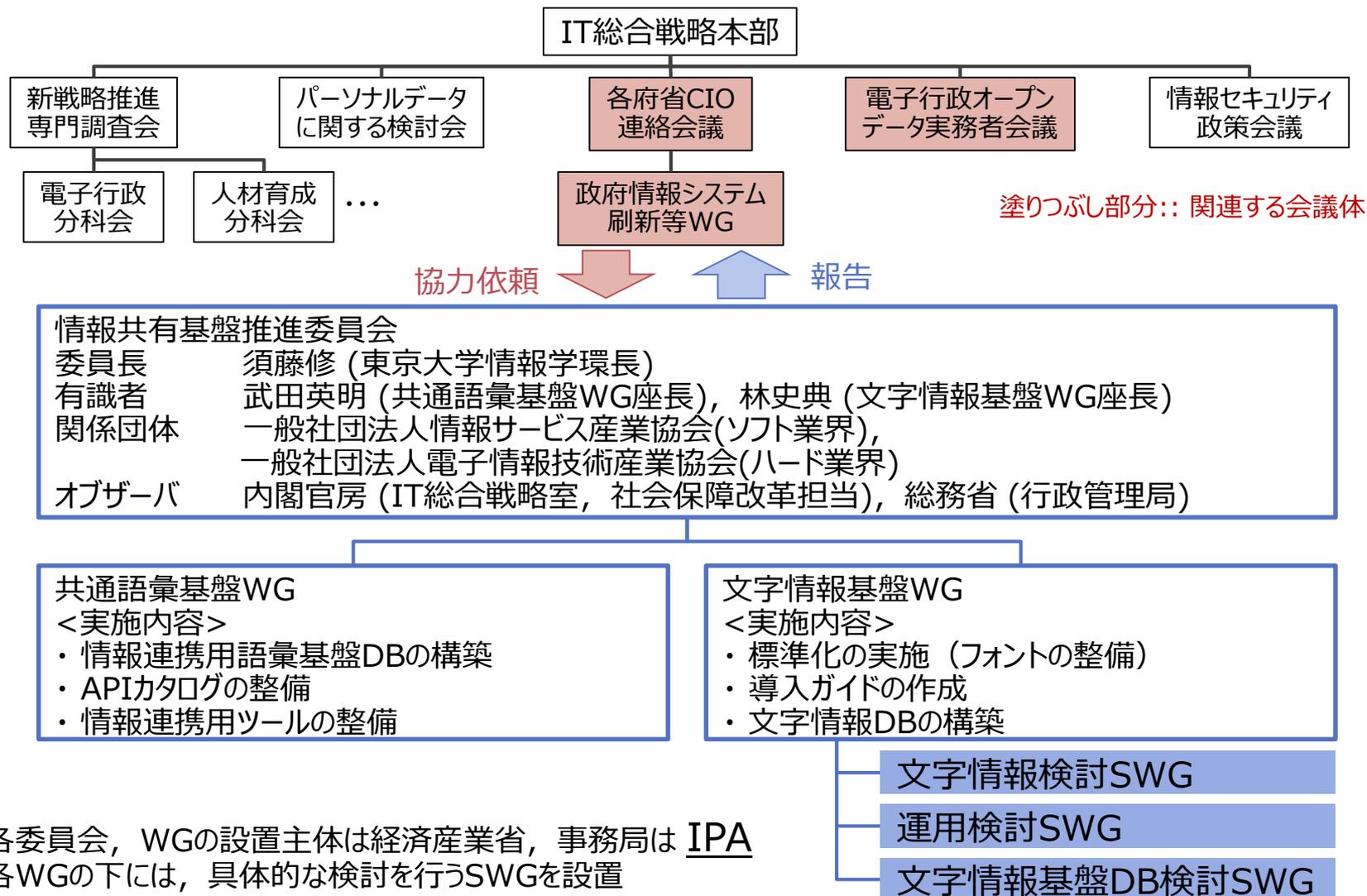
外字のないデータ流通の世界へ

(*) IVD: Ideographic Variation Database

(*) IVS: Ideographic Variation Sequence/Selector

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■文字情報基盤事業とは・・・

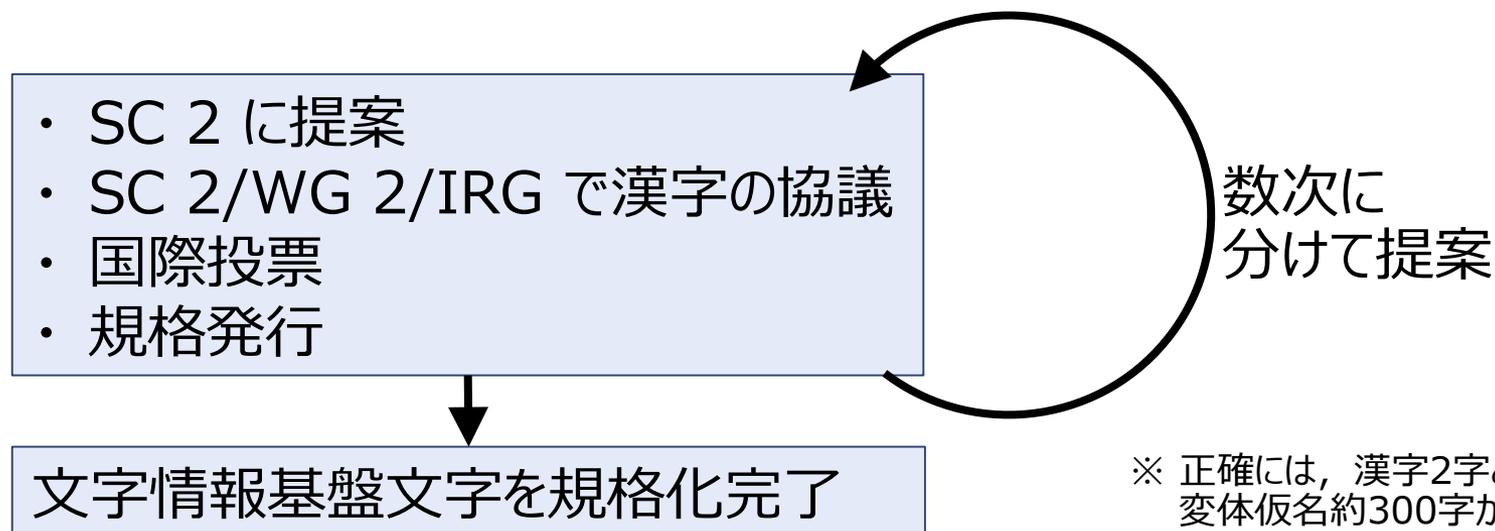


※ 各委員会, WGの設置主体は経済産業省, 事務局は IPA
 ※ 各WGの下には, 具体的な検討を行うSWGを設置

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■文字情報基盤事業とは・・・

- 2017年12月20日付で、国際標準で58,840字が利用可能に
