Minutes - Québec
ISO/IEC JTC 1 SC22/WG20
Meeting #18 - Internationalization

May 26, 2000

DATE: May 22-24, 2000

TIME: 9:00 a.m. to 5:00 p.m

LOCATION: Secrétariat du Conseil du trésor du Québec
Room 3.20 [A et B]
875, Grande Allée Est
Québec, Canada

1. Introduction and announcements by Convenor

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

Participant	Affiliation	Country
Alain LaBonté	Secrétariat du Conseil du trésor du Québec	Canada
Baldev Soor	IBM	Canada
Keld Simonsen	RAP	Norway
Marc Küster	University Tübingen	Germany
Masayuki Takata	Edogawa University	Japan
John Clews	Sesame	UK
Ken Whistler	Sybase	USA
Arnold F. Winkler	Unisys	convenor

3. Appointment of chairperson, secretary, and drafting committee

Winkler as chair and secretary, Simonsen and Küster as drafting committee.

4. Approval of prior meeting's minutes

710	Minutes of the WG20 meeting in Copenhagen, November 8-12, 1999	Winkler	1999-11-17	admin

Keld: was happy with the minutes. Meeting number for Copenhagen is incorrect, needs to be 17.

Minutes are approved with changes as above.

5. Future Meeting Schedule and Plans

October 30 – November 1, 2000	Malvern, PA	Unisys
May 8 – May 10, 2001	Tübingen, Germany	University
October or November 2001	Brussels	Together with CEN/TC304
May 20-22, 2002	Tromsø, Norway	RAP and NTS

It was generally agreed to shorten the WG20 meetings to 3 days.

Possibility to have overlapping meetings with TC 304 in Europe. TC 304 as well as WG20 seem to need 3 days rather than a full week. This would allow to join one day within a week.

Also idea to meet with SC35.

Too many difficulties in scheduling:

Keep Malvern but shorten it to October 30 – November 1, 2000 (shorten to 3 days).

Keep Tübingen from May 8-10, 2001 (shorten to 3 days}.

Add Brussels for Oct. or Nov. 2001, depending on agreement with CEN/TC 304. **AI Marc** for coordination with CEN/TC 304, for rooms and other arrangements.

Add Norway for May 20-22, 2002 pending confirmation. **AI Keld** for confirming meeting in Norway.

6. Recognition of new documents and assignment to agenda items

	Document	Source	Date	Agenda
746	Québec, participants, May 22-26, 2000	Winkler	2000-05-22	#2
747	Québec, final agenda, May 22, 2000	Winkler	2000-05-22	#7
748	Québec, resolutions, May 26, 2000	Winkler		#18
749	Québec, minutes, May 26, 2000	Winkler		Admin
750	Document register N701-N750	Winkler		Admin
751	Convenor's report 2000 to SC22	Winkler	2000-05-24	#17.1
752	Issues list #4 for ISO 15435	Keld Simonsen	2000-05-15	# 15
753	DTR 14652 – specification methods	Keld Simonsen	2000-05-21	#13
754	Draft ISO 15435 – Internationalization APIs	Keld Simonsen	May 2000	#15
755	Draft ISO 15897 – Register for cultural elements	Keld Simonsen	May 2000	#14
756	Statement of mutual benefits for a liaison between SC22/WG20 and LI18NUX	Keld Simonsen	2000-05-22	#8
757	Resolutions from the SC2/WG2 meeting in Beijing	SC2/WG2 N2204R SC2 N3422	2000-03-30	admin

7. Approval of Agenda

747	Québec, final agenda, May 22, 2000	Winkler	2000-05-22	admin

Agenda is approved with addition of documents N751 through N755.

8. Liaison Reports

8.1 Additions/deletions/changes to liaisons

AI Keld to write a statement of mutual benefits (N756) and contact Kido on LI18NUX to agree on the text. Preferably during the Québec meeting.

8.2 SC22/WG4, COBOL

Keld: Request from Ann Wallace that COBOL would like to know about case mapping and case folding. Ken and Keld responded – no more communication with Ann since then. COBOL has meeting this week.

8.3 SC22/WG5, Fortran

No report.

8.4 SC22/WG14, C

Keld: C standard has been approved. C-Locale will be registered according to 15897.

