This presentation was prepared for the 2016 IUC conference. It describes the mission, membership, and various working groups and committees that make up the Consortium.
Unicode Growth

Usage of character encodings for websites, 15 Dec 2014, W3Techs.com
Common Locale Data Repository (CLDR) Growth Over Versions
Mission of the Unicode Consortium

Enable all the world's languages for use on computers everywhere, mobile to supercomputers

• Non-profit organization
• Funded by membership fees and donations
• Our specifications and data freely available

Work done by volunteers with key funding through Adopt-A-Character
Membership

- **Full:** Voting members, run the consortium
- **Institutional:** Voting members in technical committees
- **Supporting:** Voting members in technical committees (1/2 vote)
- **Associate:** Non-voting, contribute to technical work
- **Individual and Student:** Non-voting, contribute to technical work
- **Liaison:** Non-voting, contribute to technical work
Where is the office?

Inside the Microsoft Silicon Valley Campus Building 5, Mountain View
Bird’s Eye View

North of Highway 101
Who are the employees?

ELLEN

JULIE

RICK
Who are the officers?

Executive Officers
- Mark Davis
- Lisa Moore
- Michel Suignard
- Peter Constable

Technical Officers
- Rick McGowan
- Deborah Anderson
- Ken Whistler
- Markus Scherer
- Peter Edberg

Technical Committee Chairs / Vice Chairs
- Steven Loomis
- Craig Cummings
- John Emmons
- Yoshito Umaoka
- Markus Scherer
- Some other officers are also chairs...

http://www.unicode.org/consortium/directors.html
How is the Corporation Organized?

Board of Directors (BOD)
- Elected by Full Members

Executive and Technical Officers (EMG)
- Appointed/ratified by the BOD

Full Members
- Dues paying members of the corporation elect the BOD
- and empower the officers

Employees
- Hired by the officers
Editorial Committee

Many of the regular members
August 2015
Technical Committees

Unicode Technical Committee
Responsible for the Unicode Standard and its associated standards

CLDR Technical Committee
Responsible for Unicode locale development and the LDML specification

ICU Technical Committee
Responsible for ICU development

ULI Technical Committee
Responsible for Improving localization interoperability

Emoji Subcommittee

Script Ad Hoc Group

For more detail, see this webpage
Perhaps the most interesting of the committees... because we get guests!

Some UTC members with guests at UTC #141 in Sunnyvale
Unicode Technical Committee - 2

Break time during a more typical meeting...
How can I find out what UTC is doing?

UTC working documents publicly available:
• [http://www.unicode.org/L2/index.html](http://www.unicode.org/L2/index.html)
• Meeting agendas and minutes included
• Document submission open
• Document updates announced to members
Emoji Subcommittee

Meets exclusively by phone

Twice every week...

• The wild and the wooly
Emoji Subcommittee - Additions

1. Verify the proposed emoji not already encoded
2. Submit a proposal following guidelines in UTR #51
3. Emoji proposals reviewed by Emoji subcommittee and top candidates are forwarded to the UTC
4. Unicode (and ISO/IEC 10646) review, vote on acceptance, and publish
5. Fonts and keyboards then created and software updated
CLDR - TC

Meets exclusively by phone every week...

- CLDR design & changes
- Bug database
- Survey Tool
CLDR-TC...continued

- Common Locale Data Repository
- XML / JSON data
- LDML Specification (UTR #35)
- Tools (Java) conversion utilities and CLDR utility classes.
- Survey tool for data input

Meets every Wednesday at 8:00 AM Pacific Time.
- (also on Friday as necessary)
CLDR 29 & 30

Unicode 9.0
  ◦ Also UCA & Unihan 9.0

Display names for more languages
More likelySubtags, scripts for more languages
Cantonese locale added (+nds & ckb in ICU)
“3 Fridays ago”, “this hour”
Which units used in which country/region
Improved emoji annotations
Many more transforms; all have BCP 47 IDs
+9.2% data, 5.9% fixes
CLDR – TC...continued

CLDR Versions 29 & 30
• Unicode 9.0
• Also UCA & Unihan 9.0
• Display names for more languages
• More likelySubtags, scripts for more languages
• Cantonese locale added (+ nds & ckb in ICU)
• “3 Fridays ago”, “this hour”
• Which units used in which country/region
• Improved emoji annotations
• Many more transforms; all have BCP 47 IDs
• +9.2% data, 5.9% fixes
ULI - TC

Meets exclusively by phone
bi-monthly

- Focus on interchange of localization assets
- Hosts archives of GMX-V and SRX
- Suggests extensions to TMX and XLIFF
Sentence Break:

<table>
<thead>
<tr>
<th>Thanks to Dr. Jones for this effort.</th>
<th><em>(UAX 11)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Thanks to Dr. Jones for this effort.</td>
<td><em>(ULI data in CLDR)</em></td>
</tr>
</tbody>
</table>
  - “Dr.” is an abbreviation.

