

ISO
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC2/WG2

Universal Multiple-Octet Coded Character Set

(U C S)

ISO/IEC JTC1/SC2/WG2 N 904

Date: 7 June 1993

Title: Unconfirmed Minutes, WG2 Meeting #23 (24-25 May 1993)
in Athens, Greece.

Source: Mike Ksar, Convener and Edwin Hart, Meeting Secretary

Action: Review for confirmation

Distribution: ISO/IEC JTC 1/SC 2 Secretariat and SC 2/WG 2 members
Opening and Roll Call of Participants

The convener, Mike Ksar called meeting 23 to order on Monday, 24 May, 1993 at 10:05. He thanked our host, ELOT for the fine arrangements. Mr. Evangelos Melagrakis, from ELOT, then reviewed the meeting logistics.

The WG2 meeting was attended by 25 people who represented 14 countries. The participants are listed below and the list is ordered by country

Name	Country	Affiliation
Mike Ksar, Convener	USA, ANSI	Hewlett-Packard Co.
V. S. Umamaheswaran		Canada, SCC IBM
Bo Jensen		Denmark, DS IBM
Alaa Ghoneim		Egypt, EOS IBM
Michel Signard		France, AFNOR Microsoft
Paul Detmer		Germany, DINSiemens Nixdorf
Joachim Friemelt		Germany, DINSiemens
Evangelos Melagrakis		Greece, ELOT
Yanis Maistros		Greece National Technical University
Masami Hasegawa, Editor	Japan, JISC	Digital Equipment Corporation
Kato Shigenobu		Japan, JISC
Takayuki Sato	Japan, JISC	Hewlett-Packard Co.
Tae-Jin Kang		Republic of Korea, KBS Han Computer Inc.
Kyongsok Kim		Republic of Korea, KBS Pusan National
Won Sun Lim		Republic of Korea, KBS Ministry of Culture &
Johan van Wingen		Netherlands, NNI Independent consultant,
Sten G. Lindberg		Sweden, SIS - ITS IBM
Jürgen Bettels	Switzerland	SNV Digital Equipment Corporation
Ali Ihsan Baaran		Turkey, TSE
M. Ümit Karaka		Turkey, TSE University of Beytepe

Bruce Paterson	UK, BSI	Independent consultant
Mark Davis	USA, ANSI	Taligent and Unicode
Edwin Hart	USA, ANSI	SHARE, Inc. (Johns Hopkins)
Edwin Smura	USA, ANSI	Xerox Corporation, SC18 liaison, AFII liaison
Arnold Winkler	USA, ANSI	Unisys

Approval of the Agenda

Mike Ksar repeated that the meeting objective is to develop a plan for our work rather than to do technical work.

Republic of Korea, observed an error in ISO/IEC 10646-1 and wanted to add addressing 10646-1 errors to the agenda. Mike Ksar asked for papers on the specifics and also remarked that JTC1 Directives specify a procedure to correct errors in a published standard. The procedure is for countries to submit a defect report to ITTF in Geneva.

J. Friemelt asked if item 5 included tasks and schedule. Mike Ksar responded that the goal was to generate a list of issues, and order them. If time were available, WG2 will also define a schedule.

Takayuki Sato asked that liaison reports be added to the agenda. Mike Ksar agreed to include liaison reports from SC22/WG20, the SC22/WG20 ad hoc meeting, SC18/WG8, and CJK-JRG under other business.

Mark Davis noted the US request for Unicode as liaison and Mike Ksar agreed to discuss it under other business.

J. Friemelt asked if he could give a report from CCITT, TC304, character technology but then decided to wait until the SC2 Plenary to give his report.

Mike Ksar decided to discuss item 6, Review Disposition of Comments, before item 5, Develop a plan to publish future editions of ISO/IEC 10646-1.

The agenda was adopted as modified.

Approval of Minutes of Meeting 22 in Seoul, Korea

Document N871 is the minutes. Mike Ksar asked Bruce Paterson, who took the minutes for Meeting 22, to lead the discussion. Mike Ksar noted that he and Masami Hasegawa had helped to review the minutes for accuracy prior to distribution.