Ken: has there been any feedback from C regarding APIs? Keld and Arnold will look for documents.

8.5 SC22/WG15, POSIX

Austin group: OpenGroup, IEEE, JTC1 cooperate. ISO/IEC 9945-2b standard is in WD state.

Important work are LC-Collate and regular expression.

Question about case sensitivity in regular expressions, e.g. delete functions from A through B includes upper and lower case which for some is undesirable (comments to draft). Agreement that the comparison is the sequence of the lines in the locale – Baldev does not like this idea at all – needs source code everywhere…

AI Keld: make the resulting paper from the teleconference or meeting last week on sorting and regular expressions available to WG20.

Other comments: is the SHELL locale dependent ?

Next POSIX meeting is in Nashua, NH in July.

8.6 SC22/WG21, C++

Nothing to report.

8.7 SC2/WG2

Ken Whistler reports about the meeting in Beijing: Ballot comments for CD 10646-2 were resolved. Few changes, nothing removed, add about 500 compatibility Han characters. FCD is issued before the meeting in Greece. Little change is expected due to comments on the FCD. Could go for DIS in autumn – major extension.

Gothic (ancient script) is included, as are Old Italic (formerly Etruscan) which covers more than just Etruscan, Deseret, Mathematical symbols, Plane 14 identifiers, and Musical symbols. The same repertoire will be added to Unicode as Unicode 3.1 as soon as it is approved.

Maintenance for 10646-1: large set of 950 mathematical symbols (from the US STIX project) was accepted. Will become an amendment after Greece, including 16 Cyrillic Sami characters. Some symbols in JIS X0213 are also included. More Yi radicals, Chinese experts were present.

What to do with Han characters from JIS 0213, and some more from HK? Canadian proposal to separate about 1000 characters and encode them in the BMP was defeated.

Korea, incl. North Korea was presented in force, pushing their proposal to re-encode Korean (Jamos, names, sorting). No support from anybody. The NP proposal was also voted down.

Peso sign and German Penny were added, miscellaneous other symbols fro Japan. Uighur activities for additional presentation forms were stopped as unnecessary.

8.8 GUIDE/SHARE Europe

Nothing to report.

8.9 JTC1 WG5 == SC35

Alain reports: 5 resolutions interest WG20:

CLAUI meeting October 19-20, 2000

ISO 7000 ICONs and Keyboard symbols need to be aligned (e.g. symbol for HELP).

National keyboard information data base was approved, Alain actioned to develop template.

Looking for volunteers to write TR on CLAUI.

Who is the chair of TC45 ?

8.10 CEN/TC 304

Clews reports that BSI has opted out of TC 304, also out of SC2.

Market study is in preparation for the "success" of CEN/TC 304. Call for consultant to do that study. Great variety of projects makes this a complex task. Result will be important for the future of TC304.

Projects in discussion:

Companies will be allowed to participate directly. Preferred is NB participation.

IOR have passed as ENV 13710, no printed version yet.

IOR-2 extension for more exotic scripts is on its way, draft being written as a revision of ENV 13710.

Fallback standard – comments received from various countries, downgrade to CEN report, no ENV.

Guide to character sets with comments from USA is in publication, will be a CEN report later. Considered to be a "living document", the CEN report being a snapshot thereof..

Ken asks TC304 to invite comments, if they are desired. Marc responds that the opening of the process will support this request. Unicode will be able to comment directly.

Matching and browsing report, also with USA comments, is presently stopped until the market study is complete.

ENV 12005 – identical to 14651.

8-bit standard for European character set has few implementations, UK had recommended to withdraw. Number is 1923.

Alpha project: List of all languages in Europe, was a workshop, not in TC304. Official languages, officially recognized languages, and dialects. Will not become a CEN report, but is on Everson's website.

Euro-locale: used by EC and other organizations. POSIX format. EOR collation sequence. Results of the ballot not yet known, expected to be approved. Will be progressed into a ENV.

Population of the cultural registry: project team will continue work, even without payment. Narrative descriptions.

Taxonomy: need to be coordinated with WG20

MES: will be a collection in 10646, passed ballots. WG2 looked at the MES: N2211 asks for collection identifiers – approved, next edition will contain. Other document to split presentation forms was rejected.

OCR-B testing process for new characters (accented Latin letter). Companies are now involved, mainly banking.

