

Minutes of the SC22/WG20 meeting Malvern, May 3-7, 1999
--

May 7, 1999

TIME: 9:00 a.m. to 5:00 p.m., Monday through Thursday
9:00 a.m. to 1:00 p.m. on Friday

LOCATION: Unisys Corporation - Customer Center
2476 Swedesford Road
Malvern, PA 19355
USA

1. Introduction and announcements by Convenor

Interesting meeting - I would like to agree on the ordering standard for a new FCD ballot. Everybody is here to help with this effort, CEN, the bibliographic community, Unicode, experts, ...
Kent can't come – we can call +46 31 10 22 44. Küster speaks for GOST on ordering of Cyrillic script. Mark Davis will be here starting Tuesday. Ed Hart will come for Wednesday and so will John Clews - I'd like to clean the agenda from other stuff before we start with ordering.
Party at the Winklers on Wednesday evening.
Many liaison requests in both directions.
Good success with reduction of paper mailings - less than 15 copies needed vs. 80+ before. Of course, many did not answer, especially the NBs libraries.

2. Introduction of national delegations, liaisons, and cooperation partners

Canada	Alain LaBonté	Secrétariat du Conseil du trésor du Québec
Canada	Baldev Soor	IBM
Denmark	Keld Simonsen	RAP
Germany	Mark Küster	University Tübingen, CEN TC304
Japan	Masayuki Takata	Edogawa University
Sweden	Kent Karlsson	IMI *** via telephone for ordering discussion ***
UK	John Clews	Sesame (starting Wednesday)
USA	Mark Davis	IBM (starting Tuesday)
USA	Tex Texin	Progress Software
USA	Ken Whistler	Sybase
USA	Ed Hart	John Hopkins University (Wednesday only)
Convenor	Arnold F. Winkler	Unisys

3. Appointment of chairperson, secretary, and drafting committee

Chair: Winkler
Secretary: Winkler
Drafting: Simonsen, Takata, and Whistler

4. Approval of prior meeting's minutes

597	Minutes from the meeting in Tel Aviv, October 18-22, 1998	Winkler SC22 N2840	98-10-22
---------------------	---	-----------------------	----------

Approved unanimously

5. Future Meeting Schedule and Plans

#18	November 8-12, 1999	Copenhagen	Denmark
#19	May 22-26, 2000	Quebec	Canada
#20	November 2000	Hawaii	USA
#21	Spring 2001		Slovenia ???

Slovenia is possibly not the best choice at this time (Kosovo).

Possibility to go to Hawaii - any interest ? YES, Ken, Alain, ...

Al Winkler – contact Tom Plum and get offers. Drop Balkan for the time being.

Marc Küster offers Tübingen as a possible location for a future meeting.

6. Recognition of new documents and assignment to agenda items

663	Ballot results and comments to 2 nd FCD 14651 (full text version is SC22 N2911 + WG20 N664)	SC22 N2911	1999-04-19	22.30.02.02
664	Swedish comments to FCD 14651 ballot (this text is not included in N663)	SC22 N2912	1999-04-07	22.30.02.02
665	Proposed editorial changes to FCD 14651.2 (proposed changes are shown)	Michael Everson	1999-04-21	22.30.02.02
666	Complete input document for the disposition of comments to the FCD2 14651 ballot	SC22 N2911 & SC22 N2912	1999-04-22	22.30.02.02
667	Registration of Charset and Languages Media Features Tags	Paul Hoffman IETF	1999-04-26	22.30.02.03
668	Thai string collation (input for an informative annex in IS 14651)	Theppitak Karoonboonyanan	1999-04-29	22.30.02.02
669	Excerpt of the resolutions from JTC1 in Rio regarding CLAU1	Winkler JTC1 N5748	1999-04-28	admin
670	Disposition of comments to FCD2 14651 (based on SC22 N2911 and SC22 N2912)	Alain LaBonté	1999-	22.30.02.02
671	Ballot results and comments on FCD 14652	SC22 N2917	1999-04-28	22.30.02.03
672	Disposition of comments to FCD 14652 ballot (SC22 N2917 = WG20 N671)	Keld Simonsen	1999-	22.30.02.03
673	Final agenda for SC22/WG20 meeting in Malvern, May 3-7, 1999	Winkler	1999-05-03	admin
674	Participants at the SC22/WG20 meeting in Malvern, May 3-7, 1999	Winkler	1999-05-07	admin
675	Resolutions from the SC22/WG20 meeting in Malvern, May 3-7, 1999	Winkler SC22 Nxxxx	1999-05-07	admin
676	Minutes from SC22/WG20 meeting in Malvern, May 3-7, 1999	Winkler SC22 Nxxxx	1999-05-07	admin
677	Business plan and convenor's report 1999	Arnold F. Winkler	1999-05-07	admin
678	Swedish annotations to the ballot comments to FCD 14651	Kent Karlsson	1999-05-02	22.30.02.02
679	Interim working document - Table for 14651 (without Canadian delta) == WWW only	Ken Whistler	1999-05-03	22.30.02.02
680	Cultural registry – plan for the revision of IS 15897	Keld Simonsen	1999-05-03	22.15897
681	Default sequence of GOST	GOST	1999-05-04	22.30.02.02
682	Japan's supplementary comments on FCD 14651.2	NB Japan	1999-05-05	22.30.02.02
683	European Ordering Rules, third version	Marc Küster	1999-02-19	22.30.02.02
684	3 rd FCD 14651 – International string ordering for FCD ballot	Alain LaBonté		22.30.02.02
685	Statement of mutual benefits and responsibilities – liaison between W3C and WG20	WG20	1999-05-07	admin

7. Approval of Agenda [N673]

Add new documents to the listings, 680 #14; 678, 679, 681 under #11; 654 also to #11
Agenda is approved with above changes.