既存のISO/IEC 10646のコードにない文字について、



2017-06-20 Unicode 10.0.0 発行

2017-12-14 Unicode コンソーシアムにIVD登録

2017-12-20 ISO/IEC 10646:2017 発行 (Unicode 10.0.0相当)

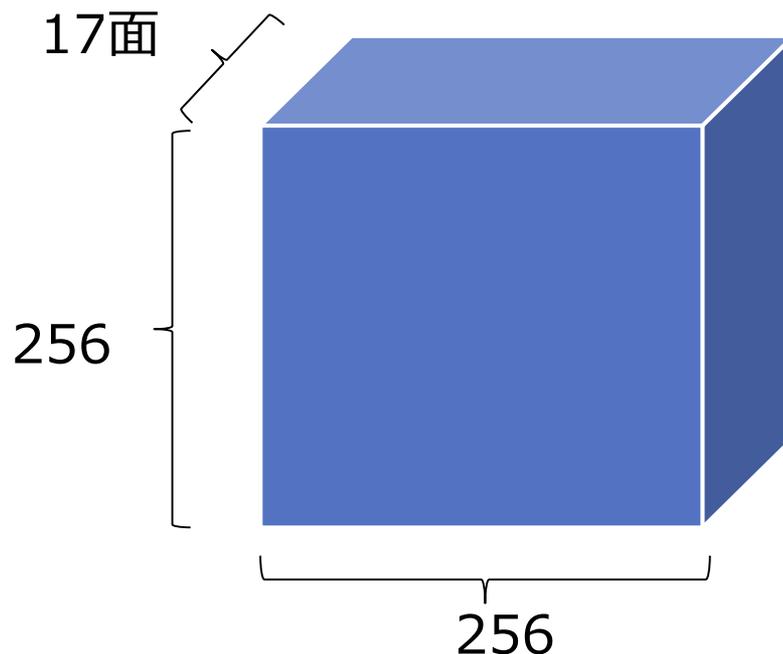
5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■文字情報基盤事業とは・・・

●ユニコード (ISO/IEC 10646) の全体像

21ビットの空間に文字を定義

- 全てが2バイトコード
(と2バイトコードの組み合わせ)
- 一つの面に65536の
コードポイント
(16ビットで表現)
- 0面～16面の合計17面を使う
(21ビット)
- 0面に定義された文字しか
使わないなら,
2バイト固定長と同等



