



User interfaces

ISO/CEI/JTC 001/SC 35

Date:
2011-08-16

Doc. Number:
N 1685

Assistant:
Amelle MOURADI
Direct line : 01 41 62 84 28
amelle.mouradi@afnor.org

Your contact:
Philippe MAGNABOSCO
Direct line : +33 1 41 62 85 02
philippe.magnabosco@afnor.org

Response to the NP ballot on Specification methods for cultural conventions

FOLLOW
UP

For information

Title: Response to the NP ballot on Specification methods for cultural conventions

Source: Keld Simonsen, editor 30112

Date: 2011-08-16

Document reference: N1649

The project was approved as project 30112 as reported in SC35 N1649, with a majority of national bodies approving the project and scope, and with 6 national bodies committing to active participation. A number of comments were received for the NP. This document gives responses to the received comments.

US-01 ge

First, the WD document accompanying the NP ballot is a minor revision of a nearly decade-old document that was the object of an unsuccessful project to create a TR of the same title in the context of JTC 1/SC 22. (See <http://std.dkuug.dk/JTC1/SC22/WG20/docs/projects#14652>.) No consensus could be reached at that time due to extensive technical concerns with the draft as well as concerns other larger concerns questioning the value of the TR. As observed in ballot comments at that time, "Germany sees little use in this DTR. It has only very limited support in the industry (not even in the Linux community..." (See "Summary of Voting on JTC 1 N 6721", ISO/IEC JTC 1/SC 22/N3405.)

The same technical and general concerns exist today. The document is inextricably tied to POSIX conventions that were already dated in 2001 and have little interest in 2011. Moreover, in the meantime different technical solutions have emerged that have generated a strong following and are defining best practices in a way that render the present document irrelevant and obsolete.

Beyond the specifics of the WD document that accompanied the NB ballot, the US has more general concerns with this project proposal. On the one hand, there is merit to the idea of specifying methods to describe various culturally-dependent conventions for one or many cultures and, in particular, of doing so in ways that facilitate easy development of user interfaces or other types of applications so as to be culturally-relevant for users of many different cultures. That is an objective that the US strongly supports. However, we strongly object to this NP because it completely overlooks the fact that this objective is already being achieved in the context of existing and vigorous standardization activities outside of ISO/IEC JTC 1/SC 35. This project would directly compete with the Common Locale Data Repository (CLDR) project of the Unicode Consortium. This project in SC 35 would needlessly duplicate efforts, and it is doubtful that it would be possible to match the depth of participation and collaboration of the CLDR effort.

The CLDR project has been underway since 2002. It is widely adopted across the software industry; it has support from standards associations in several countries and is building on the ongoing contributions of hundreds of contributors. The project has compiled and provides data specifying a wide range of cultural data categories of use in user interfaces and applications with coverage spanning 187 languages and 166 territories. In addition to the data repository, the CLDR project has developed a standard that specifies a data format for interchanging cultural-convention data, "Unicode Technical Standard #35, Unicode Locale Data Markup Language" (LDML). Moreover, the proposed project covers only a small fraction of the depth and scope of types of data covered by LDML.

The justification for the NP states,

"The purpose of the proposed specification is to make it possible for a program to adapt itself to the cultural and linguistic environment that it is being used in. It would thus be possible to write programs, that can via these specifications be given a number of cultural and linguistic user interfaces, without changing the program itself."

The reality is that this purpose, thanks to the success of the CLDR project, has already been largely achieved, and that the CLDR approach is widely considered to represent best practice in this area. Nearly a decade of experience with CLDR has shown that this is an area that is eminently suited to its chosen approach, which is a broad-based, open source collaboration of cultural and industry experts from all areas of the globe. Because of this, it is unlikely that a competing specification developed by JTC 1/SC 35 could add significant value and it is equally unlikely that there would be any significant adoption of such a specification by industry. Instead, if such a project were to be approved, the attempt could lead to a fragmentation of effort and waste of resources.

In summary, the WD submitted is not an appropriate basis to develop best-practice recommendations today, and the very project would stand in direct competition to existing standards efforts elsewhere that are widely endorsed.

Therefore, this proposal should not be approved as an SC 35 project.

Response: noted.

The NP draft follows very closely the GNU libc implementation, which in turn has been implementing the previous version of the NP (TR 14652) closely. On the <http://sourceware.org/glibc/> - latest updated 2011-06-08, it says:

"The GNU C library is primarily designed to be a portable and high performance C library. It follows all relevant standards (ISO C 99, POSIX.1c, POSIX.1j, POSIX.1d, Unix98, Single Unix Specification). It is also internationalized and has one of the most complete internationalization interfaces known. "

The people behind GNU libc thus states that its i18n capabilities are reasonably advanced, and not outdated. The NP WD enhances this interface, and is planned to be the base for further enhancements.

