

令和4年度

教育関連データのデータ連携の実現に向けた実証調査研究  
(教育における広域なデジタルコンテンツの  
利活用環境整備と連携)

<事業者実証レポート>

外字チェッカー

令和5年3月31日

エヌ・ティ・ティラーニングシステムズ株式会社

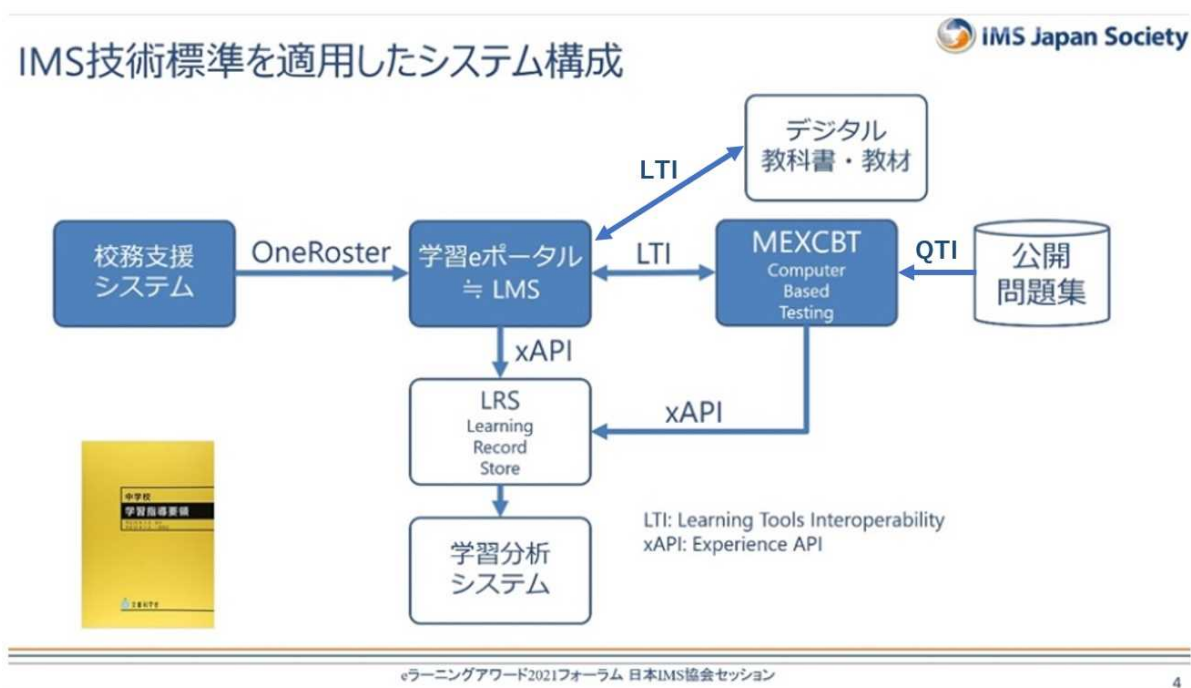
# 目次

<b>1 本事業における役割と具体的な役割</b>	<b>1</b>
1.1 校務支援システムと学習 E ポータルの接続	1
1.2 ONEROSTER JAPAN PROFILE	1
1.3 外字チェッカーの仕様	2
1. 概要	2
2. 入力画面	2
3. 出力画面	2
4. ログ収集	2
5. 環境	3
1.4 画面例（暫定版）	3
1.5 サンプルデータでの外字チェック	4
1.6 コンサル業務	4
<b>2 役務を遂行する過程で明らかになった成果と課題</b>	<b>5</b>
2.1 校務支援システムの現状把握	5
2.2 UNICODE 内の漢字の扱い	5
2.3 ガバメント・クラウドとの関係	5
2.4 ガバクラでの文字セット	6
2.5 MJ+ 問題	6
<b>3 次年度事業への提言</b>	<b>8</b>
3.1 フィールド単位の細かなチェック	8
3.2 API の公開	8
3.3 教育委員会に対するサポート	8
3.4 保護者からの JIS X 0213 文字、申請サイト	9
<b>付録：コンピュータで使える漢字の種類について</b>	<b>11</b>