5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■文字情報基盤事業とは・・・

- ユニコードの代表的な符号化方式

- UTF-16

- 0面

- 2バイトの固定長

- 1面～16面

- 4バイトの固定長（2つのコード位置を組み合わせる）

〔2つのコード位置の値のビット列を組み合わせて
1面～16面の21ビットのコード位置を表現〕

- UTF-8

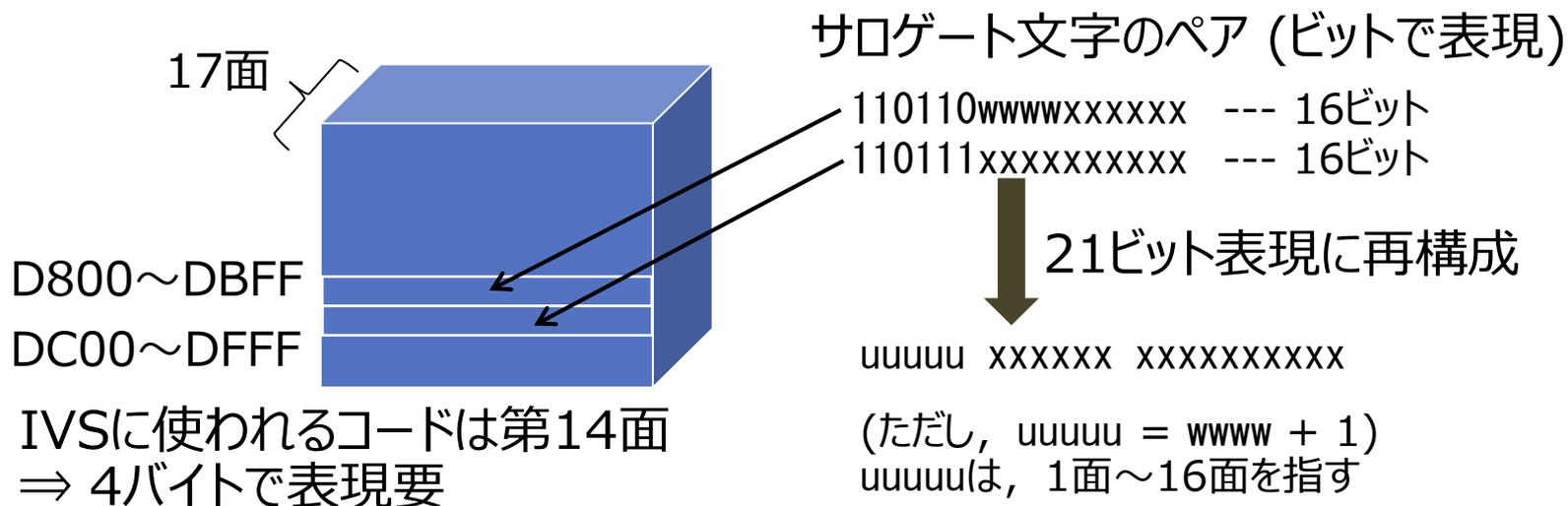
- 0面～16面

- 1～4バイトの可変長で表現（若いコード位置ほど短い）
(説明略)

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■文字情報基盤事業とは・・・

- ユニコードの符号化方式 UTF-16
 - 0面の文字は, 2バイトで表現
 - 1面～16面の文字位置を, サロゲート文字のペアで表現 (2バイト+2バイト=4バイト)
 - 0面のD800～DBFFの領域と, DC00～DFFFの領域を特別に確保
 - 1面～16面の文字の表現
0面のD800～DBFFから1つ, DC00～DFFFから1つを拾い, 両者を合わせて1面～16面の21ビットのコード位置を表現



5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■文字情報基盤事業とは・・・

- ユニコードの代表的な符号化方式の整理
 - ・包摂（統合）した文字の元の形を区別するには、IVS (Ideographic Variation Selector/Sequence) を指定する
 - ・IVS は、第14面の240個のコード位置（E0100～E01EF）として定義されている（UTF-16では、サロゲートペアの4バイトで表現）

符号化方式	ベースの文字の 符号位置	必要バイト数		
		ベースの文字	IVSの指定	合計
UTF-16	0000～FFFF	2	4	6
	10000～10FFFF	4	4	8
UTF-8	0000～007F	1	4	5
	0080～07FF	2	4	6
	0800～FFFF	3	4	7
	10000～10FFFF	4	4	8

