Minutes from the SC22/WG20 meeting #18
Minutes from the SC22/WG20 meeting #18 **Copenhagen, November 8-12, 1999**

November 17, 1999

DATE: November 8 - 12, 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: DANSK STANDARD
                   Kollegievej 6
                   2920 Charlottenlund
                   Denmark
                   Tel: +45-39-96-61-06
                   Fax: +45-39-96-61-02
                   E-mail: DANSK.STANDARD@ds.dk

AGENDA ITEMS:

1. Introduction and announcements by Convenor

Welcome.
Highest priority is IS 14651, followed by IS 15897, and all other subjects on the agenda, specifically TR 14652.

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

Canada	Alain LaBonté	Secrétariat du Conseil du trésor du Québec
Canada	Baldev Soor	IBM
Norway	Keld Simonsen	RAP
Germany	Marc Küster	University Tübingen, CEN TC304
Japan	Masayuki Takata	Edogawa University
Sweden	Kent Karlsson	IMI
UK	John Clews	Sesame
USA	Ken Whistler	Sybase
Convenor	Arnold F. Winkler	Unisys

3. Appointment of chairperson, secretary, and drafting committee

Chair   Winkler
Secretary   Winkler
Drafting committee   Simonsen, Takata, LaBonté

4. Report from the SC22 plenary in Berlin

701	Resolutions of the SC22 plenary in Berlin	SC22 N3013	1999-09-23	admin
702	Resolutions from the SC22 plenary with impact on the work of SC22/WG20	Winkler	1999-09-23	admin

Keld reports about the ad-hoc group on participation in standards work in SC22 working groups. Arnold invites contributions, how participation can be increased.
The DoC (Japan issue) will be discussed in 18.1. – read document N697

5. Approval of prior meeting's minutes

676	Minutes from SC22/WG20 meeting in Malvern, May 3-7, 1999	Winkler SC22 N2924	1999-05-07	admin

Keld: 8.5: replace "Swedish" by "Canadian" , also 14766 is the guidelines for locales. Lines about EN 12005 should be moved to CEN report section.

8.4: mentioning legacy character set is incorrect, we speak about repertoire, not coded character sets.

11.8.2: not "basic" letter, but "common" letter of the alphabet…

Keld suggests, that minutes are provided for comments by WG20 before they are distributed to SC22. Timeframe for comments will be 2 weeks.

Approved with changes indicated above.

6. Future Meeting Schedule and Plans

#19	May 22-26, 2000	Quebec	Canada
#20	Oct. 30 – Nov. 3, 2000	Unisys (MV or Malvern)	USA
#21	May 7-11, 2001	Tübingen	Germany
#22	Autumn 2001	Asking for hosts	

Hawaii is out – no enthusiastic support

Arnold is going to check Mission Viejo and/or Malvern

Spring 2001 in Tübingen

Arnold invites proposals for October/November 2001 .

7. Recognition of new documents and assignment to agenda items

693	Internationalization of the File Transfer Protocol	IETF RFC 2640	July 1999	I18n
694	Sorting order of Phillipino Ng sign	SC2/WG2 N2039	1999-06-22	22.30.02.02
695	Notice and preliminary agenda for SC22/WG20 meeting in Copenhagen, November 8-12, 1999	Winkler SC22 N2979	1999-08-30	admin
696	Japan Member Body Request for Discussion of CD/FCD Processing	NB of Japan SC22 N2993	1999-09-09	admin
697	WG20 convenor's remarks to the concerns of the NB of Japan in SC22 N2993	Winkler	1999-09-21	admin
698	Disposition of comments to TR 10176 registration and PDAM ballots	Winkler	1999-10-08	22.13.01
699	TR 10176 Amendment #1 for DAM ballot	Winkler	1999-10-22	22.13.01
700	Document register N651 through N700	Winkler	1999-10-08	admin
701	Resolutions of the SC22 plenary in Berlin	SC22 N3013	1999-09-23	admin
702	Resolutions from the SC22 plenary with impact on the work of SC22/WG20	Winkler	1999-09-23	admin
703	Swedish late comments to FCD 14651 ballot	Karlsson	1999-10-20	22.30.02.02
704	Annual report of the NB of Ireland for 1999	SC22 N2986	1999-09-02	admin
705	Summary of voting and comments to PDTR 14652	SC22 N3024	1999-10-13	22.30.02.03
706	Disposition of comments to PDTR ballot 14652	Keld Simonsen		22.30.10.03
707	Final agenda for the WG20 meeting in Copenhagen, November 8-12, 1999	Winkler	1999-11-08	admin
708	Participants in Copenhagen, November 1999	Winkler	1999-11-08	admin
709	Resolutions from the WG20 meeting in Copenhagen, November 1999	Winkler	1999-11-12	admin
710	Minutes of the WG20 meeting in Copenhagen, November 8-12, 1999	Winkler	1999-11-12	admin
711	Summary of voting and comments to FCD 14651	SC22 N3025 (txt)	1999-10-25	22.30.02.02
712	New Working Draft for IS 15897	Keld Simonsen	1999-10-29	22.15897
713	Report of the convener, resolution of comments meeting on DIS 15897	Bob Follett SC22 N2811	1998-09-02	22.15897
714	New committee draft for ISO CD 15435	Keld Simonsen	1999-11-07	22.15435
715	Report to Linux I18N group about WG20 work	Keld Simonsen	1999-11-06	admin
716	Linux I18N group suggestion for liaison	Akio Kido	1999-11-06	admin
717	Browsing and matching - scoping	Marc Küster	1999-11-05	22.30.02.02