# 1 本事業における役割と具体的な役務

## 1.1 校務支援システムと学習 e ポータルの接続

下図の通り、様々なシステムが存在する既存の校務支援システムから OneRoster 経由で学習 e ポータルに提供される CSV（カンマ区切り）形式データについて、本年度の外字チェッカー（プロトタイプ版）では、デジタル庁が定める「JIS X 0213 のみ」「外字ナシ」のデータになっているかのチェックを行う。



出典：日本 IMS 協会セミナー資料

## 1.2 OneRoster Japan Profile

以下で規定されている「OneRoster CSV 項目定義書 v1.2」のデータを受け取り、外字の有無をチェックする。各フィールドのチェックは、今年度は行わない。

<https://www.imsjapan.org/post/oneroster-japan-profile>

## 1.3 外字チェッカーの仕様

### 1. 概要

このウェブアプリケーションは、単一の CSV ファイルを入力として、入力ファイルの文字セットが JIS X 0213 に適合しているかどうかを判定し、結果を画面に表示するものである。

入力ファイルに JIS X 0213 外の文字（以下、外字という。）が含まれていた場合、その行番号と文字を表示する。

### 2. 入力画面

ファイルドロップ領域にチェックするファイルをドラッグ&ドロップ、もしくは「ファイルを選択する」ボタンをクリックしてファイル選択ダイアログを開き、ファイルをサーバにアップロードする。

入力ファイルのフォーマットは CSV とする。

プロトタイプ版ではフィールド定義はチェックしない。（校務支援システムの形式に適合するか否かは不問とし、あらゆる CSV ファイルを対象とする。）

入力ファイルの文字エンコーディング方式は、UTF-8（BOM なし）とする。

ただし、UTF-8（BOM あり）、UTF-16（BOM あり／BOM なし）、Shift\_JIS のいずれかと判定された場合、許容する。

### 3. 出力画面

サーバーサイドで外字検出処理を行い、以下の情報を表示する。

- 処理日時
- ファイル名
- 文字エンコード

（UTF-8（BOM なし）以外の許容する文字コードであった場合、アラートを表示する。）

- 検出された外字の情報（表形式で表示する。以下を含む。）
  - 外字が含まれる行の行番号
  - 行に含まれる外字（カンマ区切り）
  - 行に含まれる外字の符号位置（U+XXXX 形式、カンマ区切り）

検出された外字にユーザ定義外字（Windows の外字エディタで作成した文字はこれに含まれる。）が含まれる場合も、そのまま表示する。多くの環境では、「□」などのように表示される。

入力ファイルが CSV 以外のファイル形式であった場合、または UTF-8 もしくは許容しない文字エンコードであった場合は、エラー表示を行う。

なお、プロトタイプ版の外字検出処理では、JIS X 0201 のラテン文字の範囲に限って、いわゆる半角文字と全角文字をどちらも適合と判定し、片仮名については全角文字のみを適合と判定する。

### 4. ログ収集

次年度以降の事業に役立てるため、実行ログを収集する。収集する情報は以下のとおりである。

- ID（連番）

- IP アドレス
- 処理日時
- ファイル名
- 文字エンコード
- 検出された外字の情報

個人情報を含むため、入力 CSV ファイル自体の保持は行わない。外字検出処理後、入力 CSV ファイルは破棄する。

## 5. 環境

- 言語: C# 10.0
- フレームワーク: .NET 6, ASP.NET Core
- データベース: Azure SQL Database
- ホスティング: Azure Web Apps

