# 文字環境導入実践ガイドブック

* https://cio.go.jp/guides
* https://cio.go.jp/sites/default/files/uploads/documents/1014\_moji\_guidebook.docx
* https://cio.go.jp/sites/default/files/uploads/documents/1014\_moji\_guidebook.pdf
* 2019-03-28
* 内閣官房情報通信技術（ＩＴ）総合戦略室
* 標準ガイドライン群ＩＤ: 1014

## 2. 文字に関する標準的な取扱い

### ア. JIS X 0213の文字

氏名、法人名、地名等の名称における文字は、その個別性、同一性を確認する上で、非常に重要な要素です。一方、前述のとおり、文字の表現の正確性を重視して、取り扱える文字の数量を増加させ、複雑化させることは、人間の能力の制約上、同一の文字であることを確認することに時間を要し、誤認することに繋がってしまい、結果的に、混乱の要因となる場合があります。

例えば、左の文字群は、「ベ」と表音する「邊」の異字体とされています。常用漢字では「辺」と表記されています。これらの文字は似ていますが、全て異なる文字です。これらの文字を、瞬時に、異なる文字と認識し、同一性の確認を行い、特定することは一般に困難だと考えられます。

また、検索する場合において正しく入力しないときには誤った検索結果が得られるおそれもあります。さらに、筆記された文字の場合には、略字を用いる可能性等が加わり、同一性の確認が一層難しくなります。

また、情報システムの構築及び管理の観点からは、一般に普及しているパソコン、スマートフォン、タブレットといったコンピュータで外字を取扱う場合には、個別にコンピュータに当該外字を導入しなければいけません。また、外字を扱う情報システムとの間においてデータ連携をする場合には、同じ外字を導入するか、利用可能な文字に変換しなければならず、効率的ではありません。

以上を踏まえ、サービス・業務を設計し、情報システムを構築及び管理していく上で、文字の取扱いは、日本工業規格[JIS X 0213](用語.md#jis-x-0213)の文字（約1万文字）に限定して使用することを強く推奨します。[JIS X 0213](用語.md#jis-x-0213)は、一般的なパソコンやスマートフォン等に標準搭載されており、当該文字を利用すれば、上述の問題の発生を防ぐことができるからです。

なお、[JIS X 0213](用語.md#jis-x-0213)の利用は、[デジタル・ガバメント実行計画](https://cio.go.jp/digi-gov-actionplan)に[明記](https://github.com/code4fukui/BaseRegistry/blob/main/%E3%83%87%E3%82%B8%E3%82%BF%E3%83%AB%E3%83%BB%E3%82%AC%E3%83%90%E3%83%A1%E3%83%B3%E3%83%88%E5%AE%9F%E8%A1%8C%E8%A8%88%E7%94%BB.md#2-%E6%96%87%E5%AD%97%E7%92%B0%E5%A2%83%E3%81%AE%E6%95%B4%E5%82%99%E6%99%AE%E5%8F%8A%E5%86%85%E9%96%A3%E5%AE%98%E6%88%BF%E7%B5%8C%E6%B8%88%E7%94%A3%E6%A5%AD%E7%9C%81%E6%B3%95%E5%8B%99%E7%9C%81%E8%B2%A1%E5%8B%99%E7%9C%81%E5%85%A8%E5%BA%9C%E7%9C%81)されています。

### 2.1 総論

#### 1) 文字の範囲に関する標準的な取扱い

#### 2) 文字コード及び文字の符号化形式に関する標準的な取扱い

コンピュータで文字を扱うためには、取り扱う文字の範囲を定めるだけでなく、文字をコンピュータが理解できる形式に符号化する必要があります。

符号化に当たっては、文字に与える番号である「文字コード」及びこの番号を文書ファイルの保存や交換の際にどのように表現するかを指定する「符号化形式」が必要です。文字コード及び符号化形式には、それぞれ複数の種類があるため、情報システムを設計・開発していく上で、これらの仕様を示さないと、設計・開発事業者が独自の判断で決めてしまう可能性があります。

このような場合には、情報システムの更新時のデータ移行や他の情報システムとのデータ連携を行う際に、文字コードや符号化形式の差異により、問題が発生するおそれがあります。

このため、情報システムの設計・開発に係る調達においては、文字コードとしては [ISO/IEC 10646](用語.md#isoiec-10646)を、符号化形式としては [UTF-8](用語.md#utf-8)を指定して使用することを推奨します。[UTF-8](用語.md#utf-8)と指定するだけで、[ISO/IEC 10646](用語.md#isoiec-10646)を文字コードとして使用することを併せて意味すると解釈されることもありますが、明確化のため、文字コード及び符号化形式の両方とも指定することが安全です。

[ISO/IEC 10646](用語.md#isoiec-10646)文字コードを[UTF-8](用語.md#utf-8)の符号化形式で表現したものは世界中の文字を加えた国際規格のコードで、コンピュータの基礎的な符号化形式である [ASCII](用語.md#ascii)と互換性があり、世界の多くのソフトウェアがこれに対応しています。[ISO/IEC 10646](用語.md#isoiec-10646)と[UTF-8](用語.md#utf-8)を利用することで、効率的に情報システムを構築及び管理することができます。

調達仕様書では、以下のように記載することで、漢字の仕様を明確にすることができます。

* 取り扱う日本語文字集合の範囲：[JIS X 0213](用語.md#jis-x-0213)
* 文字コード：[ISO/IEC 10646](用語.md#isoiec-10646)
* 文字の符号化形式：[UTF-8](用語.md#utf-8)

なお、[UTF-8](用語.md#utf-8)では、漢字等を１文字当たり３バイトで表現します。既存システムにおいて漢字等を２バイトで表現する符号化方式（[Shift-JIS](#用語.md#shift-jis)等）を使用している場合、[UTF-8](用語.md#utf-8)に符号化方式を移行する際には２バイトから３バイトになることを想定した設計が必要になります。