Takayuki Sato requested adding a specific note: "All requests for substantive changes were discussed and rejected." [N889] Johan van Wingen and Michel Suignard agree with the statement but think the change to be unnecessary. Bruce Paterson suggested the above wording and WG2 agreed to add them.

Japan asked for a change on page 5 to add, "For some characters, which are not on national standards, Joe Becker of the U.S. delegation provided source information." [N889]

On page 8, Bruce Paterson noted two typographic error in the third paragraph from the bottom: Replace the "m" with a comma. Take the "e" from "avoide".

Takayuki Sato, under 1.4, suggested a change to "Some national bodies require those characters". [N889]. Mark Davis suggested changing the next sentence to read, "They reflect existing industry practice."

On page 9, Takayuki Sato suggested a change under item 2.(1) to substitute "are" for "be" to read, "They are already implemented." [N889]

K. Kim from Korea asked under 5.14 to substitute "syllables" for "characters". Bruce Paterson commented that all syllables are characters and he would prefer that no change be made. After some discussion, WG2 decided to retain the original wording.

K. Kim from Korea was concerned about using the term "combining" versus "conjoining". Masami Hasegawa noted that "combining" has a special meaning in 10646 and therefore the word was defined in the standard. K. Kim noted that WG2 decided that Hangul Jamos is not combining. He agreed with the minutes but was concerned about the standard as published.

Takayuki Sato [N889] wanted to change the first line on page 19 to read, "before any other use (for example, allocation of characters)". Mark Davis suggested, "use of the O-Zone for swapping will be considered in the process of deciding how the O-Zone is to be utilized." Mark Davis's wording was adopted because it more accurately reflected the WG2 decision.

WG2 approved the minutes as amended.

Dr. Uma then questioned if the unconfirmed minutes were forwarded to SC2. The answer was that the document was SC2 document N2415.

Review Action Items from Seoul Meeting

Character-glyph model, carry over. Send letter to SC18 regarding the character-glyph model. Mike Ksar noted that the US has this action. The US X3L2 and X3V1 technical committees will be meeting in June to discuss the second draft of the character-glyph model. Mike Ksar's action is completed in practice. The US action is incomplete.

Action: US is to prepare a character-glyph model and submit to SC2 and SC18.

Combining Hangul, carry over. It is unclear if the US has started this. The US will follow-up this item. Korea noted that it wanted to participate in such discussions.

Action: US needs to have Korean and US, SC2 and SC18 experts develop a proposal for single writhing system for Hangul.

Tunisia, carry over. Mike Ksar noted that he had not written the letter to Tunisia but that he will write that letter.

Action: The convener is to write letter to Tunisia.

4. SC18 liaison request to add mathematics characters from ISO/IEC TR 9573-13:1991. The response is reflected in the disposition of comments.

5. TC46 liaison request for characters from DIS 5426. Karen Smith-Yoshimura provided the required character shapes to the editor. The response is reflected in the disposition of comments.

Review Disposition of Comments Document

Mike Ksar asked if anyone noted any errors, omissions or misunderstandings in the document.

Turkey is concerned about registration of small subsets.

Netherlands is concerned (1) that the minutes and disposition of comments were distributed late and (2) with some of the changes made to the DIS to form the international standard. NNI accepts that 10646-1 has been published and NNI will submit a defect list to ITTF in Geneva.

Korea questioned how their comments were addressed. Masami Hasegawa noted that the disposition of comments reflects the Korea comments in document N811 because it contained the Korean ballot on 10646. The minutes reflect that WG2 used the updated Korean comments in WG2 document N868.

Develop a Plan to Publish Future Editions of ISO/IEC 10646-1 (Review of 10646 Issues)

WG2 discussed document N887 which was a draft list of 10646 issues. The delegates had two separate discussions and had several comments on the document.

Action: Ed Hart is to revise N887 to reflect the discussions. (done as N887-Rev)

Decision: WG2 will address the issues in N887R (26 May, 1993) in the following order:

Strategic Architectural Issues (item 1.a)

Independent Architectural Issues (item 1.b)

Repertoire Issues (item 2)

Review WG2 Convener's Report to SC2 plenary

WG2 reviewed document N885. Mike Ksar will make the suggested change.