## 1.4 画面例（暫定版）

GaijiChecker Home Privacy チェッカー

### 外字チェッカー

ここにファイルをドロップしてください  
または

**ファイルを選択する**

処理日時 2023/03/08 10:31:39 +09:00  
ファイル名 gaiji.txt  
文字コード utf-8

エラー	あり	
2	高	U+9AD9
4	伏	U+3438
5	儼	U+34A5, U+34A0
7	撈	U+22850
9	시작이반이다	U+C2DC, U+C791, U+C774, U+BC1B, U+C774, U+B2E4

© 2023 - GaijiChecker - [Privacy](#)

※プロトタイプ版は、1年間のサーバ運営を行う。

## 1.5 サンプルデータでの外字チェック

「教育関連データのデータ連携の実現に向けた実証調査研究」に参加されている 9 社より OneRoster 用のサンプルデータをお預かりし、外字チェッカーに入れたが、サンプルのため外字エラーは 9 ファイルともなかった。

姓名の漢字は JIS X 0213 ではなく、9 ファイルとも X 0208 内に収まっており、シフト JIS + 外字での校務システムが多いと予想されることが確認できた。

## 1.6 コンサル業務

- ・ 要求仕様の整理
  - ・ 問題の解決策についてのヒアリングやシステムの分析 など
- 次年度の本格開発に向けた準備を行う。

## 2 役務を遂行する過程で明らかになった成果と課題

### 2.1 校務支援システムの現状把握

OneRoster からデータを受け取る学習 e ポータルは、新規のクラウド型システム開発で、以下の仕様となっている。

- ・ 取り扱う日本語文字集合の範囲：JIS X 0213
- ・ 文字コード：ISO/IEC 10646
- ・ 文字の符号化形式：UTF-8

しかし、校務支援システムは 1980 年代に開発されたものもあり、当時の主要な文字コードはシフト JIS で、文字セットは JIS X 0208、つまり第一水準漢字、第二水準漢字の 6355 文字に外字を加えたものが多いと思われる。この外字は学習 e ポータルなど 21 世紀に新規開発されたシステムでは扱えない。