Hjulstad has announced that 12199 ballot (as a profile of 14651) has closed and is approved.

8.11 TC37

736	Reply to WG20 liaison statement to TC37 on ISO 639, parts 1 and 2	ISO 639 JAC Rebecca Guenther	2000-02-22	22.30.02.03

Keld reports: sorting standard 12199 has completed FDIS ballot with most likely positive result.. Kelds action resulted in N736, the reply from TC37. WG20 wants stability – no code changes, no reassignments, list of codes on the web is provided. Keld is appointed as observer or "another person" to the advisory committee on language codes.

ISO 639 – 2 letter codes

ISO 639-2 – 3 letter codes (to be expanded, used in libraries). Other committee has other 3 letter codes – non synchronized.

RFC 1766.

Clews recommends a draft resolution to TC37/SC2. The opinion of the group is that the use of not standardized country and language codes is not a WG20 problem. No resolution, just liaison about our concerns. **AI Clews:** check, if TC37/SC2/WG3 wants a liaison with WG20.

8.12 ITU-T

No report.
AI Winkler: ask Stefan Fuchs, if ITU-T is still active in the area of our interest.

8.13 Unicode

Ken reports: met 2 times since last WG20 meeting.

Editorial fix for a glaring glyph error in Ethiopic.

UTC decided **how to refer to characters off the BMP**. UTC will use the U+ notation with 4, 5, or 6 digits. First character on page 1 will be U+10000 (BMP is U+xxxx). Only for user defined character on plane 16 will U+ and 6 characters be needed.

Bi-directional mirroring: Arabic mathematicians sought for opinion on mirroring.

Georgian proposal from Everson – rejected. **AI Whistler:** send response to Georgian proposal to Bunz and Clews.

Clews is interested in Thaana.

Unicode 3.01 will contain corrections, no additions. Script.txt will be informative and part of Unicode 3.01.

Technical Reports: **UAX** (standard annex) are part of the standard (e.g. normalization). **UTS** – technical standards with conformance clauses: compression (#6) and collation (#10). UTRs for the rest (e.g. case mapping). XML based character mapping table, includes fallback mappings.

8.14 W3C

732	Category C liaison between W3C and SC22/WG20	ITTF	1999-11-29	admin
735	ITTF approves liaison between W3C and WG20	ITTF, SC22 N3073	2000-02-14	admin

Arnold reports giving the W3C 11017, 10176, the sort table. The value of control and input is great.

Keld is concerned about the unusual relationship with W3C – no access to document, no participation to meetings, … Liaison to provide documents, when they become "public". **AI Arnold:** check status

8.15 TC46

Clews reports about transliteration standards for various scripts. Some for Indic scripts, also for Cyrillic scripts. Voting procedures made complicated situations. Script code standard, connected to language codes and country codes, will have 4 letter codes. Everson might become maintenance/registration agency for DIS 15924. No code re-allocation should be necessary.

Problem: TC46 is losing the secretariate (DIN). DIN is also relinguishing ISO 3166 registration of country codes.

Ken has concern that 3166 defines scripts differently from 10646. This makes the script definition a SC2/WG2 concern – TC46 and SC2 should have the same definition or at least a description of the difference in the script definitions. WG20 will not take action, but TC46 representatives need to know the danger.

Keld: TC46 has a number of SCs, they have secretariats.

9. Review of prior meetings action items

SD-5	Action item list	Winkler	2000-05-22	admin

Reviewed, marked. All done except an old one on APIs.

AI Keld: put link to L18NUX on the SC22/WG20 web site.

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

No document. Alain suggests working with SC35 to produce such a document.

Take to CLAUI.

Winkler will take the subject off the agenda until a draft is available.

Keld and Mark indicate that they might be able to provide input.

11. Amendment of TR 10176.1

737	Results from the JTC1 DAM ballot on TR 10176	JTC1 N5991	2000-02-25	22.13

US comments accepted as editorial. Amendment is in publication process.

