

Title:	Unicode Consortium Liaison Report to TC 37/SC 2
Source:	Unicode Consortium
Author:	Peter Constable
Status:	Liaison Contribution
Action:	For consideration by TC 37/SC 2
Date:	June 10, 2013

The Unicode Consortium has noted activity within SC2 to create a new ISO 639 standard that unifies the existing multiple parts of ISO 639 into a single standard and that proposes new maintenance processes, including proposals for a new registration authority.

The Unicode Consortium acknowledges possible benefits of unifying the various parts of ISO 639 into a single standard since it is imperative that the alpha-2 and alpha-3 code spaces for languages be maintained in a unified, integral manner.

However, the Unicode Consortium is also very concerned with proposals to change the maintenance processes. ISO 639 standard is among the most widely-used of ISO standards with an extremely broad range of implementations and a vast number of users of those implementations. In particular, ISO 639 has become a fundamental element in Internet and Web technologies and in all modern software platforms. ISO 639 is, of course, of particular interest to terminologists and linguists; but the vast majority of usage today of ISO 639 lies in the domain of information technology. The view of the Unicode Consortium is that the maintenance process for ISO 639 must be considered with a high priority given to stability and to impact on existing usage in information technology and on the Internet. In reviewing recent contributions to SC2 and to WG1, the relationship between ISO 639 and the information technology domain appears to be under-represented. This is of concern, and suggests a potential risk factor for a very large proportion of the existing users of ISO 639.

Before changes in maintenance processes or in the registration authority are considered, a clear business case for change should be provided. It is not clear at present why changes are being considered or what particular problems are trying to be solved. The maxim, "If it isn't broken, don't fix it," seems to us appropriate in this case.

We have seen documents that have been circulated informally (not in the SC2 or WG1 document registers), such as a presentation by Dr. Gerhard Budin, "ISO 639: Analysis of Current Situation and Proposals for Future Development", that asserts that problems exist; e.g., "Decisions to be taken often on shaky grounds...", "Registration/Maintenance complicated, outdated...", etc. However, no specific and measurable problems have been cited that make these points actionable. Thus, it is unclear how to evaluate these claims or the goals for language coding of the parties expressing these views. In any process, there may be opportunity for improvement. However, specific problems and goals should be identified, and potential mitigations or solutions should be evaluated.

It should be stressed that, rather than entertaining radical and sweeping changes, a conservative approach to changes that seeks to solve specific problems with clear business needs is a better approach to ensure on-going success of ISO 639.

The Unicode Consortium endorses the on-going role of Library of Congress and SIL International in their roles as registration authorities for parts 2, 3 and 5. These agencies have a proven track record in their ability to provide clear documentation and supporting data, and in the processes they have used for maintaining the alpha-3 code space.

Among other changes that might be worth consideration in a new edition of ISO 639, the Unicode Consortium would like to recommend that the alpha-2 code space be stabilized. The existing alpha-2 identifiers are very-widely used, but as a whole the alpha-2 code space is a legacy technology, and modern implementations in the information technology domain, particularly the Internet specification BCP 47, already supplements the existing alpha-2 identifiers with alpha-3 identifiers—that is, any language that is not currently coded in the alpha-2 code space is represented by an alpha-3 identifier. Any new additions to the alpha-2 code space would create dual coded representations that would result in significant confusion and cost to a broad range of users of ISO 639.

In summary, the Unicode Consortium is concerned that changes might be considered that could have significant and destabilizing impacts on the information technology domain, including Internet and Web technologies as well as widely-used software platforms. We strongly request that actions considered by SC2 be measured and carefully considered, with strong business justifications and clear goals identified.