718	Late Norwegian comments on 3rd FCD 14651	Simonsen SC22 N3032	1999-11-07	22.30.02.02
719	ISO/IEC 15897- Procedure for registration of cultural elements (web only distribution)	ITTF	1999-10-15	22.15897
720	ISO/IEC FDIS 12199 – Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet	ISO/TC37/SC2	1999-09-28	22.30.02.02
721	Liaison statement from WG20 to ISO TC37	WG20	1999-11-10	admin
722	Issues list #3 for ISO 15435	Simonsen		22.15435
723	Issues list #1 for ISO/IEC 15897	Simonsen		

8. Approval of Agenda

707	Final agenda for the WG20 meeting in Copenhagen	Winkler	1999-11-08	admin

Agenda as amended is approved, 18.3 has been added (CLAUI)

9. Liaison Reports

9.1. Additions/deletions/changes to liaisons

715	Report to Linux I18N group about WG20 work	Keld Simonsen	1999-11-06	admin
716	Linux I18N group suggestion for liaison	Akio Kido	1999-11-06	admin

Keld sent the report in N715 to LI18NUX (and UTF-8). This new group is run by Akio Kido, big companies are involved (IBM, Sun, …). UTF-8 is more grass-roots activities – Markus Kuhn is involved in leading position. Also included is the X-Windows group, hosted by the OpenGroup. Keld recommends to cooperate with these groups.

Marc asks for the advantages of a formal liaison over casual participation in the e-mail lists and discussions. He also explains that access to restricted documents is dependent on formal participation in NB activities in Germany (and many other countries).

Arnold explains the advantages of formal liaison – access to documents on servers.

Keld: WG20 works inside the rules of JTC1.

Baldev questions the UTF-8 group's motives.

Keld thinks that this is a misnomer, the group is particularly interested in I18N issues.

Ken thinks that we should show interest and ask a lot of questions about their organization, their meetings, etc…

AI – Keld: Keld is going to investigate: who are the groups, who are the officers, when and where are the meetings, what do you want in terms of a cooperation, etc…

Ken asks in this context that statements to other groups are as accurate as possible (what is approved, what is not approves, status of projects, etc…)

9.2. SC22/WG4, COBOL

No report.

9.3. SC22/WG5, Fortran

Arnold reports about meeting Miles Ellis and the fact that the new standards will contain a lot of I18N features. Miles will step down as convenor.

9.4. SC22/WG14, C

Keld reports: New revision is now under publication as C99. Working on new I18N ideas. C works with POSIX and X/Open group's extensions to the standard. Keld's information is incomplete. Keld promises to summarize the C minutes from the last meeting, when available. C-locale (ASCII) will be registered in the cultural registry.

Ken reports that implementations of APIs in national extensions of C are tricky and are being discussed in Markus Kuhn's e-mail group.

9.5. SC22/WG15, POSIX

Keld: POSIX and C are cooperating – POSIX locale will be registered in 15897. Slow progress in 14766 guidelines for national POSIX locales and profiles. Austin group consolidates the issues (IEEE, ISO, OpenGroup). Final draft amendment #2B ends in December. This draft includes lots of I18N stuff.

Arnold: danger of overlap with POSIX I18N work.

Ken: we should send them our TR for consideration.

Marc: we can't work properly on 14652 and 15435 without knowing, what Posix is doing in 2B.

9.6. SC22/WG21, C++

Keld: No specific knowledge from their last meeting in Hawaii. Keld has suggested to make bindings to 14652 – no results as of now. Keld promises to check the minutes and resolutions and report, if overlaps or problems show.

9.7. SC2/WG2

Marc: no direct impact on WG20. Second edition is pretty much closed.

John Clews mentions the North Korean request for a "more reasonable" coding. I have the document, but not as a WG20 number. It would be a tailoring for the Korean Hanguls and ensure that combining Jamos are collated correctly.

Ken: CD of 10646-2 has 43000 characters from China – this will have an impact on the repertoire WG20 has to deal with.

9.8. GUIDE/SHARE Europe

No report

9.9. CEN TC 304

Marc: a n umber of projects in "Localization requirements". Ordering in 13. Populating the cultural registry (15897) is a job of TC304 also. Resolution withdraws EN 12005, replacing with ISO 15897. Get data from the NBs, Erkki Kolehmainen is running this project – narrative specifications, no formal specifications. Some countries (Ireland, etc..) are deciding, what their cultural conventions are. TC304e is not a sponsoring authority – TC3-4 will ask for proxy from the NBs, unclear for pan-European contributions.

Pan-European locales (editor Keld) are under development in POSIX form. Problems with YES/NO, etc…

Ken: that is what we have to talk about.

Baldev: large amount of Canadian comments to that.

Fallback standard: specifies transliteration and ordering of the MES-2. Latin, Greek, Russian fallback (to less than ASCII).

European order rules (EOR): pretty final

OCR-B for euro and Greek.

Browsing and matching very specifically for Europe. Will be distributed. Written report will be provided. Meta-data – Dublin core points to multi-lingual problems. (John reports, Dublin Ohio).

TC304 will address speech recognition as an alternative input method.

Keyboard issue: SC35 is setting up a data base on national keyboard. WG20 needs to be informed. 15897 has room for it.

9.10.   TC37

John Clews: ISO 12199 – ordering of the Latin repertoire was suspended waiting for 14651. Reissued because 14651 was considered stable. Is a tailoring of 14651.

Marc: compatible with 14651 and EOR.

Keld: Joint advisory council on language codes is revising the 2 letter codes (ISO 639). Postponed until March. Will contain Nynorsk.

3-letter codes (ISO 639-2) are handled by TC46. Some have double codes (deu and ger, e.g.), no more changes are being allowed (iw is now for Hebrew, was he).

Keld: Internet uses 2-letter code where available, 3-letters where no 2-letter code is available. Advisory committee has not met for a long time, WG20 could request action.

Java, and IBM are using the 2-letter codes.

Discussion about the value and the status of this standard and the IANA registry. Arnold asks for a formal contribution for the next meeting.