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■文字情報基盤事業とは・・・

- IVSを加えたUTF-16表現の具体例

北

5317_E0101

UTF-16表現 ⇒ 5317 DB40 DD01

北

5317_E0102

5317 DB40 DD02

雙

29031_E0100

UTF-16表現 ⇒ D864 DC31 DB40 DD00

雙

29031_E0101

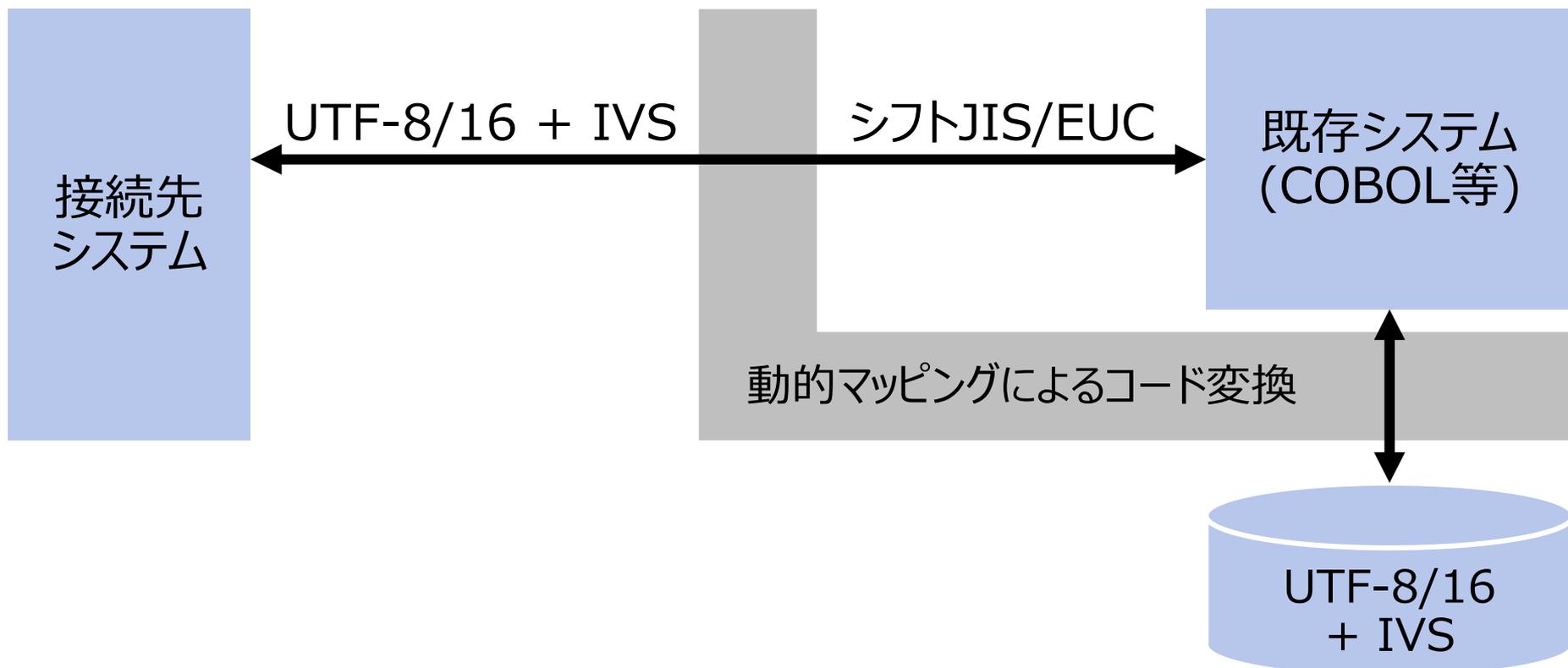
D864 DC31 DB40 DD01

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■文字情報基盤事業とは・・・

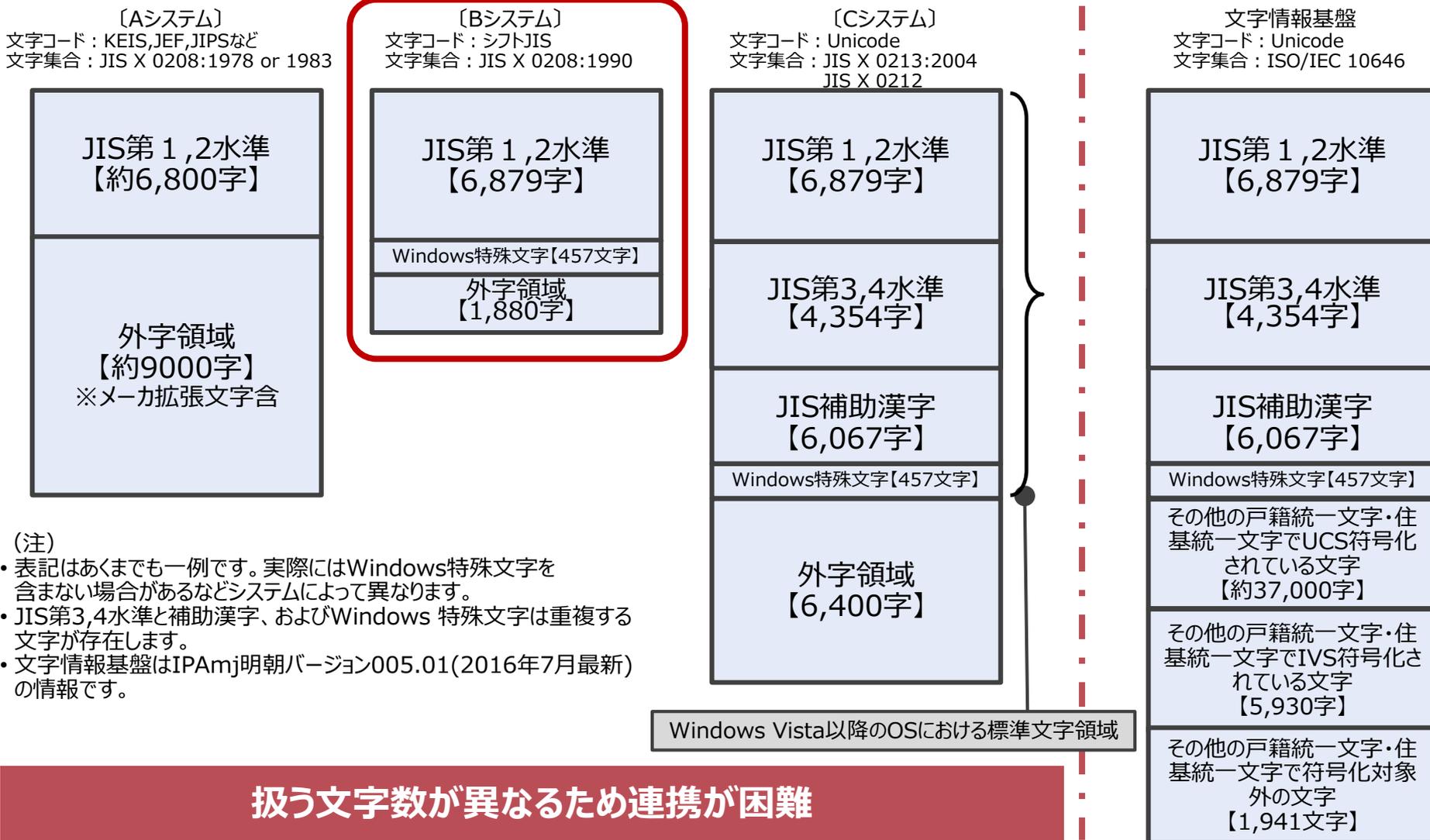
- ITモダナイゼーション

- ・既存システム（シフトJIS/EUCを処理）への変更を最小限にITモダナイゼーションして、文字情報基盤の6万字に対応可能に



5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■システム環境毎の文字環境の違い



(注)

- 表記はあくまでも一例です。実際にはWindows特殊文字を含まない場合があるなどシステムによって異なります。
- JIS第3,4水準と補助漢字、およびWindows 特殊文字は重複する文字が存在します。
- 文字情報基盤はIPAmj明朝バージョン005.01(2016年7月最新)の情報です。

