

# Registry of Cultural Conventions

**Tateo Koike**

Visiting Senior Scientist

International Information Technology Laboratory

Center of the International Cooperation for Computerization

## Standards/Projects of the Cultural Conventions

- ISO/IEC TR 11017 - Framework for Internationalization
- ISO/IEC 9945 - Portable Operating System Interface (POSIX)
- ISO/IEC TR 14652 - Specification method  
for cultural conventions
- ISO/IEC 15897 - Procedure for registration  
of cultural conventions
- Unicode: Common Locale Data Repository (CLDR)
- Data Book of Cultural Convention in Asian Countries  
(AFSIT-SIG of CICC)



## Requirements for the “Global Information Society”

- **User-friendly Interface** is required
  - Input and Output communication to/from a system is in user's language and in its character
  - The character data is processed in natural rules for user
  - Presentation of cultural conventions becomes natural to the user

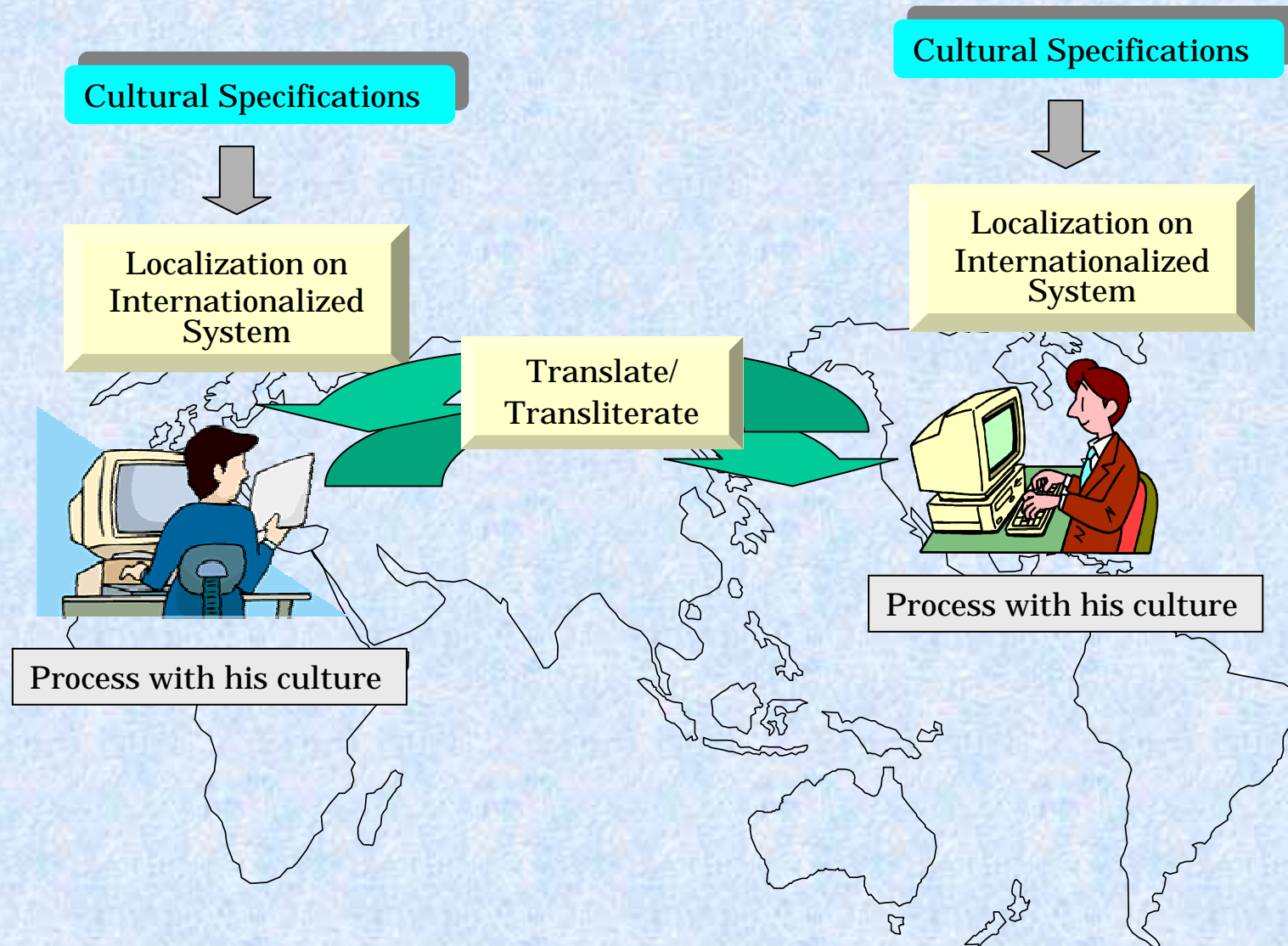
## What is cultural dependence?