外字チェッカーはこの現状を把握するためにも有用である。

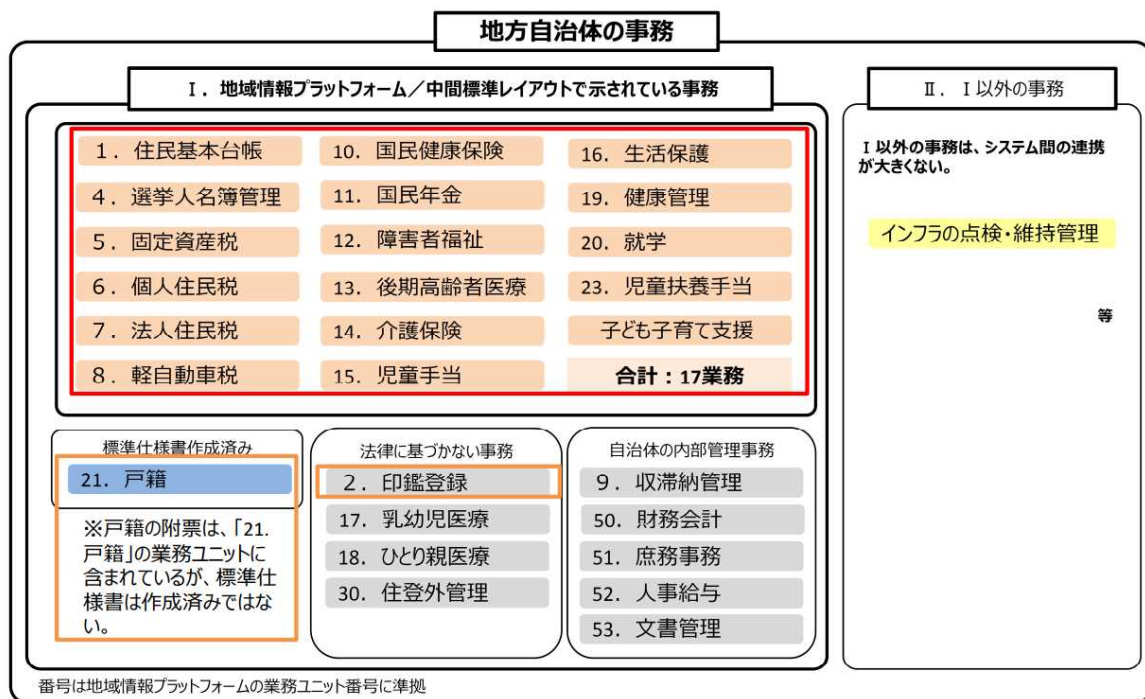
### 2.2 Unicode 内の漢字の扱い

X 0213 にはないが、Unicode には存在する日本の文字（はしご高など）の扱い

「高」、「羽」、「吉」、「閒」、「強」などについては、デジタル庁の方針が出ればそれに従い確認する。（現在の仕様ではエラーとしている）

### 2.3 ガバメント・クラウドとの関係

現在、総務省、デジタル庁が中心となって、省庁や地方自治体が利用する基幹業務のクラウドサーバー化（以下「ガバクラ」と略す）が推進されている。地方自治体では以下の 17 業務がガバクラに搭載される。



出典：デジタル庁資料「地方公共団体の基幹業務システムの標準化のために検討すべき点について（令和3年9月）」

人名については、法務省の戸籍システムから、ガバクラの住民基本台帳（住基）システムにデータ送られ、これが就学（学齢簿）にも届く。

文部科学省 教育データ標準で留意すべきは、このガバクラは2025年（令和7年）を目標に開発されており、**教育データ標準とほぼ並行して開発が進められている点**である。

## 2.4 ガバクラでの文字セット

ガバクラでの文字セットは、戸籍と住基は6万文字の漢字が扱える文字情報基盤（フォント名はMJ明朝）を使用し、それ以外のシステムでは、文字情報基盤の6万文字を、X 0213の1万文字の漢字に「縮退」させる仕様となっている。

MJ縮退マップは一般社団法人 文字情報技術促進協議会が以下で公開している。

<https://moji.or.jp/mojikiban/map/>

## 2.5 MJ+ 問題

デジタル庁では2022年8月に「[地方公共団体情報システム データ要件・連携要件 標準仕様書【第1.0版】](#)」がデジタル庁から公開され、13-14頁に「戸籍・住基以外はX0213」という明快な指針が出た。



ところが、2022 年 12 月に「[文字要件に係る今後の検討の方針について](#)」という 4 頁の資料がデジタル庁から公開され、そこには「文字情報基盤の 6 万字に新たに 5 万文字を追加することを年度内に検討する」とある。

その後、2023 年 2 月 20 日に「[地方公共団体情報システムにおける文字要件の運用に関する検討会（第 1 回）](#)」が開催され、5 万文字が X 万文字に数を減らし実施する案が提示された。

ガバクラでの MJ と MJ+問題、そして MJ の縮退マップや MJ+での代替マップなど、流動的な部分もあり、文字セットについては、今後の推移を見守りたい。

## 3 次年度事業への提言

### 3.1 フィールド単位の細かなチェック

今年度は CSV の各行について、JIS X 0213 の範囲内かどうかのチェックを行うが、次年度には「OneRoster CSV 項目定義書」の各フィールド仕様に沿って、漢字、カタカナ、数値などのチェックも有用と思われる。特に日本語固有の数字、カタカタ、記号類についての半角・全角問題もあり、細かなチェックが必要となる。特に半角カナは使用禁止なのでそのチェックは重要である。

