
Determine the character set of a MARC record

- **Product:** Voyager
 - **Relevant for Installation Type:** Multi-Tenant Direct, Dedicated-Direct, Local, TotalCare
-

Question

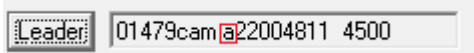
How to determine the character set (encoding scheme) of a MARC 21 bibliographic, authority or holdings record.

Answer

You may determine the character set of a record from byte 9 of the bibliographic, authority or holdings record Leader.

The value of "a" indicates the record is Unicode (e.g., UTF-8) -- an industry subset of the Universal Coded Character Set (UCS); a blank indicates a non-Unicode 8-bit character set (e.g., MARC-8).

Example of UTF-8 in byte 9 of the Voyager Cataloging Client record:



The screenshot shows a MARC record Leader field. The text is "01479cam a22004811 4500". The character 'a' at the 9th position is highlighted with a red box, indicating it represents UTF-8 encoding.

Additional Information

Only one encoding scheme may be used in a MARC 21 record: MARC-8 or Unicode.

For more information see: <https://www.loc.gov/marc/specifications/speccharintro.html>

PRO TIP: Use a tool like MarcEdit to examine the import file in human-readable ".mrk" format.

- **Article last edited:** 08-Jul-2020