1999-11-10: discussion of a proposed resolution from John Clews:

Technical problems with the compatibility and interoperability between 2-letter and 3-letter codes. Stability is important. Keld suggests to mention, that WG20 is a user of these codes. Ask for a list of codes to be available on the web. Baldev: list of 4217 – codes of currencies – this model should be followed.

1999-11-10 afternoon: proposed liaison statement is discussed, and agreed upon.

Resolution to approve N721
Action item Keld: submit N721 to TC37

9.11. ITU-T

No report

9.12. Unicode

692	Approval of liaison between WG20 and Unicode	JTC1 N5827	1999-07-01	admin

Action Ken: ask Unicode to nominate a liaison to WG20
Ken reports: Unicode 3.0 is going to press in December for a January release. Strictly synchronized with 10646, second edition. All TRs are included on the CD-ROM are being updated. New TRs are important for WG20, they have I18N implications. Regular expressions, XML is being develop with W3C.
Feedback on WG20 work show up in the US ballot comments. Disagreement with API standard, especially with the need to have such a standard.
Action Arnold: make one document that points to the URL of all UTRs.
Discussion about the actuality and stability of TRs (or any other document).

9.13. W3C

The request for liaison with W3C was approved by SC22 in Berlin [N701] and submitted to ITTF for confirmation.
Wg20 nominates Arnold as liaison.

9.14. ISO TC46/SC2 – Conversion of written languages

The request for liaison with TC46/SC2 was approved by SC22 in Berlin [N701].
John Clew: additional transliteration is being developed. "Old" transliteration standards are being re-affirmed.

9.15. SC35 – User Interfaces (formerly JTC1 WG5)

The request for liaison with SC35 was approved by SC22 in Berlin [N701]
Alain reports that there is interest in keyboard registration.

9.16. SC32/WG3 - SQL

The request for liaison with SC32/WG3 was approved by SC22 in Berlin [N701]
No report.

10. Review of prior meetings action items

SD-5	Action item list	Winkler	1999-08-24	admin

A9905-02 – requested by Keld, no answer yet.
A9905-03 – Alain wrote request, no answer was received.
A9810-12 – Baldev takes over from Ken, Ken will provide oral input about his ideas. This has been done on Monday evening, the subject needs to be dropped from the list.
A9806-04 – closed.

11. NP for a TR about the dangers of not observing I18N

12. Amendment of TR 10176.1

698	Disposition of comments to TR 10176 registration and PDAM ballots	Winkler	1999-10-08	22.13.01
699	TR 10176 Amendment #1 for DAM ballot	Winkler	1999-10-08	22.13.01

Winkler reports about problems with using the ISO templates for Word 97 in Word 2000. The preliminary correction is a change to the registry which can be obtained by sending e-mail to ITTF (perrad@iso.ch). The templates are now also available correctly from the JTC1 web site.
Action Arnold: forward N698 and N699 to SC22 for further processing.

13. International string ordering ISO/IEC FCD 14651

683	European Ordering Rules, third version	Marc Küster	1999-02-19	22.30.02.02
684	3rd FCD 14651 – International string ordering for FCD ballot	Alain LaBonté SC22 N2933	1999-06-14	22.30.02.02
686	Common Template Table to 3rd FCD 14651 in text format for electronic use **WWW only**	Alain LaBonté	1999-06-15	22.30.02.02
687	Thai sorting example from the 3rd FCD 14651 for use with Acrobat 3.0 reader	Theppitak Karoonboonyanan	1999-06-14	22.30.02.02
688	Ordering of capital letters and small letters in English dictionaries	Michael Everson	1999-06-17	22.30.02.02
694	Sorting order of Phillipino Ng sign	SC2/WG2 N2039	1999-06-22	22.30.02.02
703	Swedish late comments to FCD 14651 ballot	Karlsson	1999-10-20	22.30.02.02
711	Summary of voting and comments to FCD 14651	SC22 N3025 (txt)	1999-10-25	22.30.02.02
718	Late Norwegian comments on 3rd FCD 14651	Simonsen, Norway	1999-11-07	22.30.02.02

Discussion:
Arnold asks for general statements from the NBs, details to be discussed later.
Keld for Norway: we should try to get to FDIS in this meeting.
Alain for Canada : Canada approved the document.
Discussion about the voting modalities on JTC1 level for the DIS. Ukraine approves late.
Strategy is to go for FDIS ballot and make sure that all supporting members vote. Also, satisfy all comments that can be satisfied without endangering another vote.
Marc for Germany: document is improved. Some comments need to be resolved. We hope to resolve everything – a vote against Germany, Japan, and the USA would not be a desirable outcome.
Takata for Japan: Japan's negative vote can be resolved in this meeting, but there are other countries negative votes.
Ken for the USA: comments can be accommodated. Most are obvious errors and smaller editorial fixes. Worried by the comments from Japan, Sweden, and from France.
Clews for the UK: concern about many FCD stages. Many editorial suggestions which can be accommodated.
How precise will the final document be? How many holes can we fix without creating new problems?

Details:
13.1.   Norway:
N#1: clause 1: support backwards can be optional for small applications. This should be a conformance option. Declaration for non-support is needed.
N#2: 6.1 note 1: equivalence on 4th level differentiates the Unicode collation algorithm from the 14651 order for a combining character sequence and a precomposed character. This is a principle discussion that will come up again in other comments. Control characters are also a problem – POSIX assigns values, UTC ignores them. It seems that only level 3 ordering is the compromise – full deterministic ordering is not achievable with ignoring control characters.
N#3: 6.2.2.3: counting is done by collating-elements.
N#4: 6.2.2: "key generation and comparison" can be on-the-fly – put it in a note at 6.2.2.
N#5: 6.3.2.WF3: accepted.
N#6: 6.3.2 WF4: delete note
N#7: 6.3.2 WF6: comment is withdrawn by Keld
N#8: 6.3.2 WF12: form of prefix – should characters that notate hexadecimal be allowed ?
The fixed length is not needed, the upper case alpha restriction is removed.