It is not the impression that the project will create a lot of new work, as that work is already underway, eg. for glibc. glibc has significant following in the industry, and millions of users depend on this for everyday usage. Today 286 locales for combinations of languages and territories have been developed and distributed with glibc, about the same magnitude as for CLDR. Hundreds of developers are involved in maintaining the locale data.

The purpose of facilitating internationalized programs has been to a great extent been the success of the previous TR 14652 specification in the POSIX world.

The NP WD has a number of advantages over LDML:

1. It is character set independent, while LDML only works in Unicode. Thus it can serve also for programming languages that has support for native treatment of other character sets than Unicode. It is thus well suited also for embedded work.
2. It includes a number of advanced features, that LDML apparently does not support, such as keyboard selection, postal address formatting, and personal titles.
3. It has a specific repertoire concept, that facilitates handling of character repertoires, for example for transliteration.
4. It can be used in environments using POSIX locale syntax, and be expanded with customized keywords in POSIX locale syntax. It can be employed with POSIX/C/C++ APIs and has been done so extensively. There has been built a living and productive ecosystem, with active development, hundreds of locale data contributors. and many thousands of programmers. Activities of the locale data maintainers can be found at <http://sourceware.org/ml/libc-locale/>

There are many millions of users using programs made with help from the family of specifications around TR 14652, but it is difficult to estimate exactly how many.

The use of the platform on GNU-Linux is covering, according to Gartner Group¹, 4% of the market in 2008 while Windows-powered PCs represent more than 90% of unit sales², but "the Inquirer "considers that from 2007 to 2008, GNU-Linux use was multiplied by 28³. Canonical, besides, is telling their resellers that 11 % of U.S. businesses already downloaded Ubuntu at least once⁴.

On the opposite, the reference study NetMarketShare⁵ decreases to 1% GNU-Linux market share. Bruce Byfield⁶ and also the RedHat's interactive Open Source Index (OSI) based on 2008 research conducted by GeorgiaTech⁷, considers this figure has been disputed because based on an extrapolation of North America Web traffic. GNU-Linux coming from plethora of vendors and community projects, rather than a single corporation, requiring neither registration nor activation⁸, each download might never be used, or for a short time. Another download or simple DVD might be used to install Linux onto 100 computers⁹. It's why Byfield, by consolidating data from multiple servers' browsers, believes the share to be between 2 to 5% but admit that "the problem [...]

1 <http://ca.biz.yahoo.com/ibd/080807/tech.html?.v=1>

2 http://digg.com/news/story/GNU_Linux_Desktop_Market_Share_is_4_Gartner

3 <http://www.theinquirer.net/inquirer/news/1038132/linux-preinstalls-rocket-cent>

4 <http://www.thevarguy.com/2008/08/10/canonical-to-vars-11-of-us-businesses-use-ubuntu-linux>

5 <http://marketshare.hitslink.com/operating-system-market-share.aspx?qprid=8&sample=35>

6 <http://itmanagement.earthweb.com/osrc/article.php/3818696>

7 <http://www.redhat.com/about/where-is-open-source/activity/>

8 <http://itmanagement.earthweb.com/osrc/article.php/3818696>

can never be settled with any certainty".

On his blog, Aaron Seigo¹⁰ highlights the differences between the figures in North America and the rest of the world. Because of European governments ministries adopting GNU-Linux for their administration or the 53,000 Brazilian schools pedagogical networks using the free open source software¹¹ (FOSS) for 52 million schoolchildren or students¹², Red Flag Linux-Asianux dominance¹³ in China and other successful implementation of this kind in India or South America, he suggests that 8% are probably a bottom figure, and 10 to 12% likely the reality.

According to the Netcraft's Web Server Survey¹⁴ for July 2011, the Apache servers, in this background, are reaching a total of 65 % of the active server across all domains (235,326,985), 17 % using Microsoft IIS (i.e. 60,086,346). Moreover, Apache is increasing its leading market share by nearly 66,1% (i.e. 656,701) for top servers used to host the world million busiest sites.

This explains why it is so methodologically difficult to know the exact outlook of GNU / Linux distributions worldwide, but anyway shows that it is used for a significant part of the market and for heavy networked applications.

It is good to see that the US agrees with the scope and purpose contained in the NP in N1616. Given that there are two families of i18n support for programming languages, one which requires support for an implementation defined coded character set like in POSIX, C and C++, and another that builds all character support on ISO 10646/Unicode like Java, maybe there should also be two standardized families of i18n specifications – and then an effort to align the two approaches.