12. International string ordering -- ISO/IEC FCD 14651

727	Devanagri collation	Baldev Soor	1999-11-24	22.30.02.02
728	Gurmukhi collation	Baldev Soor	1999-11-24	22.30.02.02
730	Disposition of ballot comments for FCD-3 14651	Alain LaBonté SC22 N3070	2000-02-08	22.30.02.02
731	FDIS 14651 – International string ordering	Alain LaBonté SC22 N3093	2000-04-21	22.30.02.02

WG20 recommends to remove the 3 bullets in the explanation and to re-post the cover memo.

Keld requests links to SC22 documents in the cover memo – Arnold will provide documents and new cover memo to satisfy this request.

Keld brings up the issue of cyrillic sorting – current tables don't make deterministic order. This was a decision made in Malvern by the working group.

Marc brings up the fact that currency signs are sorted as characters – contradics the POSIX implementations and the European Ordering Rules. This treatment has been in the tables since FCD-2. This can be changed by using tailoring.

WG20 decided to establish a issue list for the next edition of 14651. Alain is the "keeper of the list". Currently on the list are:

- Deterministic ordering for cyrillic (Keld)
- Handling of currency signs in level 1 (Germany, Norway, Canada)
- Extend of tailoring for average country. (USA)

12.1 Request to ITTF to produce IS 14651 as a CD

Resolution: Publish IS 14651 as CD-ROM with separate text files for the common template table and the tailoring samples.

12.2 Request for a new work item to amend IS 14651:2000

1. Bring 14651 in sync with 2nd edition of 10646 (add 10000 character) – 10646-1:2000.

2. Repertoire of 10646-2: 2001, about >40000 characters, Han, mathematics, musical, alphabetics.

3. Add 10646-1:2000 Amd.#1 (and possibly more) for about 1000 symbols.

WG20 needs to address all of these in order:

For 1) Ken can produce a draft table for the next meeting in Malvern.

There is very little argument about the sorting order, tailoring will be needed. Section 1) is the most complicated, some of the scripts are complex and need experts. For 2) and 3) the major work is its magnitude, not the complexity. 1) and 2) should be included in amendment #1, 3) in a further amendment.

Resolution: WG20 requests subdivision of 22.30.02.02 for an amendment of 14651 to include the full repertoire of 10646-1:2000 and 10646-2. (JTC1 directive 14.5)

Resolution: WG20 requests permission of concurrent registration and PDAM ballot.

AI Winkler: put the request for project subdivision and concurrent ballots for 14651, Amd.#1 into the convenor's report to SC22 plenary.

Han sorting data are available from Unicode in Unihan.txt (16 MB). Extension will make file much bigger. Major work is in the compatibility mappings of new characters.

AI Whistler: produce a working draft (table) for the amendment.

13. Cultural convention-specifications -- ISO/IEC PDTR 14652

706 R	Disposition of comments to PDTR ballot 14652	Keld Simonsen	2000-05-18	22.30.10.03
752	Issues list #4 for ISO 15435	Keld Simonsen	2000-05-15	22.15435
753	DTR 14652 – specification methods	Keld Simonsen	2000-05-21	22.30.02.03

Whistler, Küster, and Takata mention that they had not enough time to study the disposition of comments – it will be necessary to make these comments against the new draft (N753).

Keld walks WG20 through the changes.

Language codes in address field need to be explained.

Detailed discussion of alignment with POSIX – Keld has made it so that it matches the future POSIX functionality and syntax – this might change due to comments against the POSIX standard. POSIX FCD has been approved, so Keld does not anticipate any changes to the current draft.

AI Winkler – make the ISO 9945-2 amendment a WG20 document.

Marc suggests using Unicode data instead of the tables in 15652. Keld refuses, insists on using the POSIX data.

Ken would like to see the disposition of comments (N706R) reflected in the new document – it is unclear and can not be traced. Especially the last page bunches a set of "actions" together, which are not reflected in an obvious way. Ken suggests to remove the last half page from the DoC and make them into a new document with explicit explanation of related changes in the TR.

Arnold: makes a plea for all participants to allow the implementation of the last 1/3 page without the explicit instruction through ballot comments.

Progression plan:

May 24, 2000	Keld Simonsen	Annex to the DoC (N761)

June 22, 2000	Editing group	Comments against 14652 WD4 to editor		
July 6, 2000	Keld Simonsen	Editor produces new document based on the comments of the editing group		
July 13, 2000	Editing group	Check by editing group before submission to SC22 for ballot		

14. Registration of cultural elements -- ISO/IEC 15897

723	Issues list #1 for ISO 15897	Keld Simonsen	2000-05-22	22.15897

No new document.