At this time, Greece stated that their vote has changed from no to yes. Mike Ksar asked Greece to notify ITTF of the changed vote.

Action: The convener is to revise N885 to reflect the discussions. (done as N905)

Other Business

Liaison Report from SC22/WG20 (N 890)

This document is informational. SC22 will next meet in September, 1993. If approved by SC22, SC22 will requests SC2 to:

Define a new level "2.5" of 10646 that specifies a repertoire of "valid" composite sequences.

Define "equivalences" of precomposed characters with combining sequences.

Decision: WG2 wait for the formal paper before taking any actions.

SC18/WG8 (N 891 and N 892)

The SC2 term "character symbol" and the SC18 term "glyph" appear to be equivalent. This is to be discussed in development of the character-glyph model.

SC18/WG8 needs to have the character-glyph model by the end of the year. Such a model must include a description of "How can a user need for presentation requirements (a complete repertoire) be met? Ed Smura noted that before any action on a new work item, SC18 procedures require a list of user-requirements. Mark Davis said that the intent of WG2 was to avoid encoding any more presentation forms.

Report of CJK-JRG

Kato Shigenobu reported that CJK-JRG has completed its work to unify East Asian ideographic characters for 10646-1. These characters are encoded in 10646-1. The group has decided to continue its work for future editions.

Proposal to form a special subgroup of WG2

[N888] Since the CJK-JRG had finished the work requested by WG2, Japan requested to form a subgroup of WG2 to address expansion of CJK repertoire. Jürgen Bettels noted that the JTC1 Directives appear to allow WG2 to constitute the special subgroup. Mike Ksar expressed two concerns:

Although the proposal was independent of CJK-JRG, he hoped that the group be open like CJK-JRG and that experts in East Asian ideographs from Hong Kong, Taiwan and the US would be allowed to continue participation in the special subgroup. He hoped that the special subgroup would be composed of experts in the CJK scripts.

China was not present to provide its perspective.

Action: Japan, Korea, and US are to refine the proposal to include its name, its scope and responsibilities, potential membership, etc. [completed as N902]

Mark Davis presented the refined paper, N902. WG2 discussed the paper and made several recommendations that Mark Davis will incorporate into a revised document.

Action: Mark Davis is to update document N902 as agreed.

Action: Korea delegates are to send by fax updated document back to Korea for consideration and comments. Done - No objections.

Action: Mike Ksar is to send by fax a copy of the updated document to China for consideration and comments. Done - Response from Mr. Zhang: personally no objection but will check with national body.

Unicode

Unicode Liaison to WG2 (N 878)

Ed Hart summarized the request to establish a formal liaison between WG2 and Unicode, Inc. WG2 needs to decide to recommend to SC2 to have a formal liaison with Unicode, Inc. Johan van Wingen questioned the value of WG2 making a recommendation. He stated that SC2 makes the decision and WG2 should wait for SC2 to ask for a recommendation. Others commented that WG2 should make some recommendation to guide SC2's decision. Johan van Wingen stated that the Netherlands would abstain from any vote by WG2 to recommend a position to SC2. Most of the participants thought that WG2 should make a recommendation to SC2 concerning the proposed Unicode liaison.

Kato Shigenobu observed that the implementation of a standard was separate from its development. Mark Davis replied that WG2 is trying to define a standard that can be implemented. Mike Ksar noted that the manufacturers and developers need standards that can be implemented. Jürgen Bettels noted that too many ISO standards were defined but not implemented. Having implementations being done in parallel to development of the standard provides a testbed to verify concepts and that a standard can be implemented.

Kato Shigenobu responded that he was worried that individual companies or organizations may also want to become a liaison C to WG2. Mike Ksar said that if an organization can contribute and wants to become a liaison, WG2 will consider it. No other ones have been submitted to date. Dr. Uma noted that X-Open and SHARE Europe had become approved liaisons to SC18.

Turkey asked why could not Unicode participate as part of a national body? Masami Hasegawa commented that the liaison relationship can be established if there are benefits for both parties.