CA-01 ge

Although Canada supports the intent of this work item and agrees that it should be worked upon, Canada does not accept the proposal in the attached NWI Proposal document as a sufficient definition of the new work item for the following reason:

The scope should be adjusted to deal with the following 4 remarks:

i) These documents completely ignore the work of the Common Locale Data Repository (CLDR) project (see: <http://cldr.unicode.org/>) and the Local Data Markup Language (LDML) (see: <http://www.unicode.org/reports/tr35/>) specification over the last 10 years. CLDR data is now the accepted reference for most common locale specific data, and to do anything else is pressing an agenda that is simply not sustainable.

By ignoring the work in the industry on CLDR the author(s) of the NP from Swedish NB are not presenting the complete background information on the subject matter.

Also B.3 Other Source of Standards .. says NO. This is not acceptable. There was work done before in ISO/IEC: TR 11017: 1998 then past work in CEN/TC304, as well as the work on LDML and CLDR (outside of ISO/IEC), and possibly others in the ISO/IEC arena exist.

Response: Agreed in principle.

The WD refers normatively to 15897 which again refers to LDML and CLDR and Unicode. The WD itself uses data which is explicitly described to be built from Unicode 5.0. TR 11017 is mentioned several times in the WD. Some CEN work is mentioned in the WD. It was by no means meant to ignore the Unicode work, actually the Unicode work was the only other work mentioned in the NWIP via the reference to IS 10646, the ISO version of Unicode. Reference to LDML, CLDR, and TR 11017 will be added to the bibliography.

9 <http://www.linuxloop.com/2008/08/14/linux-market-share-is-there-any-way-to-know/>

10 <http://aseigo.blogspot.com>

11 <http://hdl.handle.net/2042/9007>

12 http://portal.mec.gov.br/index.php?option=com_jfilter&Itemid=164¶ms%5Bsearch_relevance%5D=linux¶ms%5Bsearch_method%5D=exact¶ms%5Btipobusca%5D=null

13 <http://www.redflag-linux.com/en/>

14 <http://news.netcraft.com/archives/2011/07/08/july-2011-web-server-survey.html>

ii) It should be clarified whether APIs (which deals with "application coding"), whose requirements show up in the WD is a requirement for coding. This item showed that it is not clear in Canada what coding really means in the NP form.

Response: The SC 35 secretariat was consulted and it was resolved that APIs are not related to the question on the NP form on coding.

iii) SC35 should state what is the synch with the taxonomy TR already approved.

Response: Accepted. An annex will be added to document the relation of the specification to the i18n taxonomy described in ISO/IEC TR 24785.

iv) Answer to A.1 ... says DESIRABLE.

In fact no product can really market worldwide without Globalization and Localization support .. it is ESSENTIAL ... not just DESIRABLE.

Agreed.

None of the above listed responses to the Canadian comments involves adjustment of the scope of the project.

FI-01 ge

This NP should be more concise and focused to have any real value. There should be a concerted effort to avoid adding data content, and instead to refer to documents such as ISO/IEC 10646 and the Unicode Consortium's CLDR (Common Locale Data Repository). Where data items are needed, they should be synchronized with CLDR. Otherwise we are left with the untenable situation of adding another depository of data which must be kept up-to-date, as the needs of developers continually evolve.

The industry has moved towards using LDML (Locale Data Markup Language) as the specification method for the locale data that is commonly stored in CLDR. The 1.0 version of CLDR was hosted on the Open18N site before the CLDR project moved to the Unicode Consortium (already in 2004). In addition to LDML and the Repository, the CLDR Technical Committee has developed the Survey Tool to easily enter the data for a large number of language/region combinations (currently over 500 locales) by any individual or company submitters and to agree on its validity by the vetting process. (A bug reporting mechanism is also available.) The data is stored in an XML interchange format that is subsequently transformed into forms optimized for use in actual implementations based on ICU (Internationalization Components for Unicode), POSIX, OpenOffice, dojo (an open source modular JavaScript library), and others. As a result, one cannot expect that this NP would have any practical use to improve the localization situation for the users or the industry.

Response: Noted. Internationalization is by nature a many-faceted area. The project at hand given by the associated WD does try to make a concerted effort to align with other ISO and Unicode work, and other work in the industry, for example are the data directly derived from Unicode 3.0 and the sorting data is a direct inclusion of IS 14651 data. An effort has been taken to align data items with Unicode, and that process will also be taken for the future work, both with regards to Unicode and LDML and other industry work together with user requirements. The work described in the WD is reflecting work going on in the POSIX/Linux industry since the early 1990'ies, and this work is still actively being maintained today. That LDML and CLDR has good data collection tools is noted – and it is known that other mechanisms for good maintenance of internationalization data is also available from other parties in the industry like the GNU glibc community.

SE-01

Sweden recently realised that since JTC 1 adopted the ISO/IEC regulations, part 1, there are not three types of TR anymore. This document should therefore be designated a Technical Specification, TS, instead of a TR.

Response: The SC 35 secretariat will be consulted for the proper document type of the specification.

End of document