### 3.2 API の公開

今年度は UI（ユーザインタフェース）のみの公開で、特定のサイトにユーザが CSV ファイルを投げ入れる仕様となっている。

校務支援システム側で、OneRoster に入れる前処理を行い、OneRoster CSV 項目定義書に沿った形式にデータを揃える際に、外字チェッカーが API（アプリケーション・プログラム・インタフェース）を持ち、データ確認が行えれば、作業効率が向上する。

次年度には API の公開も検討したい。これに伴い校務支援システムとのシステム間認証の実装も検討する。

### 3.3 教育委員会に対するサポート

次年度、本件予算を使うかどうかは別として、オンラインでの個別相談にも対応する体制を、以下の通り、何らかの形で整えた方が良いと考える。

外字チェッカーではじかれた文字を JIS X 0213 に変更する手順については、三重県桑名市の事例が、『教育と ICT』オンライン「多くの自治体が直面する「人名外字問題」の対応が急務【後編】桑名市は外字全廃で教員の負担を軽減」（2022 年 12 月 6 月）が公開されており、これが参考になる。

<https://project.nikkeibp.co.jp/pc/atcl/19/06/21/00003/120500408/>

ここでは、「教育長名で保護者に出された通知文書」と「代替する文字の希望を保護者に聞いた資料」（以下の票）が公開されている。

しかし、これだけではカバーしきれない、個別の問題が多くの自治体で発生する可能性がある。

**学校記入** 組 齊藤 太郎 さん 保護者様

①学年・組 を記入する

校務電算化に伴う名前

②学齢簿・入学通知書・要録様式1などに表記されている字が、校務用PC上で表示されない児童生徒名を記入する

③該当児童生徒名が表示される

今後、学校で作成する各文書において、統一した記表記とさせていただきます。ご確認いただき署名線に希望の文字を記入してください。他の類似文字での表記を希望される場合は、使用可能な文字の範囲をご確認の上、希望表記欄にご記入ください。ご不明な点等ございましたら、下記担当までご連絡ください。

【使用可能な文字の範囲】  
使用可能な文字は、特別な設定（外字フォントの適用 等）をしていない、ご自宅のパソコン（WindowsならばMS明朝、ゴシック）やお持ちのスマートフォンで表示可能なものとなります。それ以外の文字は使用不可のため、類似の文字にて代替することとなります。

校務電算化での表記

齊藤 太郎

希望

齋藤 太郎

④希望表記欄に記入があった場合、校務PC上で表示できるか確認し、該当保護者へ可否を連絡。表示可であればCath取り込みフォーマットに反映させる

上記の名前表記について確認し、了解しました。

令和2年 1月

⑤新小学1年生については「教育委員会宛 返信封筒」にて提出してもらう。各校には教育委員会より転送

保護者名

1月16日までに担任にご提出ください

⑥各校で必要に応じて提出締め切りを記入する

事務担当  
桑名市教育委員会事務局 学校支援課  
教育指導係  
電話 0594-XX-XXXX

出典：桑名市教育委員会

### 3.4 保護者からの JIS X 0213 文字、申請サイト

上の票について、保護者からオンラインで申請してもらった方が、文字が確実に特定でき、書類をやり取りする手間も省ける。次年度、本件予算を使うかどうかは別として、各教育委員会等が共同利用するオンライン申請サイトの開発も有用と考える。一つのアイデアとして提案する。

教育委員会が外字を X0213 内の文字に変更する作業は、後出の「付録：コンピュータで使える漢字の種類について」で解説する通り、包摂・書体デザインなど揺らぎ問題があり、専門的な

知識も必要となる。「教育委員会⇒保護者」ではなく、「保護者⇒教育委員会」という流れの申請サイトを、スマホでも動くウェブ・アプリケーションとして開発することが望まれる。その理由は「付録：コンピュータで使える漢字の種類について」の結論で述べている通り、ほぼすべての保護者がスマホを利用し、自分の姓を外字ナシで入力しているからである。