扱う文字数が異なるため連携が困難

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■既存COBOLシステムをUnicodeデータに対応させるには・・・

シフトJIS、KEISなど：文字サイズは固定



UTF-8、UTF-16など：文字サイズは可変

シフトJISとUTF-8, UTF-16の文字サイズ（バイト数）

文字種／文字コード	シフトJIS	UTF-8	UTF-16
英数字（ASCII）	1	1	2
半角カナ	1	3	2
全角記号	2	3	2
全角カナ・かな	2	3	2
漢字	2	3～4	2, 4

- ◆ 日本語のバイト数が増加する
- ◆ 半角1バイト、全角2バイトでなくなる
- ◆ コードが変わる

影響大

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■既存COBOLシステムをUnicodeデータに対応させるには・・・

- 日本語のバイト数が増加する

例： 高橋 4バイト

シフトJISコード例

8D82

8BB4

UTF-16コード例

9AD8

6A4B

高橋 6バイト

UTF-16コード例
(IVS無しの場合)

9AD8

D86D

DF8E

⇒ **領域サイズの見直し**

- 半角 1 バイト、全角2バイトでなくなる

⇒ **文字列処理の見直し**

- ・文字数のカウント
- ・文字列の比較
- ・文字列のコピー etc.

- コードが変わる

⇒ **日本語文字コード判定方式の見直し**

**単純な修正では不可能！
全般的に再設計が必要**

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■既存COBOLシステムをUnicodeデータに対応させるには・・・

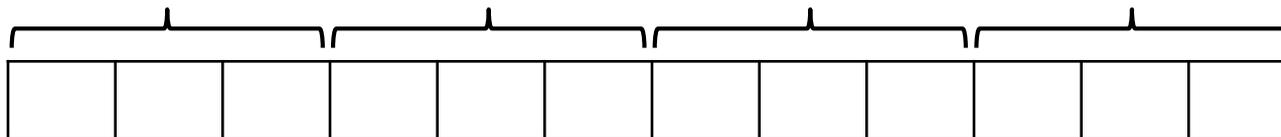
- シフトJISコード → 異体字セレクタを含むUTF-16に変更する場合

01 KANJI-AREA PIC N(3).

(文字位置1つ = 2バイト)



01 KANJI-AREA PIC N(12).

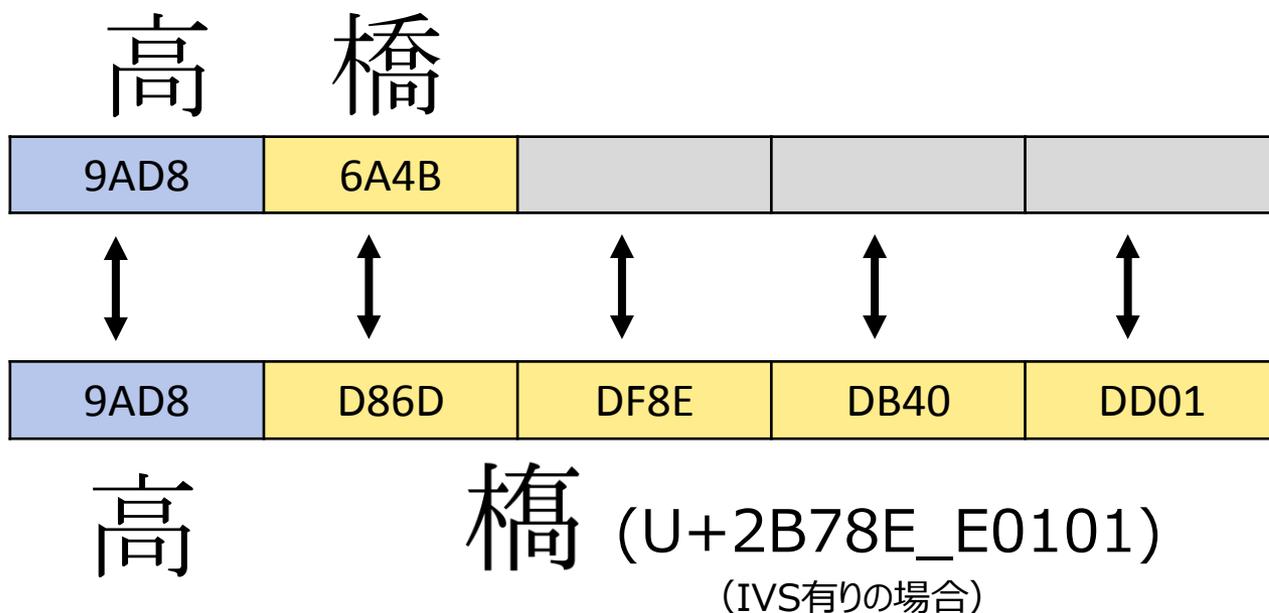


- 文字データの変数を, **4倍の領域に拡張**する必要がある
- こうした変数を含む**レコード内のレイアウトが変更**になり
レコード設計からやり直す必要がある

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■ 既存COBOLシステムをUnicodeデータに対応させるには・・・