N#9: 6.3.2 WF13: accepted, the "or" will be removed.

N#10: 6.4: declare delta syntax? No delta declaration is possible for some users.

N#11: A: weight of control characters: allowed as a delta.

N#12: A: B WITH HOOK, TOPBAR, … need to be sorted with base letter. Currently these are sorted after the character. Can be done with tailoring.

N#13: A: this is a Euro centric view. The Sxxx identifiers are auto generated and should be replaced by some meaningful other – could possibly be generated without errors: Problem is the assignment of the primary weight. Compatibility with EOR ? Study for next edition.

N#14: A: Greek should not use the same accents as Latin – this is all manual markup. Marc says that this is a compromise, but the best compromise possible. This is SC2/WG2 issue. There are no combining accents, just a few compatibility characters. Equals Unicode, can be done via delta.

1999-11-09: Marc has created a new table that uses Sxxx for each new primary weight, the hex code being the xxxx. The new table needs to be checked for utility and errors.

13.2. US comments:

7.1.1. Page 1, note 2: accepted, reference sent by e-mail on 11/9/99

7.1.2. definition 4.16: accepted

7.1.3. add references to UTR #15 – accepted, sent by e-mail

7.1.4. BNF syntax – accepted

7.1.5. page 14, note – accepted, sent UTR #10 reference by e-mail

7.2.1. Fixes for Thai – accepted

7.2.2. Fixes for archaic Greek letters - accepted

7.2.3. Palochka - accepted

7.2.4. Symbol <BLANK> - accepted

7.3 Cyrillic – withdrawn, Indic needs to be fixed

7.4 Annex E: remove Annex E. accepted, correct reference in the biography to PDTR. Explanation can be added, but no reference to 14652 and no repetition of text.

7.4.1 accepted editorial comments: resultant; 2 strings; remove "-"; ISO 10646-1, Amendment 9 and addition to the references; remove extra quotation mark; section header on page 7; correct I4 and I5; long line length.

US notes it very pleased with the accommodation of their comments and will change the vote to YES, unless in the disposition of other comments unexpected "bad" things happen.

13.3. UK Comments

GB1, GB2, GB3 noted

GB4: explanations are adequate for the intended audience

GB5: ordering of space. As intended

GB6: field descriptions: noted

GB6.1 and GB6.2: rejected, describe in delta

GB6.3: explain compatibility characters – invite UK to submit text.

GB6.4: withdrawn

GB7: noted

GB8: noted

GB9: noted, will be handled later.

13.4. Japanese comments:

JP1: 3 alternatives, bringing back POSIX references or add semantics – question if it is sufficient to add notes for that purpose. Add 6.3.1.1 at the end of the BNF "keyword usage" and add the part from Annex E that contain them. Add missing words.
Accepted in part: Alternative 1 is chosen, but not CD stage.

JP2: tailored tables will often be used as input to more tailoring. WG20 agrees that the assumptions might be wrong about the distinction that the reference method can only apply to a tailored table (null tailoring?). Therefore the comment in b) is rejected.

JP3: Extensive discussion of the differences between POSIX syntax and 14651.

a) accepted: tables will be changed to assign weights to collating symbols. This eases tailoring.

b) rejected, we believe the statement is incorrect

c) rejected, use section definition and reorder-section-after

d) for POSIX compatibility accepted

JP4: table will be made to conform to POSIX practice, charmap is implicit.

JP5: This is the fundamental problem – table will be made conforming to POSIX practice

JP6: syntax for section definition is available and can be used. The syntax is not sufficiently defined for re-ordering scripts. No big tailoring is needed. Kent needs to write text to explain the use of the Uxxx in the table.

JP7: requests text changes for the introduction: add " while retaining reasonable ordering for other scripts".

JP8:

a1) take out "context-sensitive"

a2) add " the reference comparison method"

b) Since the reference method is a logical statement of string comparison, it does not preclude an implementation from using a non-UCS character encoding directly as long as it produces results as if it was using the reference comparison method.

c) rejected, the words "visual order" and "logical order" are commonly used in the respective user communities.

d) rejected, this was introduced due to NB request. The suggested text is too restrictive.

JP9: accepted

JP10: remove the paragraph. The second part has been fixed earlier.

JP11: rejected, section definition is part of tailoring possibilities

JP12: the COLL_WEIGHT _MAX in POSIX is confusing. The word "maximum" will be removed.

JP13: accepted.

JP14: rejected, this is allowed by the syntax.

JP15: accepted, correction is the reword of WF4

JP16: accepted, reword WF4

JP17: new WF4: a tailored table shall contain one "order_start" statement. This statement has to be after symbol definition and before the symbol_weight entries. (SFFFF will be at that location).

JP18: accepted – correct by changing WF12: same number of values in the range as that of the symbol range. Fix value_range also in WF13 (identical to ..) and in I12.

JP19: discussion about re-wording.

JP20: Japan wants to treat letter as punctuation: change in table is easy to do.

JP21: rejected, this would create a major restructuring of the tables to which other NBs object. No big difference, comments accepted, no change.

JP22: accepted - changing symbols table Sxxxx to S0xxxx

JP23: accepted: Ken will remove identifier from the table; change production rule for order_start .

JP24: this seems to be a misunderstanding with decomposition. The standard is about weighting. Not accepted.

JP25: not accepted,

JP26: accepted,

JP27: accepted

JP28: not accepted. Annotated text is outside the scope of the standard.

JP29: not accepted, is correct (TR 14652). Rest is accepted, but must be inserted at the correct place after SFFFF. (also fix numbers)

JP30: incorrect example – will be corrected like the Canadian

JP31: accepted, Annex E is removed

: accepted: No change for Roman numerals.

