

**Title:** Fast-tracking Arabic Letter Mark (ALM)  
**Author:** Roozbeh Pournader, Aharon Lanin (Google)  
**Date:** 2013-01-29  
**Action:** For UTC's action

## Suggested action

- Fast-track U+061C ARABIC LETTER MARK (ALM) for Unicode 6.3, to be included with the four other bidi characters to be included in Unicode 6.3, U+2066..2069, the bidi isolates.
- Update UAX #9 Section 2.4 to mention ALM, next to LRM and RLM.

## Rationale

We are working on a consistent approach for bidi locale data in CLDR, and it appears to us that bidi isolates would be an important part of a good model to specify the definition and behavior of bidi CLDR data going forward.

For example, we plan to ask the CLDR committee to avoid trying to insert too many bidi controls to make sure the string is formatted properly in all potential scenarios, but instead assume that the string will always be displayed in a specific bidi context. When the context happens to be different or unknown, users of CLDR data can wrap the locale data in bidi isolates that give the strings the context assumed for them by CLDR.

We are still in the process of finding how best to specify this for different kinds of data in CLDR, but for example, the formatting data for a 12-hour time of the day may be able to specify that a time strings like:

2 אהה"צ

which is “2 p.m.” in Hebrew will always have the 2 appear to the right side of the PM string, even if it is included in a left-to-right context. Implementations that wrap the strings may also be able to make sure that the string doesn't change the ordering of other characters around it.

There is a missing piece for giving a string the context assumed by locale data. Locale data like dates and signed numbers behave differently when following Arabic letters (bidi class AL) as opposed to following Hebrew letters (bidi class R). For example, the logical character sequence

<DIGIT ONE, HYPHEN-MINUS, DIGIT TWO> would be displayed as “2-1” if following an Arabic letter, but as “1-2” if following a Hebrew letter or RLM.

It makes no sense for CLDR to assume a Hebrew context for locales in other scripts like Arabic (there are thirteen such right-to-left locales presently in CLDR). Rather, the assumption has to be that a string formatted for an Arabic-script locale will be displayed in an Arabic-script context. The problem is that currently, there is no way for the user to force that context when the actual context happens to be different or unknown. The ALM is the only solution to this problem.

Instead of waiting for Unicode 7.0 to include ALM (already approved by the UTC in November 2011 and in stage 6 of the ISO process), it would be extremely useful to fast-track it to be encoded with the bidi isolates in Unicode 6.3. That way, CLDR will have a complete toolbox for handling bidi locale data as early as possible.