8. Liaison Reports

8.1 Additions/deletions/changes to liaisons

8.1.1 W3C would like to establish liaison with WG20. See document N643 from Mike Ksar. We need a statement of mutual benefits.

Keld will write this document for consideration by the group.

8.1.2 ISO/TC46/SC2

John Clews suggests a liaison with ISO/TC46/SC2 - conversion of written languages. N649 is a proposed statement of mutual benefits and responsibilities.

8.1.3 NCITS/J14 – FORTH

I got a request from Peter Knaggs, NCITS/J14, programming language Forth, to consider liaison with this committee in order to support the internationalization of the Forth standards ISO/IEC 15145. I have asked for a statement of mutual benefits and responsibilities and have made some recommendations, where to participate (L2/UTC, WG20).

Generally, no big interest in Forth. Wait for input from Knaggs.

8.2 SC22/WG4, COBOL

No report from Ann Wallace.

Alain reports that COBOL is using the sort standard draft – WG4 has a desire of a named template.

Al Keld: write to Ann Wallace, offer help, ask about the status of I18N in COBOL

8.3 SC22/WG5, Fortran

624	I18N in Fortran 2000	Miles Ellis	98-11-16
---------------------	----------------------	-------------	----------

Küster reports that Fortran programs are used for text processing in his University. I18N is done on a higher level in his environment.

Miles Ellis in his liaison report says that very little is done for I18N in their standard, but Unicode is allowed as a “kind”.

WG20 accepts the liaison report.

8.4 SC22/WG14, C

Keld: C standard has been sent for FDIS.

8.5 SC22/WG15, POSIX

637	Liaison report from WG15 to WG20	Keld Simonsen	1999-01-13
---------------------	----------------------------------	---------------	------------

9945-2b is an enhanced specification in the Locale area. Based on Japanese, Danish, and Swedish comments. Common document between IEEE, POSIX and OpenGroup – combined “Common Standard” for UNIX/POSIX (successor to F1170). **Austin group** is responsible for this work – stability update to existing work. (Go via WG15 web site). Everybody is invited, no fees, open process. New C standard will be integrated. Unclear, how 15897 will fit into this picture.

TR 14644 (guidelines for locales) is being progressed by WG15 and IEEE. It mentions the cultural registry 15897.

CEN 12005 might be withdrawn (Keld).

IS 15897 has been approved, not yet published.

8.6 SC22/WG21, C++

Finalized C++, looking at WG20 API standard, binding is in the works.

Ken: there is no easy binding for object oriented APIs to C and to C. Wants to see the draft.

8.7 SC2/WG2

Ken Whistler and Mike Ksar gave oral report to SC2 in Fukuoka. Ken’s report is part of the WG2 minutes. All open WG2 ballots were accepted, minor changes to Yi radicals, a few name changes, ad few characters added and/or deleted, no new amendments – this allows revision #1 to be processed with clear repertoire. Part 2 will be populated with big chunk of Ideographs, September 99 date is still OK.

Ken: WG2 was not asking for anything, but everything in the final repertoire of revision needs to be considered in WG20, especially for the sorting standard – should eventually include all characters, not just up to amendment 7. Question of character properties is not yet addressed.

8.8 GUIDE/SHARE Europe

What is the new name of Guide/Share? Do we want to continue liaison with this renamed group?

Alain: the group seems not to be interested in I18N any more in practice – no feedback on requests from Alain. It seems that technical issues are not in their work program at this time.

AI Alain: write to Guide/Share and ask, if they are still interested in liaison.

8.9 SC35 – UI

Keyboards and Icons, chair Yves Neuville to be confirmed. Expected problems with drivers' licenses, and other "smart cards". Support of character sets on cards , etc, ...

Ken suggests preparing a TR that provides information to show the shortsightedness of the process that does not cater for internationalization. Examples: passport names, now drivers licenses, ... Can still be ignored, but at least the information is in the open. New item added under other business – 16.3.

8.10 CEN TC 304

Marc Küster: European ordering rules will be discussed in that section. Populating cultural convention registry has high priority, discussions about the format. Not yet released.

OCR-B problem: add euro sign and other characters – testing. No application for the job. Testlab should be found for euro testing in OCR-B.

Transliteration and fallback – general model. TR is suggested for transliteration, and ASCII as fallback for MES 2 which covers Latin, Cyrillic, and Greek.

Keld: Marc is taking over the cooperation, change SD-4. (AI Winkler: update SD-4)

Euro to be added to 13 language locales for EU and EFTA (Pan-European locale)

Workshop is set up in close cooperation with TC304 to collect locale information in all CEN countries (Locale 2000).

8.11 TC37

ISO 639-2 standard is completed. <http://std.dkuug.dk/i18n/iso-639-2.txt>

Working on ISO 639-1 and adding a few more languages.

These 639-2 and 639-1 values go into the cultural registries.

IETF defines to use the 2 letter code when it exists, or the 3 letter code if no 2 letter code exists.

Registration is in Infoterm, Austria, Gerhard Budin.

