

## NCITS / L2 ANNUAL REPORT

Covering the Period from June 2000 through June 2001  
Title of NCITS Subgroup L2: Codes and Character Sets, Internationalization

2001-06-13

### Informal Description of Work:

L2 is the US-TAG for codes and character sets (JTC1/SC2) and for internationalization (JTC1/SC22/WG20). Both subject areas are essential for the development of internationally usable systems and applications. They are particularly important in today's growing marketplace of WWW access and international electronic commerce.

### Executive Summary

The number of L2 members is presently 14. We attribute this strong interest to the rapidly growing application of Unicode / ISO 10646 by many companies and on the World Wide Web; and to the fact that L2 has taken on the responsibility as TAG for SC22/WG20 - Internationalization. L2 works very closely with the Unicode<sup>®</sup> Consortium - all technical meetings are co-located.

L2 meets now 4 times per year. Co-location with the Unicode Technical Committee meetings is economical because most of the members of L2 are also members of the Unicode Consortium and the subject matters overlap widely.

L2 has switched almost totally to electronic document distribution via our web site, including sets of 20 zipped documents for faster downloading. Paper documents are still needed for newly proposed scripts.

The work on ISO 10646, after publication of the second edition of ISO 10646-1:2000, has two aspects at this time. One is the continued addition of scripts and characters to the Basic Multilingual Plane, manifested in Amendment #1 to ISO/IEC 10646-1:2000, currently in FPDAM ballot; the other is the allocation of scripts to ISO 10646-2 which is currently in FDIS ballot. The US is contributing substantially to both projects, e.g. the editor is Michel Suignard from Microsoft in Redmond, he also is IR of L2.

The work on internationalization as TAG of SC22/WG20 includes contributions to and reviews of draft standards and technical reports, such as the international string ordering standard IS 14651, the specification of cultural conventions TR 14652, the internationalization API standard IS 15435, and the registration standard for cultural elements IS 15897.

### Major achievements during this report period:

For all items below, the US has made substantial contributions, for most projects we provide the editor.

ISO/IEC 10646-1:2000 (second edition) – Universal character set, Basic Multilingual Plane (BMP), has been approved and published on CD

ISO/IEC 10646-1:2000, Amendment #1 Universal character set (second edition), Basic Multilingual Plane (BMP), has reached FPDAM ballot stage

ISO/IEC 10646-2 – Universal character set, Supplementary planes, has reached FDIS ballot stage

ISO/IEC TR 10176:2000 (second edition) – Guidelines for the preparation of programming languages, has been approved and is being published

ISO/IEC 14651 – International string ordering, has been approved and published on CD-ROM and web