- COBOLでは文字列の文字位置同士を順番に比較



- 「文字」同士を比較するには、データ内容を見て、
各「文字」の開始・終了位置を認識する処理の追加が必要

5. COBOL資産を 国際標準の文字情報基盤に対応させるための課題

■ 既存COBOLシステムをUnicodeデータに対応させるには・・・

- 文字列の部分抽出には追加処理が必要

シフトJISコードの
文字列用のコーディング

```
01 K-AREA PIC N(3).
```

⋮

```
MOVE K-AREA(2:1) TO W-AREA
```

2番目の
文字位置から

1文字位置分
を取得

異体字セクタ付きUTF-16の
文字列用のコーディングへ変更

```
01 K-AREA PIC N(12).
```

⋮

```
MOVE K-AREA(k : l) TO W-AREA
```

k 番目の
文字位置から

l 文字位置分
を取得

追加処理

文字列の内容を調べ、
文字位置*k*と文字位置数*l*を求める

- ・ シフトJISでは、文字位置を使って抽出する文字列部分を指定
- ・ 異体字セクタ付きのUTF-16では、**内容を調べて、**
「開始文字位置」と「文字位置の数」を求める処理を追加

6. 既存COBOL資産を有効活用した事例紹介 ～国際標準の文字情報基盤に対応するために～

6. 既存COBOL資産を有効活用した事例紹介 ～国際標準の文字情報基盤に対応するために～

■ 既存COBOLシステムをUnicodeデータに対応させるには・・・

- 日本語のバイト数が増加する

例： 高橋 4バイト

シフトJISコード例

8D82

8BB4

UTF-16コード例

9AD8

6A4B

高橋 6バイト

UTF-16コード例
(IVS無しの場合)

9AD8

D86D

DF8E

⇒ **領域サイズの見直し**

- 半角 1 バイト、全角2バイトでなくなる

⇒ **文字列処理の見直し**

- ・文字数のカウント
- ・文字列の比較
- ・文字列のコピー etc.

- コードが変わる

⇒ **日本語文字コード判定方式の見直し**

**単純な修正では不可能！
全般的に再設計が必要**

スクラッチで作り直し ⇒ ×

既存資産の有効活用 ⇒ ◎

6. 既存COBOL資産を有効活用した事例紹介 ～国際標準の文字情報基盤に対応するために～

■文字に関するシステム仕様

- 文字コードの体系
- エンコーディング法、
- 文字セット(コレクション)

の3つを忘れずに指定することが大事。

#ときどき、ませこぜになっています。

・ 取り扱う日本語文字集合の範囲 : JIS X 0213:2012

・ 文字の符号化方式 : UTF-8 (※)

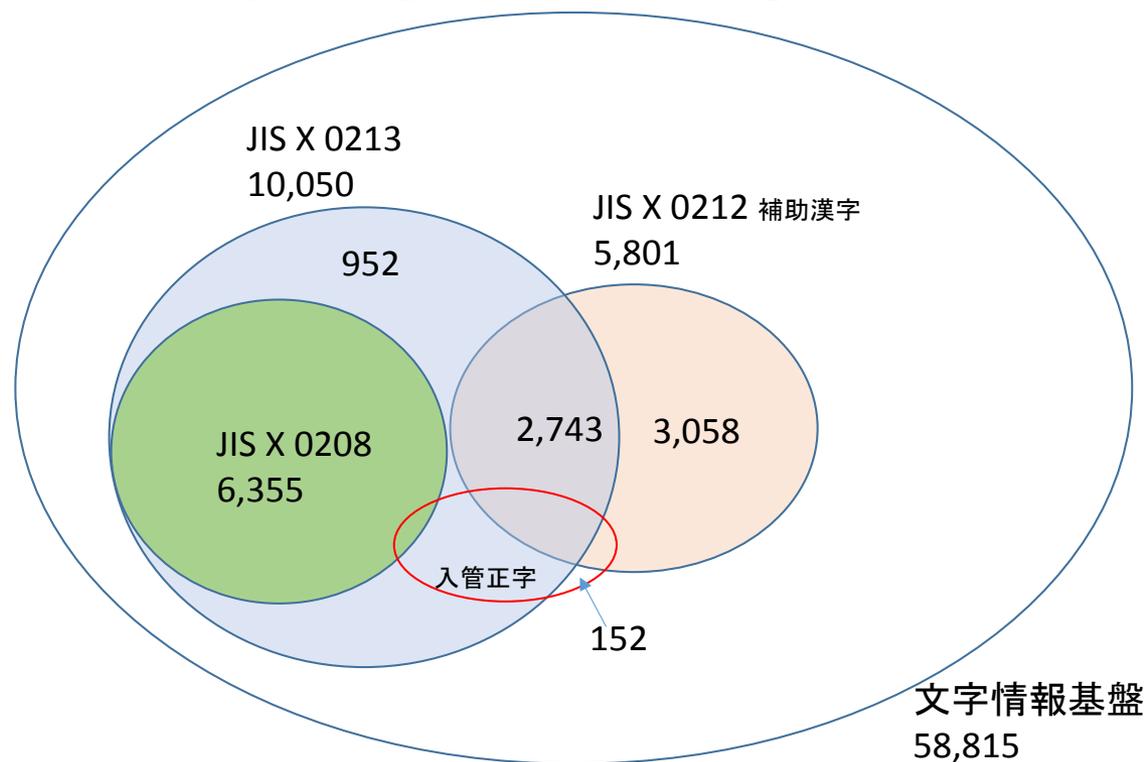
※JIS X 0221:2007 (ISO/IEC 10646 (UCS)) のUCS-4の範囲を符号化

なお、取り扱う文字集合の範囲は、将来的な変化があった場合には、各機関のシステムの対応がなされることを前提として、文字の符号化方式を従前のままとしつつ、より広い範囲の文字集合（文字情報基盤等）に変更していくことが考えられる。この場合においても、柔軟に対応できるよう、拡張性を考慮したシステム設計とすること。