Examples;

	CharaString	Some Culture	Other Culture
Date	10/12/04	Oct. 12th, 2004	Dec. 10th, 2004
Time	half seven	7:30	6:30
Yes/No sign	○	Yes	No



# “Cultural Convention System” - JTC1 i18N/L10N Model



## Category of User Cultures for the Internationalization - TR 11017

- Mono-culture/script/language
- Mono-culture/script/language, including ISO/IEC 646
- Mono-culture/language, Multi-script
- Mono-culture, Multi-script/language
- Multi-culture/script/language
  - Sequential multi-lingual
  - Concurrent multi-lingual
- Global Uniformity
- Cross-cultural friendliness



## Items of culture dependence - TR 11017

### Most Important Items;

- Currency symbols
- Calendar
- Date format
- Time format
- Number format
- Case mapping
- String ordering and comparison

## Items of culture dependence - TR 11017 (cont.)

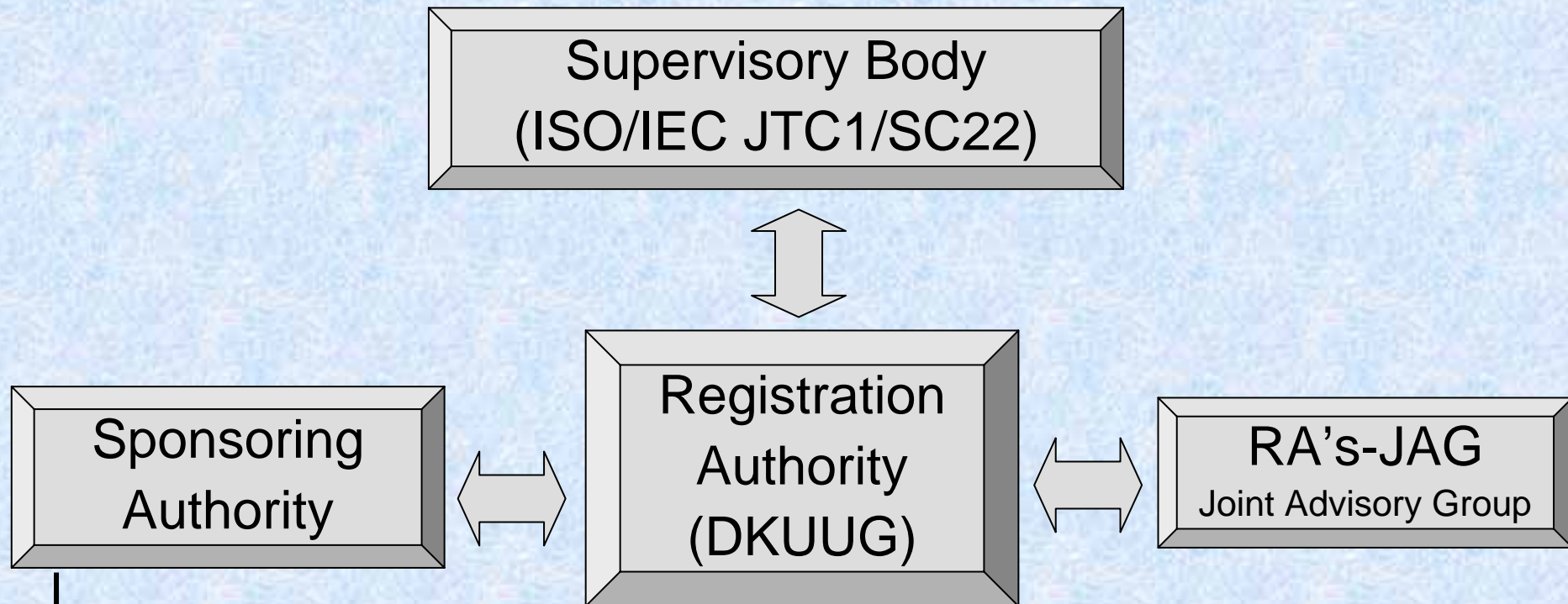
### Other items;