Sort standard has been stalled, has passed DIS ballot. Waiting for EUR project, if necessary, the sort standard can be revived immediately. Håvard Hjulstad is the editor.

AI Winkler: keep Hjulstad on the mailinglist.

8.12 ITU-T

No report. Not too active in character set work, ISO 6937 needs to be coordinated with ITU-T.

8.13 W3C

643	W3C liaison with SC22/WG20	Mike Ksar	1999-01-18
685	Statement of mutual benefits and responsibilities – liaison between W3C and WG20	WG20	1999-05-07

The required statement of mutual benefits and responsibilities is in N685.

[Resolution]

AI-Winkler: add to N677, convenor's report

AI-Winkler: forward N685 to W3C and SC22

AI-Keld: write statement, mention need for confidentiality in W3C working groups

8.14 Unicode Inc.

645	Voting summary on C Liaison with the Unicode	SC22 N2894	1999-03-09
---------------------	--	------------	------------

	Consortium		
--	------------	--	--

14:0:2:6, AFNOR abstains due to lack of resources

This request has been forwarded to ISO/IEC JTC 1 for their proposal to ISO/IEC ITTF.

No action required, wait for JTC1 approval.

Nominate liaison to Unicode: Ken Whistler is approved as the liaison to Unicode.

8.15 ISO TC46/SC2 - Conversion of written languages

649	Statement of mutual benefits for a request for liaison between SC22/WG20 and ISO/TC46/SC2	John Clews	1999-03-25
---------------------	---	------------	------------

New liaison proposed by John Clews.

Resolution to use N649 as statement of mutual benefits, asking for SC22 approval for liaison with TC46/SC2. Simonsen as liaison.

AI-Winkler: add to N677, convenor's report

Discussion about usefulness of WG20 involvement. Agreement that it can't hurt to get information from TC46 – use the same document for justification.

9. Review of prior meetings action items

SD-5	Action item list	Winkler	1999-04-29
------	------------------	---------	------------

A9711-6: take off the list, make a new one.

AI-Simonsen: Ask OpenGroup for the name of their liaison to WG20

All done except 9810-12 and 9806-4. See below.

A9806-4	Whistler (from Kung)	Suggest word-break API Breaklterator class from Java	open
A9810-12	Whistler	Provide contribution with ideas, what would make the registry for cultural elements valuable for implementors	open

10. Revision of TR 10176

618	Amendment #1 for TR 10176 to correct the list of identifiers in Annex A	K. Whistler, WG20 SC22 N2847	98-10-21
630	Request for subdivision of project JTC1.22.13 to produce amendment 1 to ISO/IEC 10176	SC22 N2846	1998-11-18
631	PDAM registration and ballot for TR 10176	SC22 N2847	1998-11-18
646	Summary of voting on request for subdivision of project JTC 1.22.13 to produce Amendment 1 to ISO/IEC 10176:1998	SC22 N2895	1999-03-09
647	Summary of voting on concurrent PDAM registration and PDAM ballot for PDAM1 to ISO/IEC 10176:1998	SC22 N2896	1999-03-09
648	Concurrent PDAM registration and PDAM approval ballot for PDAM1 to ISO/IEC TR 10176:1998	SC22 N2897	1999-03-11

The requests for subdivision and for concurrent registration and PDAM ballots have been approved, this ballot is under way now.

The publication of the second edition of 10646 will trigger a new review of TR 10176.

Arnold F. Winkler volunteers to be the editor. WG20 agrees.

[Resolution]

AI Winkler: add to N677 for approval by SC22

11. International string ordering ISO/IEC CD 14651

Kent Karlsson has agreed to be available for a teleconference at +46 31 10 22 44, when we need him for discussing the Swedish comments.

Yakupov Emil, director of Paratype, asks that Marc Küster can speak for GOST ordering specifications.

619	Text for 2nd FCD ballot for IS 14651 - ordering Common Template Table (normative 530+kB)	Alain LaBonté SC22 N2844	98-10-20
639	14651 proposal: tentative DTD for the 14651 tables	Kent Karlsson	1999-02-19
640	14651 proposal: non-normative example of tailoring ISO 14651 for Swedish and Finnish	Kent Karlsson	1999-02-19
641	14651 proposal: template collation table - example XML version	Kent Karlsson	1999-02-19
642	14651 proposal: Annex Q - order preserving subkey reduction	Kent Karlsson	1999-02-19
651	Ordering labio-velar characters in Ethiopic	John Clews	1999-03-31
652	Greek ordering (supersedes N605)	Bunz / Küster	1998-11-10
654	Draft UTR #10 - Unicode Collation Algorithm	Davis / Whistler	1998-12-17
662	Contribution on Cyrillic ordering	K. V. Chuguev	1999-04-15
663	Ballot results and comments to 2 nd FCD 14651 (full text version is SC22 N2911 + WG20 N664)	SC22 N2911	1999-04-19
664	Swedish comments to FCD 14651 ballot (this text is not included in N663)	SC22 N2912	1999-04-07
665	Proposed editorial changes to FCD 14651.2 (proposed changes are shown)	Michael Everson	1999-04-21
666	Complete input document for the disposition of comments to the FCD2 14651 ballot	SC22 N2911 & SC22 N2912	1999-04-22
668	Thai string collation (for an <i>informative annex in IS 14651</i>)	Theppitak Karoonyboonyanan	1999-04-29
670	Disposition of comments to FCD2 14651 (based on SC22 N2911 and SC22 N2912)	Alain LaBonté	22.30.02.02
678	Swedish annotations to the ballot comments to FCD 14651	Kent Karlsson	1999-05-02
679	Interim working document - Table for 14651 (without Canadian delta) == WWW only	Ken Whistler	1999-05-03
681	Default sequence of GOST	GOST	1999-05-04
682	Japan's supplementary comments on FCD 14651.2	NB Japan	1999-05-05
683	European Ordering Rules, third version	Marc Küster	1999-02-19

Voting on 2nd FCD shows 10 YES (4 with comments), 6 NO and 2 abstentions.