認証やマイナンバー入力など、デジタルですべてを完結させるシステムは、開発コストが増加するので、入力データの確認を保護者と電話や LINE で行うなど、柔軟なシステムが望まれる。

市町村コードを含んだ URL で市町村を特定し、確認メールが保護者と教育委員会などの管理者に届くものが効果的だと思われる。また、保護者がスマホから入力できるよう、レスポンスな画面デザインで、シンプルな入力項目が望まれる。

次頁以降に、一般的な漢字についての参考資料を置いた。

## 付録：コンピュータで使える漢字の種類について

今、教育現場には多くのパソコンやタブレットが導入されています。そこで使える漢字や文字の種類(以降「文字セット」と呼びます)について解説します。

コンピュータの歴史は、電算時代、パソコン時代、ネット時代に大別できます。

「電算時代」は 1950 年代の黎明期から 1980 年代のパソコン登場まで。電算は電子計算機の略で、大型計算機でコンピュータ処理を行う電算センターというジャンルの会社が登場し、大量の伝票や宛名印刷を担当していました。この時代に使えた文字は、X 0208 という JIS(日本産業規格)の番号で呼ばれる第 1 水準と第 2 水準の漢字 6355 文字でした。住所や氏名などで 6355 文字に含まれない漢字は「外字」を独自に制作し印刷していました。

「パソコン時代」は 1980 年代、個人で買えて使えるパーソナルコンピュータの登場で始まります。Shift JIS という英語と漢字を混在できる方式が広まりましたが、この方式でも使える漢字は 6355 文字のまま。足りない漢字は個々の業務システムで外字を作っていました。パソコンは「半導体の集積度は 2 年で倍になる」というムーアの法則のもと飛躍的に進化し、持ち運べるノートパソコンやタブレット PC まで登場しました。2000 年代には JIS X 0213 という新しい文字標準が普及し始め、第 3 水準、第 4 水準の新たな漢字も加え約 1 万文字が使えるようになりました。

「ネット時代」は 2000 年頃に始まります。インターネットは World Wide Web(世界を覆うもの巣)と表現されるように、世界規模での標準化が必要です。1990 年に設立された文字の世界標準を推進するユニコード・コンソーシアムが脚光を浴び、Unicode(ユニコード)という世界標準文字セットに X 0213 の漢字も組み込まれました。2007 年には iPhone が登場し、2010 年代にはスマートフォン(スマホ)が世界中で使われるようになりました。

ネット時代の現在、様々な行政システムのクラウド化が推進されています。2021 年 9 月にデジタル庁が発足し、総務省とデジタル庁が連携して「ガバメント・クラウド(ガバクラ)」という自治体が利用する様々な業務のクラウド・サービス化もその一つです。ガバクラには住民基本台帳、選挙人名簿、固定資産税、国民年金、障害者福祉、後期高齢者医療、介護保険、児童手当、就学(学齢簿)、児童扶養手当、子ども・子育て支援など、17 業務が搭載されます。これとは別に、法務省で戸籍のクラウド・サービス化が推進されています。

様々な業務システムがインターネットでつながり、連携してサービスを提供する「ネット時代」、電算時代やパソコン時代に作られた「外字」はデータ交換のさまたげになります。この問題の解決に向けて、ガバクラなどのサービスで使用する文字セット(使える漢字の種類)について、2022 年 8 月にデジタル庁から「地方公共団体情報システム データ要件・連携要件 標準仕様書」が発出され、戸籍と住民基本台帳は「文字情報基盤(書体名: MJ 明朝)」、それ以外の業務は「JIS X 0213」との方針が示されました。いずれも都道府県・市町村や関連業者のシステムごとに今でも存在する「外字をなくそう運動」です。