## L2 Annual Report 2000 - 2001

### Projects (Status June 1, 2001)

Work	Project	NCITS/ANSI	Title	International Standard	International Project
	4 - M		Control Codes for 7-Bit and 8-Bit Sets	IS 6429:1992 [1999]	02.13
	6 - M		Graphic Representation of the Control Characters of the American National Code for Info. Interchange	IS 2047:1975 [1999]	N/A
	12 - M	X3.4:1986 [R1992] [R1997]	Information Processing--Coded Character Sets--7-Bit American National Standard Code for Information Interchange (7-bit ASCII)	IS 646:1991 [1999]	02.01
x	<a href="#">13 - M</a>	X3.83:1995	USA Sponsorship Procedures for ISO Registration According to ISO 2375	IS 2375:1985 under revision	02.04.01
	103 - M		Hollerith Punched Card Code	IS 6586:1980 [1999]	02.15
	105 - M		Code Extension Techniques for Use with the 7-Bit Coded Character Set for the ANS Code for Information Interchange	IS 2022:1994 [1999]	02.02
	106 - L		USA Candidates for Registry	N/A	N/A
	107 - L		Perforated Tape Code for Information Interchange	IS 1113:1979	02.12
	240 - M		Coding of Character Sets for OCR & MICR	IS 2033:1983 [1999]	02.09 02.02033
	304 - M		Hexadecimal Input/Output to Microprocessors Using 5-Bit and 7-Bit Tele-printers	IS 6936:1988 [1999]	N/A 02.06936
x	<a href="#">392 - L</a>	A/I 8859-1:1992	Part x: Latin Alphabet No. x (8-Bit ASCII) (Revision and Re-designation of X3.134.2)	IS 8859-x:1987	02.20.0x
x	<a href="#">396 - D</a> <a href="#">(1313 - L)</a>		Multiple Byte Graphic Character Set for Processing and Interchange – Part 1 Part 2 for Planes 1-16	IS 10646-1:1993 [1999] CD 10646-2	02.18 02.18.01 02.18.02
x	<a href="#">473 - L</a>		Guidelines for the preparation of programming language standards	TR 10176 Amd. #1	22.13 22.13.01
	495 - M	A/I 4873:1991 [1993]	8-bit code for information interchange - structure and rules for implementation (Revision and Re-designation of X3.134.1:199x)	IS 4873:1991	02.08
	880 - L		Framework for Internationalization	TR 11017	22.30.01
	1037 - M		Implementation of the 7-bit coded character set and its 7-bit and 8-bit extensions on 9- track 12,7 mm (0.5 in) magnetic tape	IS 962:1974 [1999]	N/A
	1038 - M		Implementation of the 7- bit coded character set and its 7- bit and 8- bit extensions on 3,81 mm magnetic cassette for data interchange	IS 3275:1974 [1999]	N/A
	1039 - M		Information technology - Coded graphic character set for text communication (revision of parts 1 & 2) - Latin alphabet	IS 6937:1994 [1999]	02.14.01/02
	1040 - M		Information processing - Arabic 7- bit coded character set for information interchange	IS 9036:1987 [1999]	02.19
	1041 - M		Information technology - Standardized coded graphic character sets for use in 8-bit codes	IS 10367:1991 [1999]	02.22
	1042 - M		Information technology - control functions for text communication	IS 10538:1991 [1999]	02.14.03
x	<a href="#">1141 - L</a>		Operational Model for Characters and Glyphs	TR 15285	02.15285
	<a href="#">1241 - L</a>		Functionality for Internationalization	Umbrella	22.30.02.01
x	<a href="#">1241 - L</a> <a href="#">(1307 - L)</a>		International string ordering	FDIS 14651	22.30.02.02
x	<a href="#">1241 - L</a> <a href="#">(1244 - L )</a>		Information Technology - Specification Method for Cultural Conventions	PDTR 14652	22.30.02.03
	1241 - L		Internationalization API standard	CD 15435	22.15435
x	<a href="#">1275 - L</a>		Procedures for European registration of cultural elements (ENV 12005)	IS 15897	22.15897
	1279 - L		OCR-B, shapes and dimensions of the printed image	ISO 1073-2	02.26
	1359 - M	A/I 7350:1991	Registration of repertoires of graphic characters from ISO 10367	IS 7350:1991 [1999]	02.16

## L2 Annual Report 2000 - 2001

Work	Project	NCITS/ANSI	Title	International Standard	International Project
	1435 – L		Un-punched paper cards – Specification	IS 1681:1973	02.01681
	1436 – M	A/I 1831:1980 [1999]	Printing specifications for optical character recognition	IS 1831:1980	02.01831
	1437 – L		Punched paper tape – dimensions and location of feed holes and code	IS 1154:1975	02.01154
	1515-L		Extension of the Latin alphabet coded character set for bibliographic information interchange	IS 5426:1983	02.05426
	1516-L		Extension of the Cyrillic alphabet coded character set for bibliographic information interchange	IS 5427:1984	02.05427
	1517-L		Greek alphabet coded character set for bibliographic information interchange	IS 5428: 1984	02.05428
	1518-L		Documentation - African coded character set for bibliographic information interchange	IS 6438:1983	02.06438
c	1519-L		Information and documentation -- Mathematical coded character set for bibliographical information interchange	IS 6862: 1996	02.06862
c	1520-L		Information and documentation - Armenian alphabet coded character set for bibliographic information interchange	IS 10585:1996	02.10585
c	1521-L		Information and documentation - Georgian alphabet coded character set for bibliographic information interchange	IS 10586:1996	02.10586
c	1522-L		Information and documentation - Extension of the Arabic alphabet coded character set for bibliographic information interchange	IS 11822:1996	02.11822

An “x” in the first column of this table indicates that revision or development work is going on in this project; a “c” indicates re-affirmation activities.. Details can be found below in the project discussions.