Data (from DBPedia)

**English**: Mr. Mrs. Dr. St. ...

**Spanish**: Sr. Dto. Sra. Avda. ...

**Russian**: проф. февр. тел. кв. ...

Future: Word count, message format, ...
ICU - TC

Meets exclusively by phone every week...

- New member in Unicode family
  - Project management was transferred from IBM in May 2016
  - ICU 58 (Oct 2016) is distributed under the Unicode software and data license
- Providing cross-platform reference implementation of Unicode standards, CLDR and ULI in C and Java.
Projects of the Consortium

**Unicode Standard**, with Standard Annexes and data

**Unicode Technical Reports** & **Technical Notes** (UTR, UTN)

**CLDR Repository** and Survey Tool (CLDR)

**ICU Development**

**ISO 15924** Registry of Script Codes (ISO 15924)

**Ideographic Variation Database** (IVD)

**UDHR** in many languages repository
How to keep up with Unicode news

The Unicode Blog:
- [http://blog.unicode.org/](http://blog.unicode.org/)
- Announcements
- RSS Feed

Home Page Announcements section

Public Mail Lists:
- Unicode, CLDR, etc
- [http://www.unicode.org/consortium/distlist.html](http://www.unicode.org/consortium/distlist.html)
Script development
How You Can Get Involved....Proposals!
How Unicode Proposal Process Works

1. A script/character not in Unicode is identified
How Unicode Proposal Process Works

1. A script/characters not in Unicode is identified

   Verify it is not already in Unicode (or in process)
   - check latest code charts and Pipeline page on unicode.org
How Unicode Proposal Process Works

1. A script/characters not in Unicode is identified
   Verify it is not already in Unicode (or in process)
   - check latest code charts and Pipeline page on unicode.org
   - send in a query via the Unicode Contact form or ask on Unicode email list
How Unicode Proposal Process Works

1. A script/characters not in Unicode is identified and verified it is eligible and not already in Unicode (or in process)

2. A Unicode script proposal is written

---

**JTC1/SC2/WG2 N3867R2**

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

---

**Doc Type:** Working Group Document

**Title:** Proposal for encoding the Palmyrene script in the SMP of the UCS

**Source:** UC Berkeley Script Encoding Initiative (Universal Scripts Project)

**Status:** Individual Contribution

**Action:** For consideration by JTC1/SC2/WG2 and UTC

**Date:** 2010-08-17

---

1. **Introduction.** The Palmyrene alphabet was used from the first century BCE, in a small independent state established in the Syrian desert between Damascus and the Euphrates. However, Palmyrene inscriptions are also found in Hatra, in Dura Europos and other Syrian cities, and even in Rome, where
How Unicode Proposal Process Works

1. A script/characters not in Unicode is identified and verified it is eligible and not already in Unicode (or in process)

2. A Unicode script proposal is written

• If you want to write a proposal....
  ❖ Set aside time to learn about proposal requirements
  ❖ Read “Submitting Character Proposals” page
  ❖ Work with a UTC member
  ❖ Find users and involve them
How Unicode Proposal Process Works

1. A script/characters not in Unicode is identified and verified it is eligible and not already in Unicode (or in process)

2. A Unicode script proposal is written

3. Two standards committees review proposals and vote whether to accept them

Unicode Technical Committee (left) ISO SC2 (WG2) (below)
How Unicode Proposal Process Works

1. A script/characters not in Unicode is identified and verified it is eligible and not already in Unicode (or in process)
2. A Unicode script proposal is written
3. Two standards committees review proposals and vote whether to accept them
4. Unicode (and ISO/IEC 10646) publish the standard with scripts
5. Fonts and keyboards are created and software updated
How You Can Get Involved -- Review

- Review code charts (as part of beta review for Unicode release or ISO ballots)
How You Can Get Involved - Comments

• Send in comments on new script and character proposals in the Document Registry via the Contact form
How You Can Get Involved - Fonts

• If your expertise is in fonts, help create or improve fonts for scripts in Unicode,
  • Especially needed for lesser-known modern and historical scripts
Encourage Your Company/Organization to Join

- Encourage your company or organization to join, or join as an individual
Questions?