13.5.   Netherlands comments

The NL quest for a high quality standard is correct and understandable.

The group works through the editorial comments, accepting some and rejecting some.
Comma
Resultant
Insert "international"
Applied to
Strings containing characters from the repertoire of ISO/IEC 10646-1
Remove "for example"
Use "collation tables"
Remove "specific"
Remove "basic" change to "an order"
Delete "further"
... as a reference for tailoring ...
change "comparison table" to "collation table"

13.6.   Swedish comments:
SE1: Sweden has provided new text, mainly for the definitions.
Character definition will be taken from SC2.
Collation or ordering: discussion of what this standards does – Ken asserts that 14651 specifies which of 2 strings is smaller, equal or greater than the other. This means, that ordering and collation are used as synonyms. It eliminates the problems with "sorting" which involves complex algorithms.
Collation key or ordering key: accepted proposal, changes defined.
Collation subkey will be "subkey" only.
Collation key comparison: needs to stay as is, due to POSIX syntax. Question about "reference": everything is reference, why is this element specifically called so? Alain says that this is not a prescription, just a sample. Ken wants to retain the current text.
Collation weight: a numerical value, used in subkeys, reflecting the relative order of collating elements.
Collating element: same as collation element.
Collation level: the sequence number of a collation subkey.
Collation table:
Delta: take Kent's definition and add a note that in the context for this standard the delta is in reference to the CTT.
"Symbolic weight" definition goes away. Definitions 4.13 and 4.14 are removed
Stable: out of scope of this standard
5.2 and 5.3 are accepted
5.5 make more general, do not give detailed recommendations.
5.6 out of scope
5.7 not accepted
5.8 make the "normally ..." into a note.
5.9 position: rejected, is industry practice
5.10a note: not accepted
5.10b RTL script: not accepted, industry
5.11 call it "ordering key formation"
5.12 a is accepted, b and c rejected
5.13 noted

5.14 withdrawn, given the resolution of other comments

5.15 Reference comparison method.   The following describes the reference comparison method for ordering strings.

5.16 accepted in part, remove "complete"

5.17 works with subkeys – not accepted

5.18 when are two tables equivalent: change title: condition for tables to be equivalent

5.19 not acceptable, would make it imp0ossible for POSIX implementations to conform with the standard

5.20 tailoring syntax: the syntax is provided for POSIX and EOR compatibility. Not all features are in Annex A.

5.21 define "token" in 6.3.1 - add "or group of tokens" to the first 3 lines of other conventions

5.22 syntax of CTT. Not accepted, there is no technical reason to do so.

5.23 proposes new syntax. Not accepted, would make POSIX implementations nonconforming

5.24 since we don't change the syntax, this is not acceptable

5.25 withdrawn due to disposition of other comments

5.26 not accepted, but using Sxxxx. Long discussion about the advantages and/or disadvantages of script names or Sxxxx identifiers. Ken agreed to add the names of characters for primary weights.

5.27 withdrawn due to previous DoC.

5.28 text change satisfies this comment.

5.29 accepted due to other changes. Change needed in definition in 6.5 - "a delta shall…" after removing the previous sentence.

5.30 don't change unnecessarily: Add recommendation after 6.4, point 4: delta should not be bigger than necessary.

5.31 tailoring example in note is wrong: Ken will provide new table and text for the note.

5.32 not accepted, COBOL requirements contradict.

5.33 accepted for next addition, not accepted for now.

5.34 accepted in part - already discussed

5.35 Greek small sigma – should be distinguished on 3rd level. Accepted, Ken will do.

5.36 fixed, accepted

5.37 accepted, not the context. B1, B2, and B3 needs to be changed significantly – Ken !

5.38 withdrawn

5.39 B4 accepted (code points given in meeting).

5.40 accepted, correct heading numbering

5.41 moot due to German comments

5.42 accepted "as spelled out in different languages"

5.43 accepted, Word problem

5.44 deleted.

13.7.   German comments:

Problems with printing: Request paper copy if needed.

Major:
Cyrillic: B5 will be included in the CTT in GOST order. Accepted, the editor will work an appropriately modified version of B5 that satisfies Germany. Alternative is moot.

14652: remove all references to IS, remove Annex E, make TR.

Canonical equivalences should be treated equivalent – for indexing. Stop on level 3. Add note to 6.1 that normalization prior to key building solves the problem. Deterministic order was made optional in the meeting in Malvern.

Germany asks for a note that normalization is recommended for index building.

Keld would like to explain the consequences of normalizing versus not normalizing.

<u>Solution</u>: to avoid exposing encoding differences that may be invisible to the end user, it is recommended that strings be normalized according to the Unicode normalization algorithm as defined in UTR #15.

<u>Annex A:</u> not accepted to have a new schema for abbreviations for diacritics and casing.

<u>Minor:</u>
Scope: accepted
Note 1: accepted (alternative is moot)
Dash 11 : number list
Definitions: accepted.
6.2.1.2:
6.2.2: table is accepted
6.2.2.3: take out underscores – accepted
6.3.1: accepted to take out the space
6.4: replace XML with "an XML-conformant markup schema" – accepted
Annex A: use True Type font, mono-spaced,

<u>Action Arnold:</u> find out, if the document can be parts, one of which is a landscape table. Also ask about URL for table – where?

Annex B accepted.

13.8.   French comments:

1.1 help with translation is desired.
1.2
1.2.1 separate requirements from other stuff

subclause 6.2

collating element: BNF was fixed, rest is not accepted
6.2.1.1 tailoring phase: accepted, sentence reworded
6.2.1.2 position parameter. Text added to explain the position parameter.
Order-start occurrence: accepted, the current draft allows only one such statement.
6.2.2 accepted, text changed
"undefined" – the first "undefined" will be underlined – accepted
6.2.3 – accepted, has been reworded. Ken re-formulated over night and provided mathematical definition and text for Alain (not repeated in these minutes).
1999-11-11: discussion of the proposed new text.