The new table is on the web only, does not include all amendments, and not the Canadian delta. Discussion about the use of XML instead of the current format. XML is a meta-format, constructed from SGML, and thus allowable in international standards. It could be derived algorithmically from the Unicode table.

Ken: can be generated from the table. Everybody can create the preferred format from the meta-format. Problem seems to be that any drastic change in format triggers many more comments. A change to XML would possibly delay the standard. WG20 decides, not to change to XML at this time.

This is the major discussion about the comments to the ballot:

Alain: It seems that we need to go to a 3rd FCD, there are too many technical comments to satisfy for going to a DIS. General agreement with the possibility to decide after the discussions.

Discussion about the format: Should 14651 be tied to 14652 or not? Should BNF be fixed to meet the comments?

Ken - fundamental issue :

11.1 Deterministic ordering:

How to deal with controls – different weight? Strings with different content must create different key on some level. The only way is to do that on the 4th level in case of canonical equivalence.

Keld says NO to observe canonical equivalences on level 4.

Marc Küster emphasizes that things that look alike must be sorted together.

Long discussion about the conformance requirements of the Unicode collation algorithm (Java, Sybase, etc...). Standard must be flexible enough so that Unicode can be conformant. Unicode collation algorithm does not specify a 4th level of keys. 14651 should not require the 4th key, but can allow it.

Ken: the specification of the number of levels is NOT for the conformance clause, rather a given input has to produce a given output sequence for the implementation to be conformant.

Mark Davis arrived at this point and agrees with Ken – 3 level sorting keys fit into a 32 bit chunk for fast processing.

Keld: for normal linguistic processing 3 levels are OK. But there should be different conformances to 3 or 4 levels...

Agreement: 3 levels are minimum, more can be needed for specific profiles, etc... (nobody is sure about Burmese, e.g.).

The conformance clause will not require the complete repertoire of 10646 (to keep table sizes small).

Long discussion about deterministic ordering if only 3 levels are used. Danish requirement for fully deterministic ordering can be addressed in 2 ways – culturally or binary. This is a question of the interpretation of the scope of the standard.

There are 4 possibilities:

- 1 Leave status quo
- 2 4th level specials are UCA-L1 and 4th level non-specials are PLAIN
- 3 4th level specials are CHAR - UCS-CODE and non-specials are PLAIN
- 4 version 3, but 1 bit indicating PLAIN vs the Unicode value

Option 3 is tempting – makes result almost deterministic and meets Kent's requirement. PLAIN is a weight less than all other weights on that level.

Full deterministic order requires 5th level with UCS-4 code point (weight) as value.

The common understanding is to leave it as it is and use position. We might also provide a tailoring table or a character list for an option 3 implementation.

Discussion with Kent Karlsson will continue on Thursday at 9:00am over the telephone.

Marc: informative annexes should use the syntax as per formal standard.

Ken: tailoring samples such as Canadian order, Danish order, upper before lower case, etc... should be included in an annex.

Kent: added additional weight for ligatures to eliminate the problem.

Keld asks about deterministic of Arabic ligatures. It turns out that level 4 can be changed to contain the UCS code point instead of the current content.

Final Agreement: status quo with change of contents in level 4 (UCS code).

11.2 XML format for the standard:

Davis: it would be good to specify the standard in XML format, but it would delay the standard for too long. Perhaps in a TR or in a revision of the standard.

Ken: can go with this solution. Adjusting the Unicode collation algorithm to XML would also be a long job. And it must be tested very carefully.

Agreement: XML is a great idea, not at this time, but for either a future revision or for a Technical Report or as a publicly available paper on the web for implementers.

11.3 Position parameter (as per US comments)

Alain: acceptable to me.

US would like to use the right weights and sort backwards. The position parameter should not be "required" (like the 4th level).

Davis: mix of direction (e.g. accents in mixed language text) is very complex, for humans especially hard to understand. Uniform direction on a particular level would be preferred. It only applies to shared characters in different languages. Sort order is mainly to meet dictionary order.

Alain: French would be for French dictionaries backwards. Second level must allow backwards sorting.

Ken: what about backwards on the 1st level? Should be done by pre-handling, while building the keys. Position parameter is optional and only supplied on the last level.

11.4 Mix of forward and backward on the same level (as per US comments)

Alain: Egypt has this requirement in IBM implementation (back, back, back).

Davis: 14651 is essentially for Unicode/10646. For legacy character sets, pre-handling might be required (transformation, ...) before the ordering is applied. Generalize the part about 2022 type character sets.

Agreement: Use pre-handling for legacy character sets. Scanning direction is optional.