### 4-M Control Codes for 7-Bit and 8-Bit Sets

Liaison project for ISO/IEC 6429:1992. No activity.

### 6-M Graphic representation of the control characters of ASCII

Maintenance project for ISO 2047:1975. No activity.

### 12-M Information Processing—Coded Character Sets—7-Bit American National Standard Code for Information Interchange (7-bit ASCII) [SPARC/83-542]

Maintain X3.4. No activity.

### 13-M USA Sponsorship Procedures for ISO Registration According to ISO 2375

L2 decided to retain X3.83 and to propose it as a "stabilized" standard. The international standard ISO 2375 is being revised, especially the process for the registration. Editors are Michael Everson (Ireland) and Ed Hart (USA). The US has requested, and was granted, a seat on the registration review board (Ed Hart).

### 103-L Hollerith Punched Card Code (The U.S. standard X3.26:1980 was withdrawn)

No activity in this liaison project for the ISO standard

### 105-L Code Extension Techniques for Use with the 7-Bit Coded Character Set for the ANSI Code for Information Interchange -- [SPARC/83-393]

Liaison project for ISO/IEC 2022, ISO/IEC 2022 was reconfirmed.

### 106-L USA Candidates for Registry

L2 recommends the use of “collections” of ISO 10646 instead of registrations according to 2375.

### 107-L Perforated Tape Code for Information Interchange

No activity at this liaison project.

### 240-L Coding of Character Sets for OCR & MICR

Liaison project for ISO 2033 and ISO 1073

## **L2 Annual Report 2000 - 2001**

ISO 1073 is being updated to contain the euro character and made into a Technical Report.  
The SC2 secretariat in Japan has taken responsibility for the reference drawings of OCR characters

### **304-L Hexadecimal Input/Output to Microprocessors Using 5-Bit and 7-Bit Teleprinters [SPARC/77-93]**

No activity in this liaison project for ISO 6936

### **392-M 8-Bit Single-Byte Coded Character Sets**

ISO/IEC 8859-16 - Latin 10 for Romania has been approved and published. Editor is Michael Everson (Ireland)  
ISO/IEC 8859-11 – Thai is in FDIS ballot. Editor is Johan van Wingen (Netherlands).  
The character names of all 8859-x standards have been aligned with ISO 10646 names.  
The USA consistently votes against any further parts of this family of 8-bit character set standards. We recommend the use of ISO/IEC 10646 instead.

### **396-D Multiple Byte Graphic Character Set for Processing and Interchange [SPARC/83-545]**

Development of ISO/IEC 10646. This includes the addition of new scripts and characters to the BMP (Basic Multilingual Plane) as well as the development of the architecture and roadmaps for planes beyond the BMP. These are the Secondary Multilingual Plane (SMP), the Supplementary Ideographic Plane (SIP), and Plane 14 for identifiers.

**ISO/IEC 10646-1:2000 (second edition):** This edition, which includes all corrigenda and all amendments to ISO 10646-1:1993 has been approved and was published on CD-ROM. Editor is Michel Suignard, USA.

**ISO/IEC 10646-1:2000, Amendment #1:** The amendment #1, which adds characters to the BMP (Basic Multilingual Plane) is in FPDAM stage. All indications are that it will be approved. Editor is Michel Suignard, USA.

**ISO/IEC 10646-2:** Part 2 contains additional Latin characters, ancient and rare scripts, additional symbols, mathematical symbols, the plane reserved for identifiers, and additional 40.000+ ideographic characters. The project is currently in FDIS ballot and we expect, it will be approved. Editor is Michel Suignard, USA.

### **473-L TR 10176 - Guidelines for the preparation of programming language standards**

Liaison project to SC22/WG20 - Update of TR 10176 to include internationalization and handling of ISO/IEC 10646 data in programming languages.

