



Library of Congress

# *CDS Announcement*

## **Problem With Extended Arabic Characters**

MARC-8 records in the Persian/Farsi languages (and other languages that may use characters from the extended Arabic set) distributed since 3/1/07 may contain invalid or incorrect character values. Some records contain a mix of basic Arabic characters represented by the G0 set and extended Arabic characters represented in the G1 set in 880 fields. This mix of G0 and G1 will not pose a problem in some systems, but may prove to be unacceptable to others. Note that the UTF-8 files for these records are correct as issued. We recommend that customers finding these records problematic use the UTF-8 distribution files rather than the MARC-8 files. CDS continues to analyze the problem and hopes to resolve the situation soon. Once a solution is in place we will redistribute the MARC-8 records with the correct character values.

Questions regarding these records can be directed to:

Cataloging Policy and Support Office  
email: [cpsol@loc.gov](mailto:cpsol@loc.gov)

September 11, 2007