11.5 BNF Syntax – what BNF do we use ?

We studied the final text for ISO/IEC 14977, Extended BNF. We compared use of various forms and N619 and came to the following agreement for the use in 14651 and in 14652:

Terms or general rules explained	< >
Quoted literal	‘ ‘
Optional term (zero or one)	X?
adjacency	XY
or adjacency	X Y
One or more	X*
Grouping	()
Quoted terminal	N/A
Terminated lines	;
Production rule	=
Comments - in a separate line	%

Keld would use the same in 14652. Ken will use it in 14651.

11.6 Meta-Syntax:

Done in the BNF section (see 11.5)

11.7 Fractional weights:

Kent suggests using fractional weights for easier sub-key reduction, values are between 0 and 1. All numbers are with 16 decimal positions. Ken thinks that the technique is not fully described.

1. Complete ordering key --- list of numeric weights

2. A complete ordering key (COK) x is less than a complete ordering key y if and only if there is some integer i such that all (i-1) subkey in x and y are equal and the ith subkey of x is less than the ith subkey for y or the length of x is i-1 and the length of y is i or greater.

COK x greater than COK y iff y less than x

COK x equal COK y iff neither x<y nor Y<x

COK → subkey → weight

Move 6.1.2 after 6.2.2 to make the point clear.

New: preparation 6.1.1; preliminary cond. 6.2.1, key formation 6.2.2, comparison 6.2.3; BNF 6.3

Alain produced a temporary document for the re-write of clause 6.

Fractional weights are not yet decided.

Agreement: Fractional weights will not be used at this time.

11.8 Alain works off the document N670 and goes through the individual comments:

The numbers in front of the comments are from N670. The numbers for the comments discussed refer to the comment in document N670.

11.8.1 Canadian comments:

Ken recommends that the Canadian tailoring be a (very simple) tailoring of a “plain” table. IBM has implementations that are based on the Canadian standard. Alain agrees with the desire to have a totally plain table and a bigger Canadian delta.

Agreement !

11.8.2 Danish comments:

#3: Format need to be U-xxxxxxx for the 8-digit specification. Keld agrees.

#4: Escape sequences need pre-handling. C0 characters are present with names and codes, C1 characters are present but have no name. Some characters have no names specified in ISO 6429. Keld agrees with e.g. ESCAPE (in 6429).

#5a: basic letters vs. first-level letters . In Denmark, some accented letters are actually basic letters. Agreement: ... of the basic letter of the alphabets of that script ...

#5b: BNF for 14651 and 14652 should be the same. This is a major issue (see above for discussion).

#6: OK

#7: meta syntax needs to be discussed as a general issue.

#10: one or more space characters as per C language

#11: withdrawn

#14: 15897 has naming rules for languages/region. Ken points out that this table name has to be in 14652 and conform with the guidelines of 15897.

#16: self contained, Ken will work with Alain for details.

#17: put that in annex D (tutorial). Keld will provide text and location in annex D. Do not describe in detail how to do it – rather say that it can be done (see 6.17 of UTR).

#18: Option on the last level for backward compatibility. Should be written to the Austin group.

#19: not needed in 14651, allowed in 14652.

#20: conformance: English by Everson edits, consensus is needed on conformance clauses.

11.8.3 French comments

OK

11.8.4 German comments:

5.2.1 remove paragraph. New BNF will be added later as a TR or in a revision of the standard.

5.2.3 Marc Küster will provide text

5.2.4 accepted in principle

5.2.5 note is not harmful, leave there

5.2.20 5.2.8 Reference comparison added; for the method to establish an order between 2 character strings see 6.1

5.2.9 text was prepared in meeting

5.2.11 Marc will provide text

5.2.12 Mark Davis will suggest text to replace the entire paragraph (“pre-composed” needs 10646 word)

5.2.13 text developed in meeting

5.2.14 sharp s as variant of ss – Marc to provide examples in text

5.2.21 remove C3

5.2.22 make it lower case

5.2.23 change example to articles in book-names (context)

for example, part names containing date without spaces ... add to 3rd paragraph: there are

circumstances where digits do not represent numerical values, such as part numbers; these cases are not discussed below. Remove 2nd paragraph.

5.2.24 Add to warnings. Caution: preparatory steps have undesirable consequences such as the ordering of telephone books in telephone books, and should be avoided in such cases.

5.2.25 agreed in principle

5.2.28 remove VII

5.3.1 Ken made short names by tacking on (weight-)numbers to D for symbols. This created cryptic names, but consistent ones. 3 possibilities – leave as is, make all numbers, NO CHANGE

5.3.2 Variant letter shapes have been discussed earlier
Relative order of scripts is partly implicit, no change.

5.3.3 Greek – postpone until Clews is here.

5.3.4 Cyrillic: GOST has responded and asked Küster to represent them. He carried documentation with him, and we got N681 with supporting documentation (distributed to participants). There are a few characters that have not yet been accepted in 10646 – they will be excluded from the ordering. The proposal matches the EOR proposal (European Ordering Rules) from the CEN/TC304 work. Differences are in letters with diacritical marks. The new table introduces challenges due to canonical ordering and the contraction of more than one character to one weight. The current BNF does not allow these contractions, must be added ! Add a small list of characters to the table.

5.3.5 Georgian: no additional information. Based on input from Everson and Clews implemented in table from Ken.

Wednesday, May 5:

Takata provided a supplement to Japan's comments to 14651.2. The doc# is N682 and Winkler asked that the comments be considered in the disposition of comments, as if they had been submitted with SC22 N2911.