Issue list discussion:

1. TR normative referencing: can be allowed by SC22 when requested and approved.

Ken suggests that TR 14652 be referenced as "recommended" in 15897, without making it normative. Keld suggests a compromise to use "unspecified" for categories FDDC set and charmap, self-identifying (e.g.XML, 14652). Issue resolved.

2. French as primary language, also Russian. No objection, closed. Implications on the text in 6.1. "in one of the official languages of ISO/IEC , presently ...."

3. Quality control: syntax must be followed. Recommended: look for SC2 process of character set registration.

AI Winkler: distribute the SC2 process (ISO 2375) to WG20 and check, if we want to adopt a similar process. Keep issue open

4. National bodies might not be experts in some areas of registration. Question is, who is allowed to register for a specific country.

Keld mentions that NBs are not very eager to submit registrations – in the current register are about 150 registrations, mainly charmaps, mainly Danish locales.

Marc reports that in Germany it was very hard to get reliable data.

UK point of view: over-regulation ! Keld explains that the information is free – Baldev expresses that free does not mean "usable". Ken: single point of data control is necessary, making clean, usable data – then people are prepared to pay for such data.

Registration authority must "care" and be knowledgeable, only good data have value. Example Graham Rhinde for postal formats. Keld says that the registration can not be changed – that leaves the issue unresolved and irresolvable.

5. Notes on corrections – what changes and why need to be recorded. Understood, closed.

6. Version control: how to find the correct data. Is the highest number always the latest registration? Some of the wording needs clarification. Identity crisis ! RAG needs to clearly and unambiguously identify new registrations. Versioning needs to be clearly defined – develop guidelines. Clarification of "identical" is also needed. Issue is understood and closed – regulations must be forthcoming.

7. Can organizations register for somebody else ? Presently it needs clarification. JTC1 is opening up to consortia – we need to do the same. The feeling of the group is that companies could be allowed to register, if strict quality control is applied.

Perhaps create a two part registry – one for official submissions, one as company repository (Baldev). Marc is against this "dump-in" approach. He thinks that the disadvantages outweigh the advantages by far – it would open the registry to individuals with specific ideas about some conventions.

Agreement:: delete item d in chapter 5 and not allow second class registrations.

Additional problem: too restrictive in who is allowed to submit – CEN could not submit anything for Austria, but could do so for all of Europe.

If specific rules are to be applied, the rules need to be documented in the standard (token identifier).

JTC1 working groups should be allowed to register for specific countries, not only for "large territories" or groups of countries.

Delete d) and change c) to take the restriction off. Make a note about priorities.

Issue closed.

8. Annex C: allows ASN.1 for specification. Noted and closed.

9. Contributions are welcome. Sweden asks for separation of normative and informative parts into different annexes.

10. Reorder after: applies not to 14651, but to POSIX locales. Needs to be maintained. Example should be separated. Leave open for further consideration.

New issues:

11. Create one normative and some informative annexes.

12. Removal of registrations ???

Ken has a question on example registration – more annotation is needed (copy en-dk ?).

Progression:

New draft for the next meeting. New issues list will be produced.

15. Internationalization API standard -- ISO/IEC 15435

722	Issues list #3 for ISO 15435	Keld Simonsen	2000-05-22	22.15435
754	Draft ISO 15435 – Internationalization APIs	Keld Simonsen	2000-05-22	22.15435

New draft distributed in the meeting.

Keld reports about the changes to the earlier document. All references to 14652 are informative. Errors are corrected, some functions removed.

Ken: main response is that nothing substantial structural was addressed in this new draft. Ken and Kent made recommendations in Tel Aviv and in Copenhagen, these are not addressed. Ken suggests that there is no justification for that work. Keld points out that at least 5 member bodies have approved the project. Marc adds that the document is not useful from German's point of view. US and Sweden are very critical of the project. Ken says that Unicode and L2 are opposed to the work. Clarification of the process: if the CD ballots fails multiple times, the WG has the option to withdraw the project or make the document a different type (e.g. TR).

Marc: there might be a need for generic functionality description, not a C-specification.