Bruce Paterson stated that the WG2 was worried about the future development of 10646 rather than the existing approved standard. Kato Shigenobu agreed with Bruce Paterson. Kato Shigenobu suggested that Unicode can use its

members' implementation of 10646 to provide an implementer's perspective for extensions to 10646.

Mike Ksar wanted to close the discussion and asked if Kato Shigenobu wanted to make a closing remark. Kato Shigenobu asked that the first two sentences of point 2 in N899 could be removed and Mark Davis agreed to do so.

Action: Mark Davis is to update document N899 as agreed. (done, N899-Rev)

Unicode Intentions with Respect to ISO/IEC 10646-1

Sten Lindberg asked how Unicode was compliant with 10646-1. Mark Davis replied that Unicode version 1.1 implements UCS-2, with the full subset. Later in the discussion, Mark Davis stated that Unicode intends to keep its character repertoire and code assignments synchronized with 10646-1 and future editions of 10646.

Johan van Wingen asked that Unicode terminology and concepts also be consistent with ISO/IEC SC2 guidelines. Mark Davis replied that the point was well-taken. Mark Davis stated that Unicode has not discussed this particular issue. However, Unicode had changed the names of the characters to be the same as ISO/IEC 10646-1.

J. Friemelt was concerned that Unicode was only concerned with implementing level 3 of 10646. Mark Davis replied that the members of Unicode were free to implement up to level 3 of 10646.

Resolution : WG2 recommends to SC2 to establish a formal relationship of liaison, Category C, between SC2/WG2 and Unicode, Inc. and requests that SC2 forward the statements of benefits and responsibilities (WG2 document N899) to JTC1.

(The Netherlands abstained from voting on this resolution.)

Ad Hoc Meeting on Character Sets (with SC2, SC18, and SC22) - N 897.

Dr. Uma summarized the meeting. Presentations were made by:

SYMBOL 183 \f "Symbol" \s 10 \h Mike Ksar (SC2), described ISO/IEC 10646-1

SYMBOL 183 \f "Symbol" \s 10 \h Alan Griffiee (SC18), described the SC18 concerns

SYMBOL 183 \f "Symbol" \s 10 \h Gary Miller (X-Open), described the IEEE POSIX model for internationalization.

SYMBOL 183 \f "Symbol" \s 10 \h Glenn Adams (US X3L2), described possible 10646 implementation issues and suggested solutions

The people at the meeting made these recommendations:

Near term (3 to 5 years)

provide support for 10646-1 in the next version of the programming language standards
address interlanguage communication of 10646 coded-data.
recommend that FSS-UTF be registered within ISO 2375
(ECMA)

Long term solutions

programming languages and supporting environments should provide support for composite sequences and CC-data-elements of 10646 as logical processing units.

need announcement mechanisms for the different encodings.

Consistent terminology

need SC22 to adopt the SC2 terms used in ISO/IEC 10646-1.

the character-glyph model is being developed by SC2 (US X3L2) and SC18 (US X3V1).

internationalization and text processing requirements indicate a need for additional terminology not in JTC1 standards.

Mark Davis said that for several years they were successfully using C and C++ to support 10646. (Although these programming languages may be used, the standard versions lack the native constructs specially designed for 10646.)

Johan van Wingen noted that the ad hoc meeting was controversial from the beginning. He did not go to the meeting because he felt it would not be worthwhile. He commented that SC22 was working to use terms consistently throughout the SC22 standards. Mike Ksar was disappointed that more people from SC22 were not present and that those present were unfamiliar with 10646-1.

Future WG2 Meetings

Fall, WG2 Meeting

WG2 has had two tentative offers to sponsor a WG2 meeting in the fall:

September 20-24 or October 4-8 in Sweden
November 1-5 in Washington, D.C.

Decision: The next meeting will be November 1-5 to allow the US to prepare the character-glyph model document.

Ed Hart noted that the US proposal to host the November meeting was tentative. He needs to obtain approval from ANSI, and that ANSI approval should occur in two to four weeks. Because of the possibility that ANSI might not approve the US proposal, WG2 decided that the first choice is the US, the second choice is Canada, and the third choice is Netherlands.