6. 既存COBOL資産を有効活用した事例紹介 ～国際標準の文字情報基盤に対応するために～

■文字を受け入れるときの注意

- どんなコードで来るか？
- どんなエンコーディング法で来るか？
- どんな文字セットが来るか？



ちまたにはいろいろな文字セットがある

6. 既存COBOL資産を有効活用した事例紹介 ～国際標準の文字情報基盤に対応するために～

■文字を受け入れるときの注意

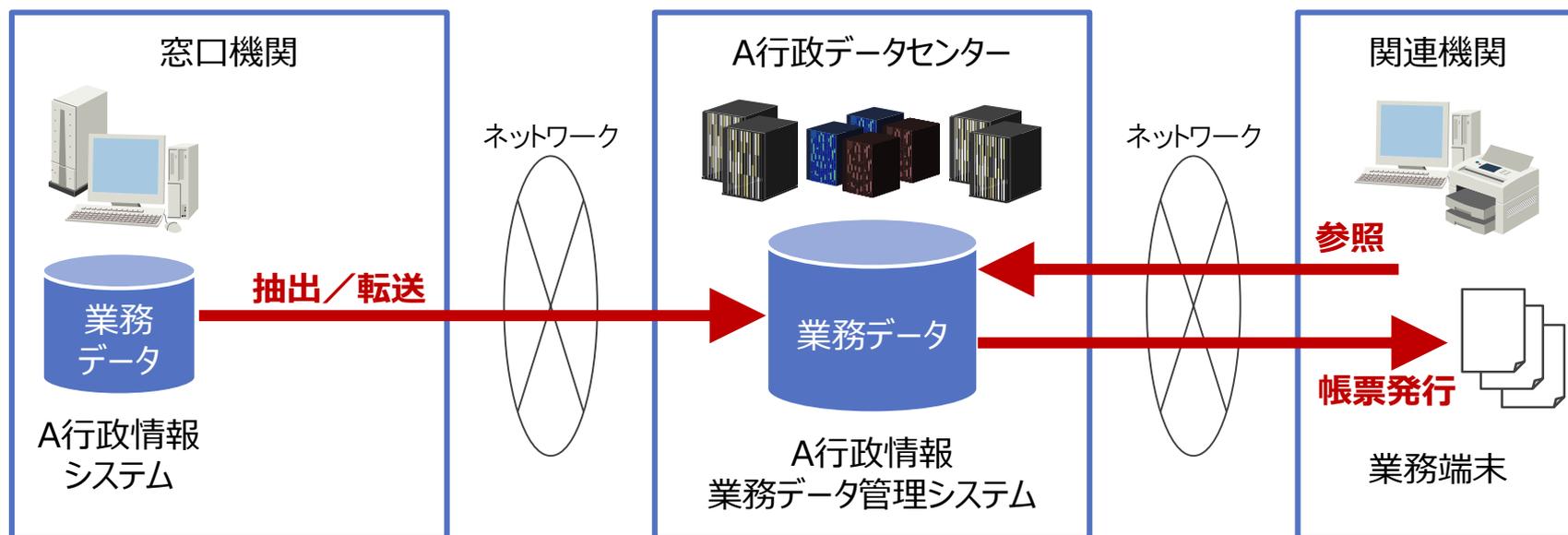
- コード、エンコードの正規化をする
- 決めた範囲を超えた文字を
入れられないようにする（丁寧に謝る）
- 一般向けに送信する文字範囲はJISの
第一～第四水準までとする(電子政府構築の方針)
- 範囲を超えた文字を入力しようとした際に、
その代替案を(親切に)ガイドする(検索、縮退)

「そのシステムで使うことにした文字のセットを明確にし、そのセットから外れた文字を受け取ったときには、警告を出してそれを拒絶するなどの機能を埋め込むことが求められる」

