Policy justification: Numeric Character References (NCRs) are common markup constructs used in markup languages like HTML and XML, where a sequence of characters will be rendered as a single character. NCRs are structured as ampersand ( & ), pound sign ( # ), lowercase letter x, four-position Unicode character code, and a trailing semicolon ( ; ). For example, &#x091A;. This policy is about the use of NCRs in MARC cataloging records in OCLC (and hence Alma).

Catalogers most often use NCRs in the context of non-Latin scripts. Catalogers may supply parallel non-Latin fields only for scripts supported by OCLC, in accordance with our Network Level Bibliographic Mandate.

Non-Latin scripts supported by OCLC:

- MARC-8 scripts (subsets of UTF-8 characters, so they are also compatible with UTF-8 Unicode): Arabic, CJK (Chinese, Japanese, Korean), Cyrillic (within the MARC-8 character set), Greek, or Hebrew scripts.
- UTF-8 Unicode only scripts: Armenian, Bengali, Cyrillic (outside the MARC-8 character set), Devanagari, Ethiopic, Syriac, Tamil, or Thai scripts. These scripts are not included in MARC-8.

Notice that Cyrillic is the one script that straddles both categories. In general, modern Slavic languages using Cyrillic (like Russian, Bulgarian, Ukrainian) are within MARC-8, and Old Church Slavic and non-Slavic languages using Cyrillic (like Kazakh, Uzbek and Mongolian) use additional Unicode characters.

If any non-MARC-8 scripts are exported in MARC-8 data format, the non-MARC-8 characters are saved in Numeric Character Reference (NCR) format (see http://en.wikipedia.org/wiki/Numeric_Character_Reference for more information). However, all settings for a shared LSP should be UTF-8 Unicode.

NCRs should NOT be used to create non-Latin scripts for scripts not supported by OCLC. Examples include Georgian, Khmer, and anything else not listed above. Non-Latin fields using NCRs to render nonsupported scripts should be deleted from OCLC master records when cataloging.

However, some scripts/characters that were not supported in the past are now supported by OCLC, so non-Latin fields using NCRs in those cases will need to be upgraded.
Old record using NCRs, should be upgraded when used for cataloging (Mongolian example):

![Old record using NCRs](image1)

Upgraded record using newly supported characters (Kazakh example):

![Upgraded record using newly supported characters](image2)

Connexion Client Help advises that non-Latin characters that are not supported by OCLC can be entered and exported to a local system using Unicode, but then deleted from the OCLC master record (NOTE: this violates SUNY policy, don’t do it); or you can enter the name of the character within square brackets using the Unicode standard name (for example, enter [schwa]), or for CJK characters, enter the reading of the character (for example, enter [yin]). NOTE: Client Help does not advise using NCRs. But AskUs@oclc.org told us via email: “However, for the stray unsupported character that appears in a supported script, then the NCR is appropriate. If the script is not supported, it should not be represented in the database, but rather be transliterated.”

Exceptions to this policy may be made in the case of large record sets provided by vendors, but SUNY members must make a commitment to using the available records that most closely adhere to this policy in such cases.

See OCLC’s International Cataloging Manual for more details:
http://www.oclc.org/content/dam/support/connexion/documentation/client/international/internationalcataloging.pdf