



1. Home (<https://www.gov.uk/>)
  2. Government (<https://www.gov.uk/government/all>)
  3. Government efficiency, transparency and accountability (<https://www.gov.uk/government/government-efficiency-transparency-and-accountability>)
  4. Open standards for government (<https://www.gov.uk/government/publications/open-standards-for-government>)
- Cabinet Office (<https://www.gov.uk/government/organisations/cabinet-office>)
  - Government Digital Service (<https://www.gov.uk/government/organisations/government-digital-service>)

Guidance

# Encoding characters

Updated 1 July 2020

## Contents

1. Summary of the standard's use for government
2. How this standard meet user needs
3. How to use the standard

Print this page



© Crown copyright 2020

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3) (<https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3>) or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: [psi@nationalarchives.gov.uk](mailto:psi@nationalarchives.gov.uk).

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available at <https://www.gov.uk/government/publications/open-standards-for-government/cross-platform-character-encoding-profile>

Use UTF-8 (<https://tools.ietf.org/html/rfc3629>), an encoding form for Unicode (<http://www.unicode.org/versions/Unicode13.0.0/>) character sets, for government digital services and technology.

## 1. Summary of the standard's use for government

Unicode is based on the ASCII (<https://en.wikipedia.org/wiki/ASCII>) character set, but expands ASCII to include characters for most written languages.

UTF-8:

- is one of the encoding forms for Unicode
- encodes all Unicode characters without changing the ASCII code

This makes UTF-8 flexible for a wide range of uses. For example, the default character encoding in HTML-5 is UTF-8.

The government chooses standards using the open standards approval process (<https://www.gov.uk/government/collections/open-standards-for-government-data-and-technology>) and the Open Standards Board (<https://www.gov.uk/government/groups/open-standards-board>) has final approval. Read more about the approval process for cross-platform character encoding (<http://webarchive.nationalarchives.gov.uk/20170605173702/https://standards.data.gov.uk/challenge/cross-platform-character-encoding>).

## 2. How this standard meet user needs

Users of this standard include:

- publishers of government data
- data scientists
- data analysts
- developers

UTF-8 is an international standard. By using it you can read, write, store and exchange text that remains stable over time and across different systems.

You will also:

- prevent accidental or unanticipated corruption of text as it transfers between systems
- save operational costs by making it easier to find and fix errors in the text
- have accurately translated languages moving between systems
- keep file sizes smaller

## 3. How to use the standard

To use UTF-8 you need to:

- save text in UTF-8 encoding to apply it to your content (<https://www.w3.org/International/questions/qa-choosing-encodings>)
- declare the character encoding, for example, W3 has an example of declaring encodings in HTML (<https://www.w3.org/International/questions/qa-html-encoding-declarations#quickanswer>)
- check your server has the correct HTTP declarations so that they do not override your encoding

Read the W3.org article on migrating to Unicode (<https://www.w3.org/International/articles/unicode-migration/>) for more information.

Print this page