Alain has produced a temporary document based on the discussion from Tuesday, the group reviewed it and recommended a few comments of mainly editorial nature.

Marc Küster provided the third version of the "European Ordering Rules" , it is now document N683 for distribution to WG20.

At 9:00am we called Kent Karlsson to allow him to contribute to the discussion of the Swedish comments.

11.8.5 Swedish comments:

Discussion on definitions. Might be continued on e-mail.

9.1 Ordering-key, collation element, etc... Anything defined in 10646 should not be re-defined in this standard (graphic character e.g.). "Preparation" does not need to be defined. A new definition of weighting table is needed (mapping from collating elements to weighting elements).
Symbolic collating element instead of symbolic collation item (also in explanation).

9.2.1 Table well formedness. This would support ISO 12999 method of ordering based on private character as tags for formatting. This NOT in the scope of this standard. This change would make the Unicode tables very different and create a lot of work. For now, we will not make the suggested change. Should be included in the future XML version.

9.2.2 accepted in principle.

9.3 accepted in principle, see new document. Position can not be removed due to existing implementations which should be conformant to the standard.

No new annex is planned for compression, so that the next ballot does not produce additional comments in the FCD ballot to follow.

- 9.4 New text
- 9.5.4 nothing new
- 9.5.7 PLUS and MINUS are currently specials; should not be changed.
- 9.5.8 Küster: do not change.
- 9.6.9 taken care of in the table
- 9.6.10 TONOS and AIGUT have the same weight correctly. Comment not accepted.
- 9.6.11 not accepted
- 9.6.12 discussed, correct.
- 9.6.13 Variants don't help, create difficulties in auto-generation of weights. Not accepted.
- 9.6.14 Ligatures as 2 or 1 ? Kent suggests to define a variant. Comment rejected.
- 9.7.15 Interesting problem, Arabic ligatures require special treatment. Should not be used in the first place, very subtle ordering impact, IBM's codes of preformed characters are not affected. No action.
- 9.7.16 Ordering of dingbats with one or two numbers. Presently they have different weights at level 4. No fix required.
- 9.7.17 Squared ligatures: could be treated like compat weights. Leave as is.
- 9.7.18 Capitals before small letters: Ken objects strongly, complexity would be introduced into level 3. Change can be done by tailoring – example will be given for that case. Not by default.
- 9.7.19 Not accepted.
- 9.7.20 The NONBREAK and VERTICAL are used in the UCA, not needed in 14651 table.
- 9.7.21 Not accepted, fonts are not a ordering criteria.
- 9.7.22 Not accepted
- 9.7.23 Is fixed.
- 9.8.24 Ignore C0 and C1 – Ken will do (BN means boundry neutral). Mark: these characters should not be distinguished for collation, only for display. Accepted.
- 9.8.25 Long discussion about deterministic ordering if only 3 levels are used. Danish requirement for fully deterministic ordering can be addressed in 2 ways – culturally or binary. This is a question of the interpretation of the scope of the standard.

There are 4 possibilities:

- 1 Leave status quo
- 2 4th level specials are UCA-L1 and 4th level non-specials are PLAIN
- 5 4th level specials are CHAR - UCS-CODE and non-specials are PLAIN
- 6 version 3, but 1 bit indicating PLAIN vs the Unicode value

There was almost agreement for option 3 – makes result almost deterministic and meets Kent's requirement. PLAIN is a weight less than all other weights on that level. Full deterministic order requires 5th level with UCS-4 code point (weight) as value.

The common understanding is to leave it as it is and use position. We might also provide a tailoring table or a character list for a option 3 implementation. Discussion with Kent Karlsson will continue on Thursday at 9:00am over the telephone.

Thursday:

Mark Davis reports about the 5 options we discussed Wednesday afternoon. There was no consensus on any of the options – this leaves us at status quo – gives UCS determinacy (not for legacy character sets).

Marc: informative annexes should use the syntax as per formal standard.

Ken: tailoring samples such as Canadian order, Danish order, upper before lower case, etc... should be included in an annex.

Mark Davis: Watch for Arabic compatibility characters that decompose in elements with similar weights – in addition to the 500+ characters which are "duplicates" with their elements' weights. Special treatment would be needed to eliminate this problem – the value might not be big enough for the effort.

Kent: added additional weight for ligatures to eliminate the problem.

More discussion on options and their values and pitfalls.

Keld asks about deterministic of Arabic ligatures. It turns out that level 4 can be changed to contain the UCS code point instead of the current content.

Agreement: status quo with change of contents in level 4.

9.9 Quality of examples. New examples will be prepared to show Canadian and Danish deltas, etc... The examples will be corrected.

Covered repertoire for Denmark does not contain combining characters. Keld will check with DK National Body about the repertoire with or without combining characters.

We thanked Kent Karlsson for his time to ensure that his comments are properly addressed and understood. Signing off a 9:50am.