Keld: we can provide the functionality description as was planned for 14652. We need portability of I18N support through standardized APIs. 15435 defines a superset of such functionality.

Ken: 15435 is not of interest to any Unicode implementer. Java has already solved portability. The standard has no relevance for Macs, for Windows, and for object oriented systems. These systems have their APIs in place, 15435 is mainly POSIX related. The portability addressed in 15435 relates to UNIX flavors and POSIX.

Marc: distinguish between functionality and prescribed formats. 15435 is for C-language, not for anything else.

Keld: Java is not a JTC1 standard, but widely used.

Clews: should fit better into a C working group. A working paper could be useful, that also includes 14652 parts.

Ken: we need input from other language groups regarding the value of these APIs for the other languages. Cross-platform libraries for internationalization are needed, not C-based APIs.

Keld claims that this standard is needed, the WG asks for proof from "customers" e.g. LI18NUX, and the SC22 working groups. Hideki Hiura reported that 15435 was rejected in the LI18NUX group, they develop their own.

Winkler suggests to question the "customers" about the perceived value of 15435 for their environment.

Possible customers to be asked for input: W3C-I18N-IG, LI18NUX, working groups in SC22.

POSIX will not address it in the revisions of the standard. The Austin group might be interested in taking API work into consideration.

Ken recommends: make it a clear C specification which can be used by other programming languages. But before that, we need to check the need for the standard at all. The standard MUST address ISO 10646. (Consider ICU class availability as an alternative).

Discussion of the current draft N754, only technical issues.

Why is API defined in this standard, is this not defined in e.g. C or POSIX or somewhere else ?

Definitions of transliteration and transcription are incorrect, use existing definitions.

Data type definition is incoherent – do a C definition clearly (stuct, array, pointer). String is a pointer. What it points to is an array. Internal and external representation should go into the definitions.

"Character encoding" is locale independent. Drop ASN.1 specification.

5.1.1, 5.1.2, 5.1.3 can be replaced by a note to the C standard.

Make a public struct .

Make returns consistent.

5.2.1 – 5.2.3 - type specific call back definitions are needed (indirect parts of the API).

Detailed comments from Ken to make the standard more useful.

5.4 needs complete struct definition.

5.4.1 newlocale: typedef needed, …

5.4.3 modifylocale: needs 2 fully formed locales – better way is the public struct …

5.4.4. and 5.4.5 are unnecessary, due to having a public struct.

6.4 no toupper and tolower in stringtrans

7. string comparison is designed different from Java, ICU, etc…

7.2 clarify IGNORE_SPECIALS

7.3 is supposed to prepare a binary string for comparison – output length is missing, what are the result values? Memcomp with length needs to be used.

8.1 message formatting needs to be clarified – passing number ?

Make results consistent.

Annex A: Rationale

More comments to the annexes

Issues list #4:

Binding to C.

Check implementability

Spelling – NO

Wordbreaks and line breaks – NO

API should be thread-safe

C++ binding – NO.

16. ISO/IEC 10646 Issues

No additional issues.

17. Other business

17.1 Convenors report 2000

Checked and corrected by the group.. Convenor will submit to SC22.

17.2 CLAUI TD coordination meeting

| 758 | Excerpt of SC35 resolutions with impact on WG20 | SC35 | 2000-05-22 | Admin |

There are plans to hold a meeting in France, October 19-20, 2000. Arnold doubts that he will be able to attend this meeting. He asks Alain to represent WG20 at the meeting, supported by Keld.

AI Alain: send TR 10176 to SC35 as additional input to their new work item on I18N horizontal issues.

AI Keld: find JTC1 decision that extends the responsibility of WG20 for I18N outside of SC22 in JTC1.

AI Winkler: distribute agenda of the meeting in France, when it becomes available, requesting input from WG20 for our representatives (Alain, Keld).

18. Review of Priorities and Target Dates

Reviewed and agreed.

19. Review of Actions Items from this meeting

SD-5	Action item list	Winkler	2000-05-24	admin

Reviewed and will be corrected with input from the group.

20. Approval of Resolutions

748	Québec, resolutions, May 24, 2000	Winkler	2000-05-24	Admin

All resolutions were approved unanimously, resolution #6 per acclamation.

21. Adjournment

The meeting was adjourned on May 24, 2000 16:25h.