Subclause 6.3

BNF accepted
Line completion accepted for trailing spaces – remove brackets.
Level_token not accepted, special keyword IGNORE is not a weight. May also be multiple weights and this introduce other confusion.
WF1: it is a complex matter: can also apply to non tailored tables. The WF1 is technically correct.
WF2: duplicates are not allowed in a symbol definition – solution: no symbol that occurs in a symbol_definition in a weight table that contains no tailoring lines may occur in another symbol definition in the same weight table. Accepted, WF2 is re-written.
U00C0 … are allowed, comment not accepted
WF6, WF12, WF13 need not to be moved or renamed, any order is OK.

WF rules : insert "...shall be followed at a later point in that tailored table ..."
WF10: accepted, needs to be re-written to contain one order_start and one order_end.
WF12: accepted, re-written
Note on value range: not accepted due to other changes to the text in WF13 and removal of the term value range.
Add to note: not accepted, is needed for existing implementations.
U-symbols: not accepted, might break implementations.
6.3.3 implication is clear enough, changes might cause other difficulties in the rules
I2 and I3 : accepted: Expand WF: the hex numeric string concatenated to the common prefix must contain the same number of digits as the hex numeric string of the first symbol.

An ordering key x is less than an ordering key y at level s where $1=s=m$
iff (there exists some integer i where $1=i=s$ such that $_xsk_i<_ysk_i$)
and ($_xsk_j=_ysk_j$ for all integers j where $1=j<i$)

A subkey v of length l_v is less than a subkey w of the length l_w
iff ($l_v=0$ and $l_w>0$)
or (there exists some integer i where $i=l_v$ and $i=l_w$ such that
(for all integers j where $1=j<i$ $_vwt_j=_wwt_j$)
and (either $l_v=i$ and $l_w>i$ and $_vwt_i=_wwt_i$)
or ($_vwt_i<_wwt_i$))

1999-11-11: continuation of discussion of French comments:

Detailed changes were made to the document to accommodate the French comments. Some changes were made to the BNF. Details are in the FDIS document.

Precomposed characters: Danish example will be expanded with a collating symbol for the combining character A with ring. Ken will draft the addition, if Keld does not stop it, it will go forward.
Gumurkhi and Devanagari – Baldev has asked expert in India, no answer yet. Baldev confirms the French comment and can provide details not before 2 weeks from now. Ken will make all known changes, rest will have to be done with tailoring.
Action Baldev: information about Gumurkhi and Devanagari ordering to Ken by November 25, 1999. Marc will contact Indian specialists and inform Ken and Baldev.
Ken will fix the text and add the combining characters in the example. Ken will also fix the tailoring example (add order_start for example).
B.4 will be moved to C
B.5 has been integrated in the CTT.
Annex E is removed – accepted.
Annex F – accepted.

Progression of FDIS 14651 :
WG20 decides to go for FDIS (Sweden is against it, all others support it)

December 3, 1999	WG members input to Alain	Ken, Keld
December 10, 1999	Alain completes draft DoC and draft FDIS	Alain
January 7, 2000	Final day of comments by editing group	Editing group
January 14, 2000	Final document to convenor for FDIS ballot	Alain

Generous offer from Kent Karlsson to check for "Word" problems is happily accepted.
Keld will set up an editing group mailing list sc22wg20ed@dkuug.dk for the comments purpose.

14. Cultural convention-specification standard ISO/IEC PDTR 14652

690	Text for ballot: PDTR 14652 - Specification method for cultural conventions	Keld Simonsen	1999-06-30	22.30.02.03
705	Summary of voting and comments to PDTR 14652	SC22 N3024	1999-10-13	22.30.02.03

In an e-mail from Keld Drube, Denmark, he says on November 3:
Regarding the TR 14652 which is going to be discussed during the meeting, then I will withdraw the Danish claim for the report to be changed into a TR 2 rather than a TR 1. Both report types have the same chance of becoming an IS at the 3Y revision, so we don't need to discuss this further.

1999-11-11:
Keld will walk us through the comments in N705:

14.1.   Canada comments:

1: Call it Technical Report everywhere – accepted.
2: consistently change terminology to TR: (shall, …) ~~not~~ accepted. Type 1 can have such language and might become a IS. Baldev does not accept this logic, he wants non-prescriptive language. Arnold points out that the non-acceptance might be a reason for NO votes. Keld says that it has to be re-edited, when it becomes a standard again. Suggestion to put some language in front of the TR that explains the reason for prescriptive language in a TR.

There will be an introduction and the effort to do all edits!

1: remove text
2: accepted
3: provided by the application or the operating system
4: group definitions – noted, better to sort them alphabetically
5-7: accepted
8: accepted where applicable.
9-10: accepted
11: accepted – empty string
12: accepted
13: Basic keywords – accepted: transliteration keywords
14: accepted
15: map: not accepted, is in POSIX and C. Baldev contradicts, Isak Korn reminds Keld, that for the world outside this meeting "map" is loaded. GNU and LINUX need it. Not accepted.
16: automatically included: not accepted, is in POSIX.
17: LC_CTYPE for transliteration: Baldev: this will break existing implementations. Use LC_XLITERATE and try to convince the POSIX folks to change to LC_XLITERATE. Also, define it outside makes it easier to implement and especially for CTS (conformance testing service). Accepted, pending check with C standard.
18: accept
19: have been reviewed 2x and should be correct
20: punctuation is "all the other crap" euro and dollar etc. Accepted: All goes into graph except what was in portable set of XPG 4. Use Unicode character properties file to assign characters to punctuation.
21:
22: accepted, yes
23: agreed to new wording
24:
25:
26: currency dates are not needed. Keld: this is an optional keyword. Distinction needs to be made between optional and mandatory keywords.