11.8.6 Irish comments

6.1.1 English language - YES

6.1.2 OK

6.1.3 Upper before lower – NO

6.1.4 OK in principle

6.1.5 Gaelic examples are there

6.1.6

6.1.5 Will be aligned up to Amd. 9, equal to Unicode 2.1 (scope, paragraph 3)

11.8.7 Japanese comments

7.1.1 NO

7.1.2 YES

7.1.3 Accepted in principle (Aip)

7.1.4 Reference will be changed, full UCS format will be allowed

7.1.5 OK

7.1.6 Together with ...

7.1.7 Name will stay

- 7.1.8 Accepted in principle
- 7.1.9 In conformance clause we will use “process” – this accommodates the comment in part.
- 7.1.10 OK
- 7.1.11 ... shall **reference** a declaration ...
- 7.1.12 OK

- 7.1.15 accommodated
- 7.1.16 Accepted in principle: do top down vs bottoms up.
- J15.2 depends on the Canadian delta, only the can we name it the Common Template Table.
- J15.3 Accepted
- J15.4 accepted
- J15.5 accepted
- J15.6 new text developed in 6.3.3, I6.
- J15.7 Ken will correct the BNF.
- J15.8 accepted
- J15.9 no description needed; rejected.
- J15.10 accepted
- J15.11 accepted
- J15.12 accepted
- J15.13 accepted
- J15.14 remove from the BNF
- J15.15 ... or symbol_definition ...
- J15.16 accepted, Ken will do BNF
- J15.17 POSIX allows that. Comment rejected.
- J15.18 agreed in principle, Ken will do that.
- J15.19 problem of identity; rephrase to ... the tailored table ...
- J15.20 Question of 2 re-order statements: complete one, then do the next. Accepted, Ken promises text for the note to I2 for clarification.
- J15.21 Multiple occurrences of the same symbol is not allowed. A wellformedness condition will be added by Ken to the BNF. “Reorder after” is also used to introduce new symbols. The term “reorder” is misleading for that case, “put” would be better.
- Ken: Add a well-formedness condition to the BNF, make text for note to I2, add condition for reorder after to include new symbols.
- J15-22 withdrawn
- J15.23
- J15.24 is not true, and BNF corrects possible occurrence
- J15.25 identifiers starting with numbers must be allowed. Ken will add well-formedness condition for such expression generically.
- J15.26 . precise definition must be written for ranges – Ken will do; accepted

- 7.1.17 Text by Alain. Accepted.
- 7.1.18 Agreed (will have “collating-element”)
- 7.1.19 Toggles are a shortcut for tailoring. Will not be added to 14651, can be done by tailoring.
- 7.1.20 Conformant is acceptable

Additional comments in N682

11.8.8 UK Comments:

- 10.1 Editorial comments, eliminated by Irish edits
- 10.2 Accepted in principle – not all was corrected by Everson
- 6.1.1 accommodated by edits
- 6.2.2.2 will be reworded, characters which have a weight assigned in level 4
- 10.3 Ken is not in favor of changing the meta-syntax to EBNF. Will be discussed in general.
- 10.3.1 Withdrawn by UK.

10.3.2 Ken shows the implications and asks, if this change had the approval of all (present) NBs. This seems to be the case, Ken will make the change. Combining comma above and reverse comma above will be moved in front of the accent acute. Also, Perispomeni will go before Diacesis. Last change of that part of the table.

10.3.3 not accepted

10.3.4 NO.

10.3.5 Not accepted.

10.3.6 Differences between EOR and 14651 will be ameliorated by EOR annexes.

10.3.7 Not accepted.

10.3.8 Text will be added to Annex A table. Note about ordering of ideographs will be provided. Additional ways exist, for specific usage.

10.3.9 Default order should be predictable in order of scripts. UCS does not specify order of scripts. UK hopes to see a specific order – origin of scripts in East-to-West direction. Symbols are not counted, just written scripts. Rejected. Clews is encouraged to write a liaison statement that specifies the request for script ordering.

Conformance:

6.2.2.2 Standard must require 3 levels minimum, that does not preclude the possibility for more levels through the tailoring statement. More text for the levels and parameters was added to the conformance statements.

11.8.9 US Comments:

11.1 Discussion about conformance, re-formulation.

11.2 Accepted

11.3 Accepted (no text addition from paragraph 4 of the US comments is needed)

11.4 D1 and D2 are ok.

Discuss D3: Alain postulates that such cases (irrelevant combinations, such a A with cedilla below and ring above) can be done via pre-handling, but the pre-handling must know about tailoring.

Discuss D4: exclude unsupported characters or sort them in code-point order? D4 can be handled by the current standard with the changes accepted in this meeting.

11.5 accepted

11.6 comment withdrawn, have been accommodated.

11.7 OK, Mark will provide location

11.8 Accepted due to Japanese comments

11.9 Accepted due to Japanese comments, Ken will attempt to provide text.

11.8.3 BNF. See discussion and agreement in general discussion 11.5

11.8.4 Accommodated by changes in Malvern and editing by Everson. Introduction being wordsmithed in meeting.

11.8.5 Accommodated

11.8.6 Discussion of the use of “weight” and “token”. It seems that they mean the same – replace all “token” by “weight”.

6.2.1.1 Alain has text for correction

6.3.4 accepted

11.9 Editorial stuff

11.9.1 Tex and the group discuss the issues identified in the US comments and Alain edits the document directly. A new draft is distributed for review over night and discussion tomorrow.

11.8.10 Netherlands comments:

The group discussed the NL comments and Alain edited the disposition document accordingly. Many comments had been addressed during the week and were not valid any more.

Marc Küster confirms that the 3rd FCD14651 will cover all GOST requirements.

11.9 Progression:

3rd FCD is necessary !

