To: UTC
From: Rick McGowan and Ken Whistler (via Debbie Anderson)
RE: Roadmap Boundary Constraints
Date: 26 October 2010

The Roadmap Committee requests members of the UTC to re-examine the constraints on the allocation of blocks on the SMP.

Background
For some time now the UTC (and WG2) have been abiding by a constraint on the allocation of scripts, such that if they can be fit within 8 or fewer columns (≤ 128 characters), those columns will not be allocated across a half-row boundary. Thus a 5-column block could be allocated as xx00..xx4F or xx30..xx7F or xx80..xxCF, etc., but not, for example at xx40..xx8F, crossing the xx7F/xx80 boundary.

While not a normative requirement, for some time this constraint has been written into the text of the standard as a guideline. (See TUS 5.0, p. 41 or TUS 5.2, p. 35.)

For supplementary characters, there is an additional allocation constraint not to cross 1,024-code-point boundaries.

The stated rationale for these constraints is:
"Such constraints enable better optimizations for tasks such as building tables for access to character properties."

At this point the application of the half-row boundary allocation constraint on the BMP is basically moot, as the BMP is essentially filled, and no new script block allocations could be made there which would violate the constraint anyway.

However, a question has arisen about the continued applicability of the half-row boundary allocation constraint, in the context of the allocations for historic Indic scripts, some of which require three, four, or five columns. The SMP could be packed more tightly with such scripts if the half-row boundary allocation constraint could be relaxed for the SMP.

As a consequence, the Roadmap Committee requests the UTC to reconsider the applicability of the half-row boundary allocation constraint in the SMP. In particular, do UTC members know of implementation dependencies or optimizations that rely on this constraint, perhaps in the following contexts:

1. Table construction and optimization for character properties. (Note that case tables are not at issue here, as the concern for allocation only involves scripts without casing.)

2. Identification of script runs in text.
3. Setting of range windows for compression. (Note that compression of most of the scripts involved here is generally not a major concern, as we are talking about very low-use historic scripts generally.)

If not, the Roadmap Committee would request that the UTC drop the half-row boundary allocation constraint for the SMP, to enable more efficient use of the encoding space for packing historic scripts.

The Roadmap Committee is **not** requesting any change to the 1,024-code-point boundary allocation constraint.