The second edition of TR 10176 has been approved in 1999 and was published in 2000.

An amendment to TR 10176 was developed to synchronize the list of characters, which are suitable for identifiers, with the growing repertoire of ISO/IEC 10646. This amendment was approved and the TR is being re-published as a new edition (there are no amendments to TRs in ITTF). Additional synchronization work is needed.

### **495-M 8-Bit ASCII - Structures and Rules [X3/84-1080]**

Maintain ANSI/ISO 4873. No activity.

### **880-L TR 11017 - Framework for internationalization**

Liaison project to SC22/WG20 - Produce a framework document to outline subsequent work on standards and TRs for internationalization of programming languages and application programs.

The project is completed and the document has been published by ITTF in English and French.

### **1037-L Implementation of the 7-bit coded set of its 7-bit and 8-bit extensions on 9-track 12,7 mm (0.5") magnetic tape**

No activity at this liaison project to ISO 962

### **1038-L Implementation of the 7-bit coded set of its 7-bit and 8-bit extensions on 3,81 mm magnetic cassette for information interchange**

No activity at this liaison project to ISO 3275

### **1039-L Coded graphic character set for text communication (revision of parts 1 & 2) Latin alphabet**

Liaison for ISO/IEC 6937:1994

## **L2 Annual Report 2000 - 2001**

The FCD ballot for updates to ISO 6937 to include the euro has concluded, the editor Keld Simonsen (Norway) is working on the disposition of ballot comments and on the next draft. FDIS ballot is under way – US approves.

### **1040-L Arabic coded character set for information interchange**

Liaison for ISO/IEC 9036. No activity.

### **1041-L Standardized coded graphic character set for use in 8-bit codes**

Liaison for ISO/IEC 10367

Now that all 8859-x standards are revised, ISO 10367 will have to be revised also to reflect the correct character names. Nobody has volunteered for this work so far.

### **1042-L Control functions for text communication**

Liaison for ISO/IEC 10538. No activity.

This standard was re-confirmed in 1999.

### **1141-L ISO TR 15285, An operational model for characters and glyphs**

Liaison project for TR 15285.

The project is completed, the TR has been published. JTC1 agreed to make this TR available on the web for free. Based on input from Japan, a new edition of the TR is being prepared. Editor is Ed Hart (USA).

### **1241 Functionality for internationalization (SC22/WG20 project 22.30.02)**

This is an umbrella project for internationalization. It contains IS 14651, TR 14652, IS 15435, and IS 15897. See the details for these projects below:

#### **1241 - L [1307 - L] - IS 14651 - International string ordering**

Produce an international standard for culturally correct ordering and comparison of strings encoded in ISO 10646. ISO/IEC 14651 – International string ordering, has been approved and published on CD-ROM.

Editor is Alain LaBonté (Canada).

WG20 is working on Amendment #1 to synchronize with the rapidly expanding repertoire of ISO 10646. The USA is making substantial contributions – template table from Unicode to ensure compatible sorting with differing algorithms. The amendment has been submitted for PDAM ballot in May 2001.

#### **1241 - L IS 15435 - Internationalization API standard**

A liaison project to specify APIs to access the services defined in the cultural specification standard.

This project is still in working draft state – no ballot has been conducted yet. Editor is Keld Simonsen (Norway).

The USA is opposed to this project.

#### **1241 - L [1244 - L] - IS 14652 - Specification method for Cultural Conventions**

Liaison to a project producing a Technical Report, how to specify cultural conventions such as formats of date, time, currency, etc. Editor is Keld Simonsen (Norway).

This project is presently in DTR ballot, the USA has prepared more than 30 pages of technical and editorial comments. We hope that the ballot will fail.

#### **1275 - L IS 15897 - Procedures for European registration of cultural elements (ENV 12005)**

Maintenance of IS 15897 was assigned to WG20. This standard was fast-tracked from EN 12005. DKUUG is the registration authority. Editor is Keld Simonsen (Norway).

WG20 is considering to adopt similar registration procedures like the ones in the currently being introduced in ISO 2375 to avoid incorrect or incomplete registrations.