May 14, 1999	LaBonté	Send disposition of comments (N670) to the web for comments and review by WG20
May 19, 1999	WG20	Deadline for comments to the disposition of comments
May 21, 1999	LaBonté	Send new draft to the web (N684)
May 28, 1999	WG20	Deadline for comments to the 3 rd FCD 14651
June 1, 1999	Winkler	Submission to SC22 of N670 and N684 for ballot

[Resolution]

12. Cultural convention-specification standard ISO/IEC CD 14652

634	FCD 14652 - Specification method for cultural conventions - for FCD ballot	Simonsen SC22 N2869	1998-12-21
638	ISO 3166-2 in relation to IS 14652 and IS 15897	John Clews	1999-02-16
671	Ballot results and comments on FCD 14652	SC22 N2917	1999-04-28
672	Draft disposition of comments to FCD 14652 ballot (SC22 N2917 = N671)	Keld Simonsen	1999-05-06 (draft)

Voting on 2nd FCD shows 8 YES (1 with comments), 4 NO and 3 abstentions.

Friday, May 7, 1999

Keld points out that many NBs did not vote or abstained. The number of comments is much lower than in the ballot for the first FCD.

Based on N672, Keld leads through the proposed disposition of comment.

Discussion about the usefulness and the application of POSIX methodology for internationalization. **USA** is not only not supporting, but even opposed to it .

Clews suggests to change the project to a TR – it is easier to approve and update is needed after 3 years.

Keld agrees that this is a route that could overcome the obvious opposition to the standard. **Clews** expects that a TR would get more input from the NBs and more support.

Germany was against the work item from the very beginning. Germany had no rapporteur for WG20 and thus abstained for a long time. Now, with Küster representing Germany, it opposes the work item, a TR would be acceptable.

Keld considers 14652 as the pivotal standard for I18N and opposes turning it into a TR.

Soor shows comparison with Java's implementation of I18N – make the information available to more than POSIX based systems without the need to "copy" the format. Points out the need for a "consistent cultural data information" in a company.

Keld: we could do that – TR with example syntax.

Ken: character set work should be done in SC2 (e.g. properties)

Marc: Example syntax is not necessary for Germany, what is needed is the collection of cultural conventions.

Keld points out that this might already be in IS 15897.

Marc: TR with sample syntax, taken from 14652, could be an option.

Ken: give it to the UNIX/POSIX community for completion. Same for API standard.

US agrees with all comments from Japan, except for J22 and J73; all German, and all Swedish comments.

WG20 will make 14652 a TR (Winkler to check the possibilities with the SC22 secretariat)

12.1 Progression:

May 21, 1999	Simonsen	Send disposition of comments (N672) to the web for comments and review by WG20
May 26, 1999	WG20	Deadline for comments to the disposition of comments
May 28, 1999	Simonsen	Send new draft to the web (N684)
June 4, 1999	WG20	Deadline for comments to N684
June 8, 1999	Winkler	Submission to SC22 of N672 and N686 for ballot as PDTR

13. Internationalization API standard ISO/IEC 15435

This project was not addressed at the meeting in Malvern, May 3-7, 1999.

14. Registration of cultural elements ISO/IEC 15897

680	Cultural registry – plan for the revision of IS 15897	Keld Simonsen	1999-05-03
-----	---	---------------	------------

This document describes the plan for the progression of IS 15897.

The US request that the standard will be revised not as an amendment, but through the FCD route. We need to ask SC22 to approve concurrent registration and FCD ballot, when the standard is ready.

[Resolution]

AI-Winkler: add to N677

15. ISO/IEC 10646 Issues

This agenda point was not addressed at the meeting in Malvern, May 3-7, 1999

16. Other business

16.1 Business plan and convenor's report 1999

677	Business plan and convenor's report 1999	Winkler	1999-05-07
-----	--	---------	------------

[Resolution]

AI-Winkler: submit to SC22

Discussion about the documents on other servers. We have to deal with the fact – version control is needed. Sometimes it might be necessary to keep the original document on our web site, while still pointing to the latest on the host site. Legal implications need to be explored.

16.2 Cultural and Linguistic Adaptability and User Interfaces (CLAUI)

625	Identification of CLAUI related areas	France, JTC1 N5608	1998-11-19
626	Draft guidelines to address cultural and linguistic aspects in International Standards	France JTC1 N5609	1998-11-19

627	The BSR: a tool for cultural adaptability	France, JTC1 N5610	1998-11-19
628	Elements of reflection on cultural and linguistic adaptability of products and standards in the field of information and communication technologies	France JTC1 N5611	1998-11-19
633	Report from the CLAU meeting in Paris, 12/98	SC22 N2862 JTC1 N5629	1998-12-07
669	Excerpt of the resolutions from JTC1 in Rio regarding CLAU	Winkler JTC1 N5748	1999-04-28

The text in the NP form is changed so that I18N issues are addressed in the NP. Mode of operation in the DT is 3+, the 3 committees (SC22/WG20, SC35, and SC2) are to present a common report to JTC1 and that these committees need to work with each other in pursuit of I18N. Conveners must co-operate and meet regularly to coordinate the program of work (e.g. who is going to do character properties?).

16.3 New work item for a TR about the dangers of disregarding I18N.

AI-Winkler: put on agenda for the next meeting in Copenhagen

Winkler asks WG20 for contributions and to contact liaisons for additional input.

17. Review of Priorities and Target Dates

Reviewed and approved.

18. Review of Actions Items from this meeting [SD-5]

Action items are approved as per SD-5

19. Approval of Resolutions [N675]

N675 is approved.

20. Adjournment

Meeting is adjourned