文字情報基盤は、自治体が個別に開発・運用してきた電算時代やパソコン時代の「外字」があると、ネットワークでのスムーズな情報交換が進まないのが、全国各地の個別システムで使われていた外字を整理統合した約 6 万文字の文字セットです。文字情報基盤のフォント「MJ 明朝」は Windows、MacOS で表示できますが、スマホやタブレットの Android、iOS、Chrome OS では表示できません。

スマホやタブレットに「外字」や「フォントをインストールする」という概念はなく、スマホはインターネットへの接続を前提としているので、ネットで常に使える Unicode 互換の JIS X 0213 の約 1 万の漢字が標準となっています。

デジタル庁の評語は「誰一人取り残されないデジタル社会」、文字が読みづらい人は人口の 1 割近いと言われており、デジタルだとコンピュータでの自動読み上げが可能になります。しかし、外字は読めません。「個別最適化された学び」にも X 0213 での標準化は重要です。また、画面上で文字列をコピーし他の文書に貼り付ける「コピー＆ペースト(コピペ)」でも、外字は貼り付けできません。



（出典：文字情報技術促進協議会）

文字関連では様々な用語が使われていますが、簡略にまとめてみました。

#### ■ 一般用語 ■

- ・フォント：コンピュータで扱う文字のこと
- ・ビットマップフォント：16x16 や 24x24 などの点で文字を表す方法
- ・アウトラインフォント：ベクトル曲線で文字を表す方法、拡大しても美しい
- ・JIS：日本産業規格、ネジから自動車まで 1 万種類以上の国家規格
- ・ISO：国際標準化機構、JIS の国際版、国際調達などで使用される
- ・書体(Typeface)：明朝体、ゴシック体など、文字の形をデザインしたもの
- ・スタイル：太字、斜体、下線付きなど文字の修飾。書体として別に存在する場合もある
- ・グリフ：個々の字の形。字の形を表す考え方として字形、字体などもある
- ・環境依存文字：コンピュータシステムに依存する丸数字、半角カタカナなど
- ・ラテン文字：ラテン語などを表記するためのアルファベット、ローマ字
- ・半角/全角：半角：8bit で表現できる文字、全角：16bit 以上で表現できる漢字など
- ・外字：利用システムの文字セットに含まれない文字、旧来は現場で外注していた
- ・縮退：文字を特定の文字セットに対応させ、その範囲に収めること
- ・包摂：文字の複数の字体が相互に区別されずにひとつの符号位置に対応すること