#### **1279-M ISO 1073-2: Character set OCR-B, shapes and dimensions of the printed image**

Liaison project for ISO 1073-2

ISO 1073-2 OCR-B will be changed to a TR. Editor is K.I. Larsson (Sweden)

ISO 1073-1 OCR-A was re-affirmed in 2000.

#### **1313 - L ISO/IEC 10646 Part 2: see project [396-D](#)**

The project for part 2 was combined with part 1 of ISO/IEC 10646.

#### **1359 - M Registration of repertoires of graphic characters from ISO/IEC 10367:1991**

Liaison project for ISO 7350. No activity.

## **L2 Annual Report 2000 - 2001**

### **1435 – L Un-punched paper cards – Specifications**

ISO 1681 was re-affirmed in 2000

### **1436 – M Printing specifications for optical character recognition**

ISO 1831 was re-affirmed in 2000

### **1437 – L Punched paper tape – dimensions and location of feed holes and code**

ISO 1154 was re-affirmed in 2000

The following project have been added lately due to a transfer of the international projects from ISO TC46 to SC2/WG3:

### **1515-L Extension of the Latin alphabet coded character set for bibliographic information interchange**

ISO 5426:1983: This standard was transferred from ISO TC46 to SC2/WG3. No action.

### **1516-L Extension of the Cyrillic alphabet coded character set for bibliographic information interchange**

ISO 5427:1984: This standard was transferred from ISO TC46 to SC2/WG3. No action.

### **1517-L Greek alphabet coded character set for bibliographic information interchange**

ISO 5428: 1984: This standard was transferred from ISO TC46 to SC2/WG3. No action.

### **1518-L African coded character set for bibliographic information interchange**

ISO 6438:1983: This standard was transferred from ISO TC46 to SC2/WG3. No action.

### **1519-L Mathematical coded character set for bibliographical information interchange**

ISO 6862: 1996: This standard was transferred from ISO TC46 to SC2/WG3.

L2 recommends re-affirmation and setting the status to "stabilized standard".

### **1520-L Armenian alphabet coded character set for bibliographic information interchange**

ISO 10585:1996: This standard was transferred from ISO TC46 to SC2/WG3.

L2 recommends re-affirmation and setting the status to "stabilized standard".

### **1521-L Georgian alphabet coded character set for bibliographic information interchange**

ISO 10586:1996: This standard was transferred from ISO TC46 to SC2/WG3.

L2 recommends re-affirmation and setting the status to "stabilized standard".

### **1522-L Extension of the Arabic alphabet coded character set for bibliographic information interchange**

ISO 11822:1996: This standard was transferred from ISO TC46 to SC2/WG3.

L2 recommends re-affirmation and setting the status to "stabilized standard".

## L2 Annual Report 2000 - 2001

### 3. Committee Activities

#### a. Previous Years' Meetings (Dates and Locations)

##### 2000 - 2001

# 181	August 8-11, 2000	Boston (Progress Software)
# 182	November 7-10, 2000	San Diego (NCR)
# 183	January 29-31, 2001	Cupertino (IBM)
# 184	May 21-24, 2001	Pleasanton (PeopleSoft)

#### b. Next Year's Planned Meetings (Dates and Locations)

##### 2001 - 2002

# 185	August 7-10, 2001	Redmond (Microsoft)
# 186	November 6-9, 2001	Baltimore (SHARE)
# 187	February 11-14, 2002	Bay Area (Sun or Microsoft)
# 188	May 7-10, 2002	Pleasanton (PeopleSoft)

#### c. Officers

Chair	Arnold F. Winkler (Unisys)
Vice-Chair	Lisa Moore (IBM)
International Representative	Michel Suignard (Microsoft)
Vocabulary	open

L2 will need a new chair in 2002.

#### d. Membership

##### Status: April 12, 2001

L2 has presently 14 members: Apple Computer, Inc.; Compaq; Hewlett-Packard Company; IBM Corporation; Microsoft Corporation; Oracle Corporation; PeopleSoft; Progress Software Corporation; The Research Library Group, Inc.; SHARE, Inc.; Sun Microsystems, Sybase, Inc.; Unicode Inc., Unisys Corporation.

