

Title: L2 Codes and Character Sets Liaison Report

Author: Dave Birdsall
Status: Informational
Date: January 22, 2001

1. Introduction

L2 is the NCITS committee that deals with codes and character set issues. More recently, L2 has expanded its focus into internationalization issues.

Most of the information from this report comes from the L2 web site http://www.ncits.org/tc_home/l2.htm,¹ and from the L2 annual report covering June 1999 through June 2000, <http://www.ncits.org/archive/2000/it000440/it000440.pdf>. I also found L2-related information at the Unicode® Consortium web site <http://www.unicode.org/index.html>.

I have begun an e-mail dialogue with the L2 chair, Arnold Winkler. The concerns expressed in Section 4 come from that dialogue. I plan to attend part of the next L2 meeting, January 29 - February 2 in Mountain View, California. I may post a revision to this paper later reflecting what I learn there.

2. Organization

L2 works closely with the Unicode® Consortium. L2 meets four times a year, co-located with Unicode Technical Committee meetings. As the annual report puts it, this is “economical because most of the members of L2 are also members of the Unicode Consortium and the subject matter overlaps widely.” A listing of up-coming meetings can be found at the Unicode web site, <http://www.unicode.org/unicode/timesens/calendar.html>.

Within ISO, this work is spread over two committees, JTC 1/SC 2 for codes and character sets, and JTC 1/SC 22/WG 20 for internationalization. L2 holds US TAG responsibility for both.

L2’s web site currently lists 16 members. Among them are computer vendors (Apple, Compaq, Hewlett-Packard, IBM, Sun, Unisys), DBMS and other software vendors (Microsoft, Novell, Oracle, Progress Software, Sybase), and a diverse variety of other companies and industry groups (Bell Labs, which I assume is Lucent Technologies, Research Libraries Group, Reuters, Share, and Unicode).

The L2 officers are:

1. Sometimes web search engines can be wonderful. A search on “L2 NCITS” reported this URL as its very first hit.

Chairman - Arnold Winkler, Unisys Corporation
Vice Chairman - Lisa Moore, IBM Corporation
International Representative - Michel Suignard, Microsoft Corporation

Papers are distributed electronically from their web site. Papers are zipped in sets of 20 for faster downloading.

3. Program of Work

3.1 Past Publications

L2 has a very long history. The annual report lists over 30 projects.

Besides the well-known character set standards (e.g. IS 646 also known as ASCII, various flavors of IS 8859, and IS 10646 also known as Unicode), L2 has standardized just about anything imaginable that involves characters. Examples include: *Hollerith Punched Card Code* (IS 6586:1980), *Perforated Tape Code for Information Interchange* (IS 1113:1979), *Printing specifications for optical character recognition* (IS 1831:1980), and *Code Extension Techniques for Use with the 7-bit Coded Character Set for the ANS Code for Information Interchange* (IS 2022:1994). The last example is one that H2 members may remember from SQL3 discussions on “form-of-use.”

3.2 Current Work

L2 is currently maintaining the ISO 8859 (8-bit character set) and 10646 (multi-octet character set) standards, and re-affirming other older standards.

L2 has embarked on new standards in the internationalization area. These include string sorting, specification of cultural conventions (e.g. date and time formats), registration of cultural elements, and Internationalization APIs.

Because of the large number of very stable standards that L2 maintains, L2 has driven an administrative innovation within NCITS. Standards can now be declared “stabilized,” meaning that they do not need to be continually re-affirmed.

3.3 Future Plans

A new 8-bit character set, 8859-16, has been proposed for Romanian, but L2 intends to oppose other new 8-bit standards. L2 prefers that new characters be added to ISO 10646 instead.

4. Relevance and Relationship to H2

Our SQL standards, ISO 9075, contain these normative references to the L2-developed standards:

ISO/IEC 646:1991, *Information technology — ISO 7-bit coded character set for information interchange*.
ISO/IEC 2022:1994, *Information technology — Character code structure and extension techniques*.

ISO/IEC 10646-1:1993, *Information technology — Universal Multi-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*.

In my e-mail dialogue with L2 chair Arnold Winkler, I heard two concerns.

The first was a feeling that a proposal in the international arena, BHX-080, reflected a misunderstanding of the ISO 10646 standards. I do not yet have specifics. I have relayed to him that this paper, and the follow-on paper HEL-052, have been reviewed, and passed along what I understand to be H2's opinion on these papers: that no proposals are expected to come of them. (Please feel free to correct me if this is not the proper understanding!)

The second concern was a desire that the SQL standards support ISO 10646 Parts 1 and 2 “without restrictions”. I'm not aware of any areas where the SQL standards fall short in this regard, but I will probe further to see if there are specific issues.

Other L2-developed standards and reports that may have future relevance to our work are:

TR 10176 — *Guidelines for the preparation of programming language standards*

TR 11017 — *Framework for Internationalization*

IS 14651:2000 — *International string ordering*

5. Contact Information

Arnold Winkler, the L2 chair, can be contacted by e-mail at arnold.winkler@unisys.com.

- End of paper -