Action: Ed Hart is to follow-up so that ANSI can formally submit an offer to host the WG2 meeting.

January, SC18/WG8 and SC2/WG2 Joint Meeting

SC18/WG8 offered to sponsor a joint meeting between itself and SC2/WG2 in January in Japan. However, the representatives from Japan were concerned about meeting logistics, and other delegates were concerned about having meetings so close together.

Decision: WG2 will decide WG2 representatives to the SC18/WG8 Meeting at the November WG2 meeting.

Action: WG2 is to decide its representatives to the SC18/WG8 Meeting in January 1994 and notify SC18/WG8 of its representatives.

April, 1994

Decision: With confirmation by Turkey, WG2 will hold its meeting on April 18-22, 1994.

Action: Turkey is to submit a formal offer to host the proposed April 1994 WG2 meeting.

Liaison from WG2 to SC22/WG20

Resolution : WG2 will establish a formal liaison to SC22/WG20 Internationalization Working Group and appoints Takayuki Sato as the SC2/WG2 liaison to SC22/WG20.

Architectural Issues

WG2 started discussing the architectural issues regarding the pros and cons of alternatives for extending the architecture.

Ed Hart presented five alternatives for expanding the repertoire available in 16-bit UCS-2 mode. He discussed the visuals included in document N898 to encourage the delegates to consider these and other alternatives within their national bodies before the next WG2 meeting.

Use combining characters where appropriate
Use UCS-4 for everything
Use UCS-4 for processing but compression for interchange
(storage and communication)
Use Enhanced UCS-2
Use swapping into the O-zone

He was concerned that WG2 take a longer-term view where technology should make UCS-4 feasible in 5 to 10 years. He felt that experience with existing practice with the 7-bit ISO 646 code the resulting garbled text data would likely carry into the fifth alternative. He also felt that the fourth alternative added unnecessary complexity. His conclusion was that implementing either the fourth or fifth alternatives now would cause it to be unwanted legacy software and data in 10 years.

Mark Davis then led a discussion of the advantages and disadvantages of the different alternatives proposed in N898. He noted in describing the enhanced UCS-2 alternative that the 4-octet addresses could be mapped, for example, to contiguous areas of the code space. The discussion quickly centered on the fourth and fifth alternatives.

Bruce Paterson suggested that WG2 try to agree on some very simple objectives and suggested the following objectives:

Need more than 16-bits but much less than 32-bits for encoding.

The need is to encode more characters than code positions are available in the Basic Multilingual Plane (64K characters) but less than the 2,000,000 available with the UCS-4 mode. By how much should we expand the repertoire for the long term?

Upward Compatible with UCS-2.

Developers cannot afford to introduce UCS-2 knowing that customers could face another migration. Customers must understand that WG2 is not introducing the need for another conversion after the one to 10646-1.

Unambiguous Encoding.

Each code position must be complete in itself. A code position must be locally complete in the data stream without outside context. (Each encoded character in a CC-data-element must be uniquely identified by knowing only the 16-bit character-boundary locations before and after the coded-character.)

The fourth alternative was similar to Shift-JIS implementations but appeared simpler (because of unambiguous encoding) than Shift-JIS and simpler than ISO 2022.

One worried about any architecture extensions that create variants.

With this context, WG2 needs to first consider the allocation problem and guidelines outlined in document N884. Then WG2 could size the number of additional characters needed for the first objective. Only then could WG2 consider the architecture issue.

The fifth alternative (swapping) would need some type of announcement mechanism. What if the announcement were lost? This alternative appears to be not compatible with the objective to be upward-compatible with UCS-2.

The enhanced UCS-2 (fourth alternative) appeared to be easier to implement than swapping (fifth alternative).

It is possible, in theory, to implement both the fourth and fifth alternatives.

The need to encode characters outside of the BMP may not occur for several years. So backward/upward compatibility may not need to be implemented for several years.

The enhanced UCS-2 (fourth alternative) could be the compression method used for interchange of the third alternative. That is, it is possible to use UCS-4 for an internal processing code while using extended UCS-2 for an interchange code. This would maintain data compatibility.