6. 既存COBOL資産を有効活用した事例紹介 ～国際標準の文字情報基盤に対応するために～

■A行政分野ユーザ様の事例

【課題】短期間で、既存資産を有効利用して開発コスト抑制、
業務プログラム・業務データ品質の確保



事例の 特徴

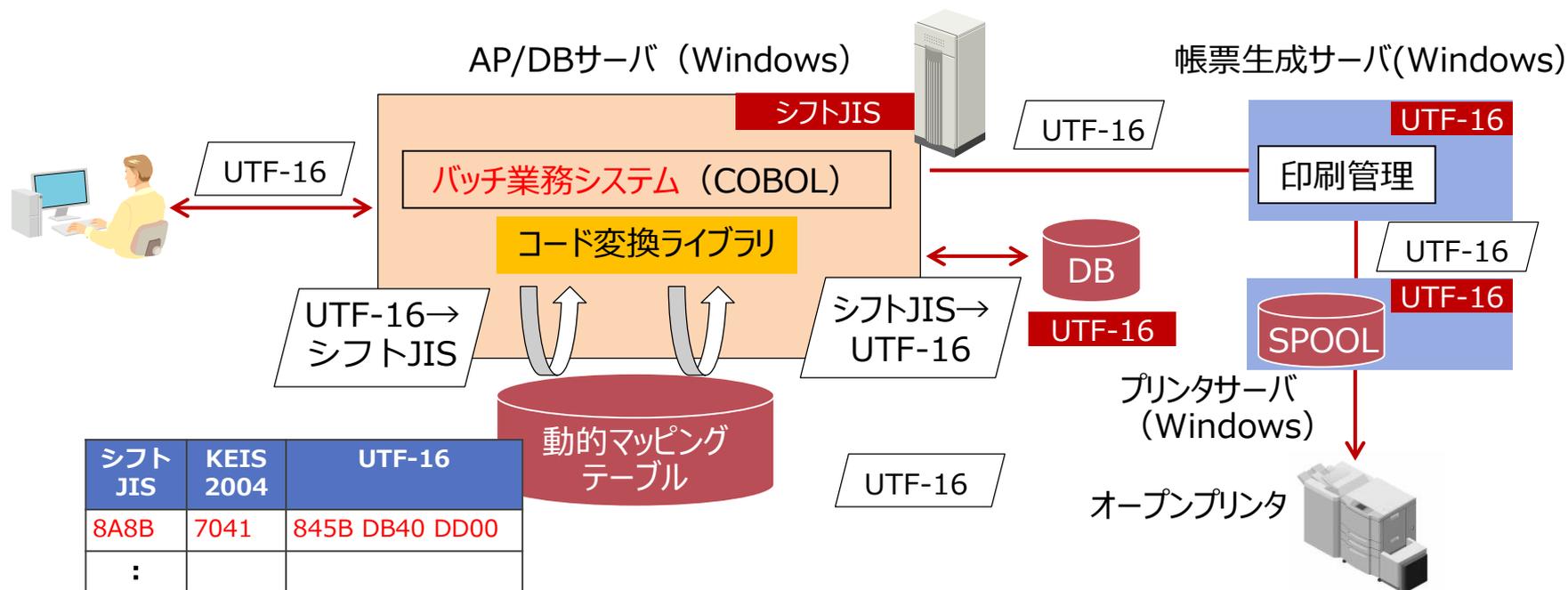
- 既存COBOL資産を有効利用し、Unicode(IVS)対応を実現
- 既存COBOL資産への改修を最大限抑制する動的マッピング採用
(COBOL資産：約900本、250Kstep)

全体6.0ヶ月

6. 既存COBOL資産を有効活用した事例紹介 ～国際標準の文字情報基盤に対応するために～

■ 既存COBOLシステムをUnicodeデータに対応させるには・・・

- バイト数を意識するCOBOL資産は、IVSのマルチバイト対応には多大な工数を費やす。
- レガシー資産を有効活用するため、従来のシフトJISをそのまま使用。



バッチ業務で扱う文字コードを変更せずに対応

7. まとめ

Society5.0
基盤構築の加速

政策動向の動き

実現に向けた
府省などの取り組み

データ連携のための
相互運用性の確保

データ等の標準化

データ流通基盤

IMI

共通語彙基盤

文字情報基盤

UTF-8/16 + IVS

シフトJIS/EUC

既存システム
(COBOL等)

接続先
システム

動的マッピングによるコード変換

UTF-8/16
+ IVS

ITモダナイゼーション
ソリューション

モダナイゼーション技術についての問い合わせ

■ 株式会社 日立製作所 サービスプラットフォーム事業本部 IoT・クラウドサービス事業部

事業推進本部 ミドルウェアソリューション窓口

E-mail: midsol-p@itg.hitachi.co.jp

関連するURL

■ ITモダナイゼーション ソリューション

http://www.hitachi.co.jp/mid-sol/search_s/modernization/

■ COBOL2002

<http://www.hitachi.co.jp/soft/cobol/>

ご清聴ありがとうございました

ITモダナイゼーションSummit 2018 ～どうする？最後のレガシー～

2018/4/13

株式会社 日立製作所
公共システム事業部 全国公共システム第一本部

河合 孝志

- UCS-2 : UCS-2とは、国際符号化文字集合(Universal Coded Character Set)の部分集合のこと。
符号化文字集合の国際規格の1つであるISO/IEC 10646として定められており、2バイトの値として定義されている。
- UTF-8 : UCS/Unicodeを8ビット系システムで扱えるよう、8ビット×1～6バイトに変換するエンコード方式。(ASCIIと同じ文字セット部分は1バイト、それ以外は2～6バイト)
- UTF-16: UCS/Unicodeにおいて、1文字を16ビット一つまたは二つで表現するエンコード方式、16ビットを二つ使うことで16面拡張する。
(サロゲート以外の文字はUCS-2の2バイトと同じ16ビット、拡張16面の文字はサロゲートペアの32ビットで表現)
- EUC : UNIX上で日本語の文字を扱う場合にもっとも多く利用されている文字コード(符号化方式)のひとつである。
- JEF : JEF漢字コード(Japanese processing Extended Feature)は、1979年に富士通が策定した文字コード。
- JIPS : JIPS(Japanese Information Processing System)はNECが開発した日本語処理システムの名前であり、システム上で使われる漢字コード。
- KEIS : 「Kanji processing Extended Information System」は日立製作所が開発した日本語処理システムの名前であり、システム上で使われる漢字コード。

- HITACHIは、株式会社 日立製作所の商標または登録商標です。
- Microsoft、Visual Basic、Windows、Windows Server、およびWindows Vistaは、米国Microsoft Corporationの米国およびその他の国における登録商標または商標です。
- OracleとJavaは、Oracle Corporation 及びその子会社、関連会社の米国及びその他の国における登録商標です。
- その他記載の会社名、製品名は、それぞれの会社の商標または登録商標です。

製品の改良により予告なく記載されている仕様が変更になることがあります。

HITACHI
Inspire the Next 