

German National Library
30 January 2012

Changes to UTF-8 character set

Dear Sir or Madam

The German National Library is to revise the conversion service for its data provision interfaces. This will result in changes in the UTF-8 coding of our data services. This affects all XML versions of the data formats, the MARC-21 formats and GND-MARC 21 in the UTF-8 character set. The changeover will take place simultaneously with the introduction of the Integrated Authority File (GND) in mid-April 2012.

In concrete terms this means that the data records in UTF-8 will be issued in the original character set without any conversion processes. The uniform encoding form will now be the decomposed form, meaning that special characters will consist of the basic letters and the relevant diacritical marks. Umlauts will e.g. be issued as a basic letter and an appended "combining diaeresis" on the vowel (U+0308).

Please do not hesitate to contact us should you have any questions.

Best regards,

Heike Eichenauer

--

Heike Eichenauer
German National Library
Digital services
Adickesallee 1
D-60322 Frankfurt
Phone: ++49 69 1525 1074
Fax: ++49 69 1525 1636
<mailto:h.eichenauer@dnb.de>
<http://www.dnb.de>