#### ■ 文字セット ■



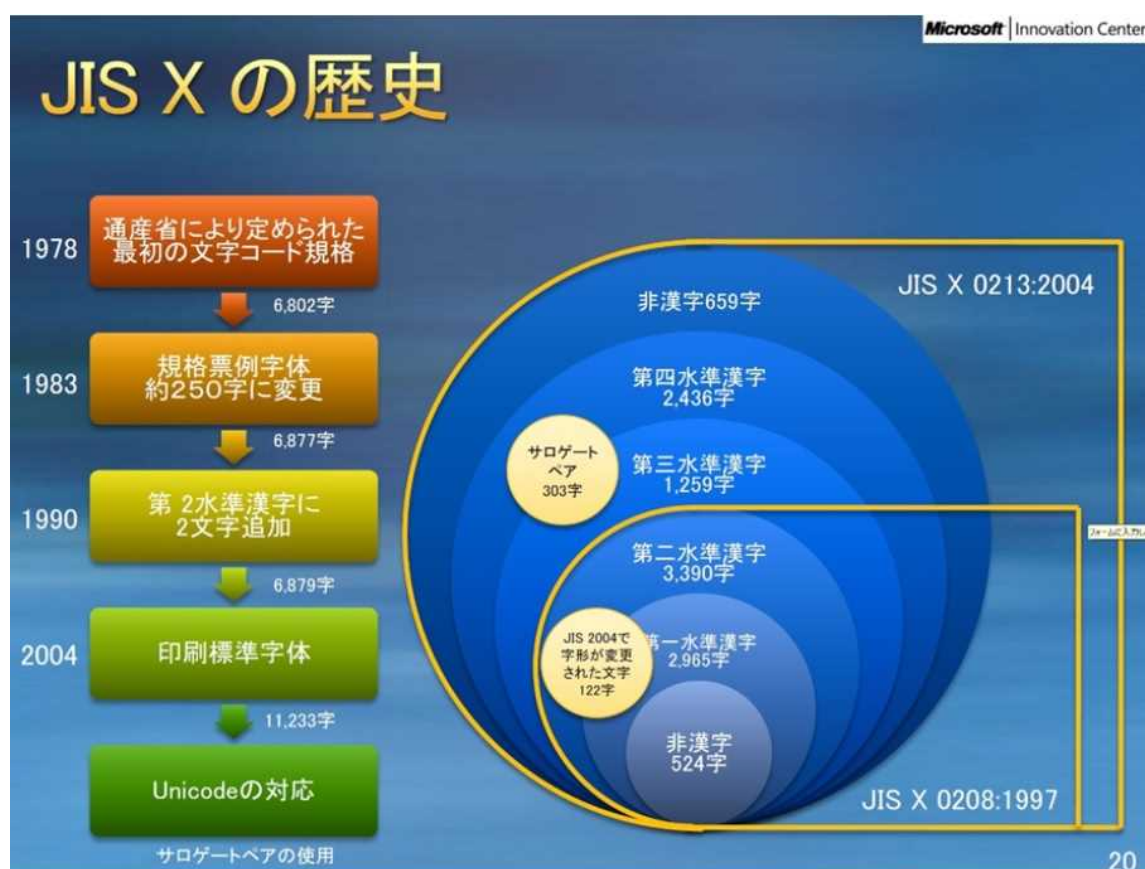
- ・文字セット: JIS X 0208、ISO10646 など、文字集合及び関連規則を定めたもの
- ・JIS X 0201: 後出の ASCII に仮名文字 (俗称: 半角カナ) を加えた 8 ビット文字セット
- ・JIS X 0208: 数字、ひらがな等に加え、6355 文字の第 1、第 2 水準漢字を含む文字セット
- ・JIS X 0213: 0208 に第 3、第 4 水準漢字を追加、全 11233 字の文字セット
- ・Unicode: 国際的な文字標準化団体が定義する文字セット、ネットと共に広まった
- ・常用漢字: 法令、公用文書、新聞などの文章を書き表す漢字の目安、2136 字
- ・人名用漢字: 戸籍に名として記載できる漢字で常用漢字に含まれないもの
- ・非漢字: 平仮名、片仮名、ラテン文字、ギリシャ文字、算用数字、記号など

## ■文字コード■

- ・文字コード: JIS、ISO、Unicode の文字セットを符号化する方式
- ・ASCII: 情報交換用米国標準コード、7 ビットで表現するラテン文字や数字、制御文字
- ・Shift\_JIS: X 0201 を 1 バイト、X 0208 等を 2 バイトで符号化する可変幅符号化方式
- ・UTF-8、UTF-16: Unicode の 8/16 ビットの符号化方式、UTF-8 がネット環境での標準

JIS の文字数などは下図を参照してください。X 0213 には JIS 第 1 水準から第 4 水準漢字までが入っており、この 1 万文字は世界標準の Unicode にも入っていて、Unicode がインターネットの標準文字セットなので、世界中のパソコンやスマホ、タブレットで、X 0213 の漢字やひらがな、カタカナなどが表示できます。

戸籍統一文字の構想がスタートした 20 数年前、コンピュータ(パソコン)の利用者は国民の数パーセントでした。今では、当時のコンピュータの数百倍の能力を持つスマホを誰もが所有し、日々オンラインショッピングなどで氏名を入力しています。コンピュータで使える漢字への国民の理解は格段に深まっています。



(出典: 日本マイクロソフト社 セミナー資料より)