TITLE: Summary Report on Minority Scripts
SOURCE: Deborah Anderson, UC Berkeley (liaison member to SC2)
DATE: 8 April 2007
ACTION: For consideration by WG2
DISTRIBUTION: ISO/IEC JTC1/SC2/WG 2

Summary
With ISO/IEC 10646 Amendment 2, two new modern scripts were encoded, N'Ko and Balinese. However, at least twenty-seven modern minority scripts are not in ISO/IEC 10646, impacting electronic communication with these populations in their native script, and impairing access to their historic and cultural documents. The affected populations have shown considerable interest in becoming members of the digital world, but various factors have prevented them, including: (a) the need to more actively involve countries and script experts from such countries in IT standards' development; (b) lack of funding for research and writing of script proposals, for the development of ISO/IEC 10646 and Unicode compliant fonts, and for instruction in the how to develop such fonts and input methods/keyboards.

1. Unencoded Minority Scripts
At least twenty-seven minority scripts not encoded in ISO/IEC 10646, coming from four geographical areas: Africa, Southeast Asia, the Indian subcontinent (comprising India, Nepal, and Bangladesh), and the Middle East:

> *Africa*: Bagam, Bamum, Bassa Vah, Kpelle, Loma, Mende
> *Southeast Asia*: Batak, Cham, Dai Bopomofo, Fraser, Javanese, Nushu, Pahawh
>   Hmong, Pollard, Tai Viet
> *India*/*Nepal*/*Bangladesh*: Balti, Chakma, Grantha, Kaithi, Maithili, Modi,
>   Newari, Sharada, Sorang Sompeng, Tulu, Varang Kshiti
> *Middle East*:   Samaritan

A number of other scripts could be included; additional information is needed to fully evaluate their eligibility. For details, see the Script Encoding Initiative research page (http://linguistics.berkeley.edu/sei/USR.html).

2. Obstacles to Encoding Minority Scripts
Several obstacles stand in the way of encoding minority scripts, but two predominate:

(a) Greater Participation in IT Standards' Activities

Because the unencoded minority scripts are less well-known, the involvement of members of the minority user communities and script experts is critical. Ideally, these communities should directly participate via their national body representatives in IT standards' activities, but the cost for national body members to travel to meetings -- often held throughout the world – is a serious impediment. For example, 37 African countries are members of ISO,[1] but none are actively involved in discussions of ISO 10646, and no African country is a member of the Unicode Consortium. It is extremely important to engage experts in work on Unicode and its implementations, and to include them formally into the script and language standards process of JTC 1, which is in keeping with the ISO Action Plan for Developing Countries 2005-2010.[2] For African languages, one option might be to include representatives of ACALAN (African Academy of Languages) in technical committee deliberations dealing with Unicode and language issues. This organ of the African Union can't speak for each nation, but may be able to bring the various countries closer together. Another option would be to encourage members to participate via online email discussions (such as on the relevant Unicode email lists). A third possibility is for national body members to help locate experts who can answer questions on script proposals.

(b) Funding

All stages of work on unencoded scripts need support, for tasks ranging from performing research and writing character encoding proposals to font development. Unfortunately, raising funds for these tasks can be very difficult, particularly as script encoding and font development is not a trendy topic that can be easily explained to funding bodies. The UC Berkeley Script Encoding Initiative, a project that works with user communities to help get scripts encoded, tried three times before successfully receiving a grant from the US National Endowment for the Humanities. The current NEH two-year grant has no guarantee of funding beyond two years.

On the educational side, work needs to be done to provide instruction on how to develop Unicode fonts and create converters (from legacy fonts to Unicode-compliant ones), as well as how to develop input methods or keyboards.

One other project that should be supported is to ensure that a member of the user community submits locale data for a given language to the Common Locale Data Repository (CLDR, http://www.unicode.org/cldr/) which will, in turn, help with localization development.

In sum, stable funding is necessary to assure ongoing work on script proposals, promoting the development of fonts and input methods, and locale data submission.

---

[1] "ISO and Africa" report, June 2006, http://www.iso.org/iso/en/comms-markets/developingcountries/pdf/iso_and_africa.pdf

[2] http://www.iso.org/iso/en/prods-services/otherpubs/pdf/actionplan_2005-en.pdf