27: same as 26 – optional. Ken: should be removed. Discussion about the accessibility of currency exchange rate for uses – not without system knowledge. Should be GUI based and user accessible. (Careful: this is only for frozen rates, can not be used for variable currency rates). Keld insists in maintaining the current proposal.

28:

29-31: accepted.

32: not accepted, there is an ISO directive.

33: not accepted, see 33.

34: accepted, will be clarified.

35: accepted, new text from Baldev.

36: further options are needed. Rule needs to be defined how to allow more time zones.

37: accepted, restrictions are stricken.

38: Baldev will provide text. Accepted.

39: Add well known sizes, e.g. A4, letter, etc... Long discussion about the value of a default paper size. Most applications can set their own paper size, many depend on the Windows APIs and the printer drivers.

Remove LC_PAPER. Keep it in the cultural registry.

40: improvement to the comments – need indication, if 2 or 3 letter codes. Why in LC_ADDRESS ? Locale defines the language and country. Define a LC_MISC for all the miscellaneous codes?

Discussion about the usability of the language code in a locale – this is better done by APIs and the cultural registry. Delete everything for a country.   Accepted.

14.2.   Danish comments:

Denmark wanted to make it a Type 2 and voted NO therefore. They have changed their opinion since then and agree with the Type 1.

Change IS, standard, etc.. to TR terminology – accepted.

Sort definitions – not accepted.

Maintain reference of definitions from other standards –accepted where differing definitions exist in various, normatively referenced standards.. Ken points out that the POSIX definitions might differ from definitions in 14651 – it might make sense to either align this TR with 14651 (collation and collating element), or specifically mark where references come from.

FDCC definition – accepted.

Where do the table come from – accepted, POSIX standard.

Plans for future: accepted, in note.

FDCC set description: accepted.

Shall, define, ..: Type 1 can contain these words. Accepted.

References: accepted, efforts to be consistent.

… shall be escaped: accepted, addition escape character

… wth: accepted, will be corrected.

14.3.   German comments:

TR instead of a standard – accepted

LC_CTYPE – address with the US comments.

Re -visit DoC ? not accepted, comments have been disposed of.

14.4.   Japanese comments:

Why TR – accepted.

Proposed text for TR introduction and a few paragraphs .

Proposal to have an **annex** that shows recurring problems (issues list). These issues that have been rejected a number of times (principal items) ought to be in the annex, as well as a list of gory details.

Comments and suggestions are accepted.

14.5. US comments

Ken thanks the editor to for adding the line numbers.

1-2 dealt with

126: 10646-1 is constantly changing, due to amendments. Suggestion to reference a given Unicode standard or as 14651. – rejected.

169: rejected, timing problem.

270: syntax and meta-syntax. – accepted, the editor promises to make it consistent.

304: portable character set: rejected. Ken: there is no need for character specification in a TR on cultural specifications. Keld: needed for Linux, for POSIX, backward compatibility.

491: not accepted. Comments are not allowed everywhere, due to Japanese input.

585: in use in the industry – rejected

677: praise accepted, noted.

717: accepted

738: empty string, accepted

740: accepted.

795: Ken suggests a new way of achieving the same detail – rejected.

874: accepted

926: "map" out of FDCC type, as per Canadian comments. Accepted in part.

976: accepted in part

1091: make description correctly – East Asian symbols, ideographs, Hangul PU area,

1315: accepted

1369: as per Canada. TR does not need to specify charmaps and repertoiremap. Rejected.

1848: text removed. Accepted.

1888 and 2081: accepted. Delete text.

(2085: correct to: … exceeds the number of levels, the keyword supports)

1964: accepted

page 31 ff: rejected, needed for POSIX.

2014: noted

2121: accepted: for canonical decomposition …

2203: accepted

2226: rejected, needed for POSIX

2293: accepted

page 38 ff: partly accepted, equivalences will remain, correct BLANK to BASE

3233: 2 options for the YES expression (+ or 1) and 2 options for the NO expression (- or 0) – needs clear explanation.

3237: text removed

3262: rejected

3298: will be changed

3200: professional title used in name formulation: string. Remove example.

3274: example

3305: remove text

3318: this is for localedef utility (also in other locations). Add text at the beginning that explains then need for this format for the localedef.

3237- page 57: moot

page 57: charmap – used in implementations. Rejected.

3502: rejected

3593: octal and decimal representation should be deprecated.

3625: rejected

3622: accepted

page 62: repertoiremap – not accepted, is a basic principle

3823 –3848: accepted in principle.

Page 92 – Rationale: rejected, too much work

6138: 16-bit bytes – rejected, is used in C

B.1.3: too much irrelevant information. Explain LC_COLLATE . No tutorial. Differences from POSIX.

6315: accepted, will re-write

B.1.3.3: SPECIAL will be fixed

6753: noted

6762: spell out API work

page 104: character set - rejected

page 105: repertoiremap – rejected

section 4.5: monetary – rejected

2616 and 2621: euro implementation specific.

Editorial: all accepted.

1999-11-12:

Ken found additional issues with 14652.

To align 14652 with the new POSIX draft, changes need to go into 14652 and BNF:

Example is char_class instead of class.

L_NUM should be added to 14652.

LC_CTIME has added error parameters …

Format statements for time

New sentence for portable character set in locales – characters are allowed. Keld objects.

BNF problems in 14652:

Comments inconsistencies: 7125 - at the same time change grammar of LC_COLLATE

7078 – paper – remove

7102: address keyword string must be fixed

6999: re-order section end – remove

6985: collating element is undefined

6986: collation element is unused

check ellipsis

6970: spece+ ???