The enhanced UCS-2 could be considered as another UTF for compression.

Mike Ksar closed the discussion and solicited papers on this topic before the end of September to be sure they are in the WG2 mailing for the November, 1993 WG2 meeting.

WG2 Resolutions

Resolution 1: WG2 will establish a formal liaison to SC22/WG20 Internationalization Working Group and appoints Takayuki Sato as the SC2/WG2 liaison to SC22/WG20.

Resolution 2: WG2 recommends to SC2 to establish a formal relationship of liaison, Category C, between SC2/WG2 and Unicode, Inc. and requests that SC2 forward the statements of benefits and responsibilities (WG2 document N899) to JTC1. (The Netherlands abstained from voting on this resolution.)

Resolution 3: WG2 endorses the creation of an Ideographic Rapporteur Group (IRG) under WG2 to add to the repertoire of East Asian ideographic characters, subject to confirmation by the National Bodies of Korea and China. The Document WG2 N902 describes the charter and membership of this proposal..

Resolution 4: WG2 thanks AFII for its effort to obtain the fonts and print the code tables for ISO/IEC 10646-1:1993 and looks forward to a continuing good working relationship..

Resolution 5: WG2 thanks the 10646 editor, Masami Hasegawa, for all of his efforts of the last several years in producing the many drafts of 10646 and the final standard. WG2 also wants to thank Digital Equipment Corporation for its support of Masami Hasegawa's work as editor of 10646.

Resolution 6: WG2 thanks our convener, Mike Ksar, for his efforts in leading the final stages of development of ISO/IEC 10646 to become an approved international standard, in keeping the working group focussed on its primary objective, and in keeping ISO/IEC aware of progress and time-table to completion. WG2 also wants to thank the Hewlett-Packard Company for its support of Mike Ksar's work as convener of WG2.

Resolution 7: WG2 thanks our most gracious host, E OT, for arranging the fine support and facilities for this WG2 meeting.

. Resolution 8: WG2 thanks Ed Hart for being the meeting secretary.

Closing

The convener thanked the delegates and experts for their participation, thanked E OT and its staff for thier support during the meeting and for being such a fine host. He then adjourned the meeting on Tuesday at 17:35.

Documents Distributed at Meeting 23

N885 Ksar Report to SC2 Plenary

N886 Ksar Mailing List

N887RHart List of 10646 Issues

N888 Japan Proposal on Formation of a Subgroup

N889 Japan National Body, Japanese Comments on SC2/WG2 N871, Unconfirmed Minutes WG2 Seoul Meeting (June-July 92)

N890 LaBonté, Sato Request for Normalization of 10646 Level 3: Draft Liaison Report from SC22/WG20 to SC2/WG2

N891 JTC1/SC18/WG8 Liaison on Character, Graphic Symbol & Glyph Terminology

N892 SC18 Smura Additions to 10646

N893 Haralambous Khmer Script

- N894 Friemelt Graphic character repertoire of Indo-European languages for inclusion into the BMP of ISO 10646
- N895 Korea Comments on 10646-1
- N896 X/Open File System Safe UCS Transformation Format (FSS-UTF)
- N897 Draft Meeting Report and Recommendations [from the Ad Hoc Meeting on Character Sets, 21-23 April, 1993, Copenhagen, Denmark]
- N898 Hart Take Long-Term View for the 10646 Architecture
- N899RDavis Liaison Request from Unicode, Inc.
- N900 Ksar (reserved for) Document Registry, Updated
- N901 Ksar (reserved for) Mailing List, Updated
- N902RDavis, Kato, Kang Proposal for Creation of an Ideographic Rapporteur Group within SC2/WG2
- N903 Ksar & Hart SC2/WG2 Resolutions of Meeting 23 in Athens, Greece, 24-25 May 1993
- N904 Ksar & Hart (reserved for) Unconfirmed Minutes of WG2 Meeting 23, Athens, Greece
- N905 Ksar (reserved for) WG2 report to SC2 plenary (update of N885)

[End of Document]

Unconfirmed Minutes WG2 Meeting Number 23 in Athens Greece WG2 N 904