- Script (from the process viewpoint)
- Character encoding and handling
- Write direction
- Multiple forms of characters
- Hyphenations, Spacing and Punctuation
- Expression of numbers as word
- Character classification
- Message and Dialogues
- Text length
- Spelling
- Documentation
- Layout of text on a page
- Character size, Line size and Line spacing
- Font requirement
- Drawing of boxes in text
- Icons and Symbols
- Mathematical symbols
- Paper size
- Data input method
- Voice message
- Number rounding
- Price expression
- Telephone number format
- Postal address format
- Measurement system
- Color signification
- Function names
- Personal titles
- Taboo words
- Regulatory requirement regarding IT



## Procedure of International Registration - ISO/IEC 15897

Authority



{ Member bodies of CEN or JTC1  
 { CEN/TC304  
 { SC and WG of JTC1

DKUUG: Danish UNIX-systems User Group

## Type of cultural specifications

Type 1: Narrative Cultural Specification

Type 2: POSIX Locale

Type 3: POSIX Character Map

Type 4: POSIX Repertoire Map

Type 5: Machine-parsable\* cultural specification

Type 6: Machine-parsable\* coded character set

\*: specified in TR 14652, XML or SGML

## Media for application form

- Internet
- Other electronic media
- Paper



## Registration procedure

- Sponsoring Authority(SA) sends application form to the Registration Authority(RA)
- RA checks the validity of the application form
- RA circulates the application form to JTC1/SC22 members, liaisons and RA-JAG for 3 month reviewing
- RA assigns the numeric and token-ID
- RA announces to JTC1/SC22 members and liaisons the cultural specification registered

## Narrative Cultural Specification

- Narrative Cultural Specification is the explanatory note of the Locale/FDCC-set specification
- For implementation of the culture dependence without misinterpretation, the person in the culture should describe the Narrative Cultural Specification clearly
- Because, the Narrative Cultural Specification in one culture can be described only by a person who belongs to the culture



## Required Items

Clause1: Alphanumeric deterministic ordering  
(list the national standards)

Clause2: Classification of characters

Clause3: Numeric format

Clause4: Monetary format

Clause5: Date and time conventions

Clause6: Affirmative and negative answers

## Additional Items

- Not related POSIX Locale

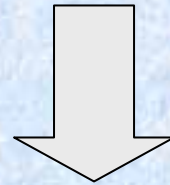
- Clause7: National or cultural information technology terminology
- Clause8: National or cultural profiles of standards
- Clause9: Character set considerations
- Clause10: Sorting and searching rules
- Clause11: Transformation of characters
- Clause12: Character properties
- Clause13: Use of special characters
- Clause14: Character rendition
- Clause15: Character inputting
- Clause16: Personal names rules
- Clause17: Inflection
- Clause18: Hyphenation
- Clause19: Spelling
- Clause20: Numbering, ordinals and measuring systems



## Additional Items (cont.)

- Clause21: Monetary amounts
- Clause22: Date and time
- Clause23: Coding of national entities
- Clause24: Telephone numbers
- Clause25: Mail(postal) address
- Clause26: Identification of persons and organizations
- Clause27: Electronic mail address
- Clause28: Payment account numbers
- Clause29: Keyboard layout
- Clause30: Man-machine dialogue
- Clause31: Paper format
- Clause32: Typographical conventions

## Sample of the Narrative Cultural Specification



See <http://std.dkuug.dk/cultreg/registrations/number/2>



## FDCC-set (Locale)

### Formal Definition of Cultural Conventions-set

The FDCC-set defines the user's IT environment that depends on language, territory, or other cultural customs.