6968: collating element is missing : "from" and "space" – also does not allow symbol weights before reorder start - must be allowed

6977: order-start: correct to match 14651

6980-6982: order options – these allow illegal constructs

Generally: incompatibilities between BNFs – UCS symbols, ranges, weight symbol "empty" differs from 14651.

Collating element statement – one or more as in 14651.

Alignments with 14651 BNF from the new FDIS draft are needed.

Progression:

Marc recommends looking at the new document in Quebec before we send it for ballot, Ken supports this, so does Takata. Canada voted yes.

No request for input of other NBs for the issue list.

January 14, 2000	Provide National issues list to the editor	All NBs
January 14, 2000	Draft disposition of comments	Simonsen
January 21, 2000	Comments to draft DoC to Keld	Editing committee
January 28, 2000	New draft for next ballot and final disposition of comments.	Simonsen
March 3, 2000	Comments by the participants of the WG20 meeting in Copenhagen	Editing committee

15. Registration of cultural elements ISO/IEC 15897

577	Table of replies and comments to Fast Track ballot on ISO/IEC DIS 15897 (EN 12005)	SC22 N2717 (paper only)	98-05-14	22.30.02.03
592	Disposition of comments to fast track ballot for ISO/IEC DIS 15897 - Procedure for registration of cultural elements	SC22 N2812	98-09-02	22.15897
712	New Working Draft for IS 15897	Keld Simonsen	1999-10-29	22.15897
713	Report of the convener, resolution of comments meeting on DIS 15897 – Registration …	Bob Follett SC22 N2811	1998-09-02	22.15897

Arnold explains the history of this standard – WG20 NB rejected, CEN done as EN 12005, fast track to IS 15897 with unresolved issues.

Ken explains that the US objects to normative references from this standard to non-normative Technical Reports. This would in fact make the TR 14652 normative.

Alain wants that narrative descriptions of cultural elements can also be provided in French and in Russian.

Keld: this is the collection of an issues list. The group agrees that such an issues list is a good idea and the editor asks for additional issues to be added to the list.

List of issues:

Ken: registration schema is very similar as the registration for character sets. Problems are that no quality control is possible – the NBs elements have to be registered, however lousy they may be.

Marc: additional problem is that very few countries submit data to be registered.

Keld: uses POSIX syntax

Ken: the fast track process made EU policy a world-wide problem, the problem being "member body control".

Marc: quite a few elements are not addressed by national bodies, or that they are not qualified to handle them. Some of the required cultural conventions are not in the jurisdiction of standards bodies – inflection rules (linguists), telephone numbers (there are no rules), etc…

Ken: there is no need to fix errors, even if they are detected. The collection of unchecked data is without value.

Alain: request for correction is needed.

Ken: Collection rules : NBs will not be able to input collation rules for their ordering rules according to 14651 and 14652.

Marc points out that Island has not been able to provide POXIF formatted specification. All that can be expected are narrative specifications. In Europe, the EU is providing money to "populate the registry" with independently checked data.

Baldev: From Canada, the registry only contains "locales" with its 6 categories. Not likely to ever get anything more.

Keld: registry is open to companies for input.

Baldev: version control.

Ken: in Annex C a totally incomprehensible normative reference to ANS.1

Discussing N592, the disposition of comment of the fast track ballot

Canada, Denmark, Ireland have agreed

Finland is mentioning that the standard needs to be revised.

Japan: asking for comments by e-mail.

US:

US-1 was in fact rejected. Scott was the representative.

US-3

US-4 re-order after is now specified in 14651, should not be specified again.

Action Keld: send out 15897 issues list by December 15, 1999

Action Takata: send 15897 issues list by December 1, 1999

Comments on issues list from all members is requested.

16. Internationalization API standard ISO/IEC 15435

714	New committee draft for ISO CD 15435	Keld Simonsen	1999-11-07	22.15435
722	Issues list for			

Ken's comments on N722:

11 word break – use UTR #14 for line breaking information, not word breaking. Look at break-iterator class of Java for input. Hyphenation and spelling is another issue. All these services need dictionaries etc… Recommendation, not to do that.

1: bindings – do not do !~

4: strings – decide for one or the other

6: thread-safe – this is a hope, very problematic. Call-back is dangerous.

8: struct ? Impossible in C. Example was given from Keld by Ken. Bad examples are 6.2 and 6.3, these are not re-entrant, also what is a string.

5.1.2 – binding wrong.

Inconsistency terminology, suggest "result" consistently.

17. ISO/IEC 10646 Issues

No time for action.

18. Other business

18.1. Japan's NB contribution to SC22 plenary in Berlin

696	Japan Member Body Request for Discussion of CD/FCD Processing	NB of Japan SC22 N2993	1999-09-09	admin
697	WG20 convenor's remarks to the concerns if the NB of Japan in SC22 N2993	Winkler	1999-09-21	admin

Arnold reported that the Japanese contribution was discussed in Berlin's SC22 meeting. Results are reflected in the resolutions from Berlin in N701 and N702. Important points:

- No multiple FCD ballots

- In the disposition of comments we need to provide rationales for acceptance and rejections

- The next draft should, if possible, be submitted with and without change marks.

18.2. Ireland's NB report to SC22

704	Annual report of the NB of Ireland for 1999	SC22 N2986	1999-09-02	admin

Arnold explained the background and postulated that at this time no action from WG20 will be taken to move project 22.30.02.02 (ISO 14651) to SC2.

18.3. CLAUI meeting

There will be a CLAUI meeting in Spring 2000, perhaps in California. Unisys has suggested to hold it in Malvern.

19. Review of Priorities and Target Dates

Done in the respective projects.

20. Review of Actions Items from this meeting

SD-5	Action item list	Winkler	1999-11-12	admin

Reviewed.

21. Approval of Resolutions

709	Resolutions from the WG20 meeting in Copenhagen, November 1999	Winkler	1999-11-12	admin

The resolutions were approved unanimously.

22. Adjournment