



ISO/IEC JTC1 **SC36 N0314**

ISO/IEC JTC1 SC36
Information Technology for Learning, Education, and Training

Title:

Enquiry to SC22 Regarding Recommendations for Locale
Specification for Internationalization and Localization

Source:

SC22 Liaison

Project:

All

Document Type:

Working Paper

Status:

FYI

Date:

2002-08-22

Action ID:

FYI

Distribution:

P, O, & L Members, WG Conveners

There is a need in SC36's field of work (Information Technology for Learning, Education, and Training) to specify "locale" information for use in internationalization (I18N) and localization (L10N) contexts. These contexts may include the L10N services (e.g., causing IT systems to adapt to the requirements of a particular "locale"), but might also include describe "content" (e.g., a locale is associated with or attached to some "content", such as the XML "**lang**" attribute). Furthermore, within certain IT environments, it is necessary to maintain multiple localized variants (e.g., a multi-lingual/cultural message list — for each kind of message, there is an array of **{string, locale identifier}** pairs, one for each locale).

Based on preliminary investigation and discussion on the SC22/WG20 (Internationalization) E-mail reflector, the following specifications appear to be relevant:

- ISO/IEC 9945-1 (POSIX, Part 1) specifies a locale identifier, using the pattern **{language code} + underscore character + {country code}**. UNIX and Java systems use this format.
- IETF RFC 3066 (a revision of RFC 1766) specifies a language identifier, using the pattern **{language code} + hyphen character + {country code}**. Note: A good number of developers of RFCs 1766 and 3066 also participate in SC22/WG20 and also participate in the SC22/WG15 (POSIX) development.

In the WG20 E-mail discussion, it appears that the subtle distinction is that the locale identifier in 9945-1 describes the "user's" environment, while the locale identifier in 3066 describes the language. In the E-mail discussion (message number SC22WG20.4131), it appears that one can write a locale identifier such as "en-GB_US", which might mean "the language is British English, but the user is operating within the US environment".

There appear to be a variety of features one would want to incorporate into such a locale identifier, including language subcodes (e.g., "en-US-philadelphia"), jurisdictional domains as described by ISO 3166-2 and ISO/IEC CD 18038 (e.g., "US-NY"), and personal or group locale-specific preferences.

While the 9945-1 and 3066 specifications have partially address this solution, it is important to have a common standard that addresses a more complete solution. When discussing the problem with SC22/WG20 via the E-mail reflector, they had little interest in solving this kind of problem. When discussing the problem informally with participants in other JTC1 SCs (e.g., SC32, SC25, SC29, SC34), there seemed to be interest in a common solution, but all participants felt this problem was an "internationalization" issue and, thus, properly belongs in the domain of SC22/WG20.

The following are requests to SC22:

- Can SC22 determine if this kind of work belongs in SC22/WG20? If not, where should this work be developed? Should this request be forwarded to the CLAU Technical Direction?
- Considering the cross-SC relevance, should this be a topic of discussion at the JTC1 Plenary?
- Can SC22 make recommendations on relevant standards, technical reports, and other documents that would better inform users of this kind of technology and would identify harmonization and compatibility concerns?

Thank you in advance for your consideration and your advice.

Sincerely,
Frank Farance ("frank@farance.com", tel: +1 212 486 4700)
Liaison from SC36 to SC22