LC_IDENTIFICATION	Versions and status of categories.
LC_CTYPE	Character classification, case conversion and code transformation.
LC_COLLATE	Collation order.
LC_TIME	Date and time format.
LC_NUMERIC	Numeric, non-monetary formatting.
LC_MONETARY	Monetary formatting.
LC_MESSAGES	Format of informative and diagnostic messages and interactive responses.
LC_XLITERATE	Character transliteration.
LC_NAME	Format of writing personal names.
LC_ADDRESS	Format of postal addresses.
LC_TELEPHONE	Format for telephone numbers, and other telephone information.

## Example of FDCC-set coding

### Example

```

LC_TELEPHONE
% This is the ISO/IEC TR 14652 “i18n” definition for
% the LC_TELEPHONE category.
tel_int_fmt
    “<U002B><U0025><U0020><U0025><U0061><U0025><U0074
>/
<U0025><U006C>”
% “+%c %a%t%1” which is
% +country area local
END LC_TELEPHONE

```

Note: The line which begins with the ” % “is a comment line.

Sample of the FDCC-set  <http://std.dkuug.dk/cultreg/registrations/number/7>



## charamap

- The charmap defines the mapping information between symbolic character name in the FDCC-set and actual encoding of coded character set that is used in the application system.

### Example 1

```
<escape_char>          /
<comment_char>        %
% The 7-bit charmap defines both control
% and graphic characters
<code_set_name>       "eastern7bit"
<escseq2022> "c0";"c0","/x21/x40"
<escseq2022> "go";"g0","/x28/x48"
<escseq2022> "g1";"g0","/x28/x48"
<escseq2022> "g2";"g0","/x28/x48"
<escseq2022> "g3";"g0","/x28/x48"
```

```
CHAMAP
<tab>          /x08
<newline>     /x0D
<a>           /x61
<b>           /x62
.
.
END CHAMAP
```

Sample of the charamap

<http://std.dkuug.dk/cultreg/registrations/number/11>

## Repertoiremap

- The FDCC-set and charmap sources may be specified in a coded character set using symbolic character names.
- The relation between the symbolic character names and characters may be specified via a repertoiremap.
- The repertoire mapping is defined by specifying the symbolic name and the ISO/IEC 10646 code position in hexadecimal form and optionally the long ISO/IEC 10646 character name.



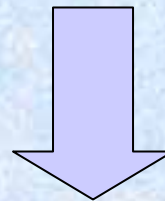
## Example of repertoiremap coding

<A>	<U0041>	LATIN CAPITAL LETTER A
<B>	<U0042>	LATIN CAPITAL LETTER B
<C>	<U0043>	LATIN CAPITAL LETTER C
<D>	<U0044>	LATIN CAPITAL LETTER D
<E>	<U0045>	LATIN CAPITAL LETTER E
<F>	<U0046>	LATIN CAPITAL LETTER F
	.	
<A->	<U0100>	LATIN CAPITAL LETTER A WITH MACRON
<a->	<U0101>	LATIN SMALL LETTER A WITH MACRON
<A(>	<U0102>	LATIN CAPITAL LETTER A WITH BREVE
<a(>	<U0103>	LATIN SMALL LETTER A WITH BREVE
<A;>	<U0104>	LATIN CAPITAL LETTER A WITH OGONEK
<a;>	<U0105>	LATIN SMALL LETTER A WITH OGONEK
<C'>	<U0106>	LATIN CAPITAL LETTER C WITH ACUTE
<c'>	<U0107>	LATIN SMALL LETTER C WITH ACUTE
	.	

Sample of the repertoiremap <http://std.dkuug.dk/cultreg/registrations/number/1>

## Cultural Specification in Asia

- AFSIT-SIG discussed the cultural conventions in Asia region from 1992 to 1995.
- The result is shown in the “Data Book of Cultural Convention in Asian Countries” which was published in October, 1996.



See

<http://www.cicc.or.jp/english/hyoujyunka/databook/contents.htm>



## In conclusion

### Registered Cultural Specifications

171 of Cultural Specifications have been registered as of July 10, 2004.

--> See <http://std.dkuug.dk/cultreg/registrations/chreg.htm>

### CICC Expectations

- **Add your Cultural Specification to the Data Book**
- **Describe the Narrative Cultural Specification of your country (culture)**