Changes since last year: NCR and Xerox have dropped out.

#### e. Liaison Activities

##### Liaison Representatives to L2:

	Committee	Representative
FCC	Federal Communications Commission	D. Campbell
NISO (Z39)	National Information Standards Organization	S. McCallum
NCITS	NCITS Operational Management Committee	David Thewlis
X3J4	COBOL	A. Wallace
X3V1	Text: Office and Publishing Systems	(open)

##### Liaison Representatives from L2:

	Committee	Representative
JTC1/SC2	Character Sets and Information Coding	M. Suignard
SC2/WG2	Universal Coded Character Set	M. Suignard
SC2/WG3	7-bit and 8-bit Codes	M. Suignard
X3J11	C Language	(open)
X3K5	Vocabulary	(open)
X3V1	Text: Office and Publishing Systems	(open)

## **L2 Annual Report 2000 - 2001**

NISO (Z39)	National Information Standards Organization	J. Aliprand
WG2/IRG	Ideographic Rapporteur Group	J. Jenkins

### **f. Administrative Matters of Note**

None. L2 does not collect or disburse funds.

### **g. Web resources:**

The NCITS/L2 web site is password protected, hosted by Unicode.

Home page: <http://www.unicode.org/unicode/members/L2001/L2.html>

Current document register: <http://www.unicode.org/unicode/members/L2001/L-curdock.htm>

Membership list on June 1, 2001: <http://www.unicode.org/unicode/members/L2001/2001-06-01-SD1.pdf>

## **4. Anticipated Projects**

L2 anticipates continued maintenance work on ISO 8859 and ISO 10646, as well as re-affirmation ballots and recommendations for "stabilization" for older standards.

Significant additional work results from the internationalization standard for sorting and the guide for the development of programming language standards – both are strongly dependent on the rapidly growing repertoire of ISO/IEC 10646. A new format for publication is being investigated.

New projects will probably address character properties.

## **5. Future Trends in Area of Standardization**

ISO 10646 is used increasingly in Java, in XML, in the WWW, and in other Internet standards as the logical character set for world wide use.

Programming languages such as C, C++, SQL, COBOL, and Fortran are now enabling the use of ISO 10646.

## **6. Significant Challenges**

### **TAG for SC2:**

Present challenges are the completion of the work on ISO/IEC 10646-1:2000, Amd. #1 and ISO/IEC 10646-2. Since these standards contain as yet unavailable characters, the creation of high quality fonts, and the printing of code charts are humongous tasks for the editors.

Maintaining synchronization with the Unicode standard which is widely implemented. This is achieved through co-location of technical meetings and strong liaison activities.

Additionally, fending off new proposals for 8-bit character sets present constant challenges.

### **Co-operation with C and C++ committees:**

L2 plans to host a combined meeting of representatives from J11, J16, and L2 to develop a consistent, acceptable model for a data type that allows the use of ISO 10646 data in its UTF-16 format.

### **TAG for SC22/WG20:**

The quest for synchronization of the Unicode collation algorithm with ISO/IEC 14651 – International string ordering, needs constant US input to the WG20 project.

In-depth review and extensive comments to ISO/IEC TR 14652 – Specification method for cultural conventions, and to ISO/IEC 15435 – Internationalization API.

## **7. Expected Challenges.**

### **TAG for SC2:**

Ensure architectural consistency in both parts of ISO/IEC 10646.

Prioritize the encoding of new scripts according to market demands and technical readiness, including fonts.

Political challenges (e.g. North Korea requests the change of names and sequence of characters)

Unreasonable requests for pre-composed characters.

### **TAG for SC22/WG20:**

Ensure high quality registration process definition in ISO/IEC 15897.



## **L2 Annual Report 2000 - 2001**

Deflect POSIX-centric specification method in ISO/IEC 14652.

Prohibit the current draft of ISO/IEC 15435 – Internationalization APIs with its POSIX fixation, from ever becoming a standard.

Arnold F. Winkler (chair L2)