### multilingual/multiscript country code tags
### http://mltf.org/draft-mltf-jfcm-cctags-01.txt

Status of this Memo

Copyright Notice

Abstract

This memo documents the multilingual and multiscript country code
tags that are derived from ISO 3166-1:2006, and their extension from
other tables and standards, which can be used in applications,
protocols, BCP 47 langtags, databases, etc. in order to index,
reference, or sort country oriented information that is expressed in
any mode and language.

Table of Contents

**1**.  **Requirements notation**

   The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
   "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL"in this
   document are to be interpreted as described in [RFC2119].


**2**.   **INTRODUCTION**

   Human society has been, past and present, organized into Nations,
   located in Countries and on Territories, as well as structured in
   States that are led by Governments.  The countries and territories'
   administrative information of the world is documented by the ISO 3166
   [ISO 3166] standard.  For thirty years it has guaranteed the Internet
   international stability by providing the multinational framework of
   the ccTLD labels [RFC 920], [RFC 1591], [ICP-1], IDNA files, and
   languages tags [RFC 4646].

   Human beings have, past and present, used a multiple mode (voice,
   signal, sign, script, music, computer programming, etc.) language
   diversity, regarding their local education, usage, and law.

   There are many reasons as to why one would want to identify the local
   context of information, and the mode and language it is, or the way
   it is to be expressed.  A means of indicating this information is by
   labeling its container with a structured identifier or "tag", which
   the various lingual contexts involved in its transport, sorting,
   processing, rendering, or understanding can take advantage of, even
   if they differ from the language and script of its content.

   While remaining consistent with every other online and offline
   application and process, any national or lingual need, architecture,
   protocol, or application can now take pragmatic advantage of the
   language and script transparent, non-Internet specific, country code
   tag system as implied by the ISO 3166:2006 version, and its recently
   published administrative language and script information.

   This memo fully documents a multilingual multiscript identifier
   system (the country code tag) derived from ISO 3166-1:2006 and a
   registry that will consistently extend it along the documented
   syntax, semantics and pragmatics, as per other standards, such as IS0
   639 [ISO 639] or Linguasphere [LS].  It can be used in applications,
   protocols, BCP 47 langtags, databases, regulations, etc. in order to
   index, structure, sort, or retrieve country or administratively-
   oriented information that is expressed in a broad diversity of modes
   and languages.  It also provides the current status of the regular
   cctag registry file.

This system respects the IETF use of the English language and the
ASCII text format [RFC 2026] as unique references for operators,
developers, and users.  This is for utility [RFC 3935] and security
obvious reasons in a multilingual cross-technology operational
environment.  This system also assists database and registry
interoperability [ISO 11179], multilingual ontology interoperability
or interintelligibility, and relational sociability.


## 3.  Special mentions [to be removed]

### 3.1.  Relation to ISO 3166

As per the ISO 3166 rules, this memo along with its annexes have been
copied to the ISO 3166 Maintenance Agency.

### 3.2.  Discussion mailing list

This document is being discussed on the http://mltf.org/mail.htm
mailing list.

### 3.3.  Parallel works

The Multi-Lingual Technical Forum [MLTF] has engaged in several
parallel works that concern the world digital ecosystem (WDE)
multilinguistic support evolution within the five areas of material
plug to plug interconnectivity, end to end interoperability, brain to
brain interintelligibility (semantic interoperability), folk to folk
intersociability (pragmatic interoperability, and of Multilingual/
Polynym Ontologies, and the Multi-National Information Center
[MULTINIC] to support Multilingual Distributed Referential Systems
(MDRS).

NB.  When considering a garden of various lingually equivalent
ontologies, one tends to call them a "Multilingual Ontology" if they
adapt to the constraints of the language diversity, and a Polynym
Ontology or Polynym Multilinguual Ontology if they enforce a strict
polynymism, i.e. a bijective equivalence of metadata and data.  A
Multilingual Ontology will be said "multiintellegible" and a Polynym
Ontology will be said "multioperable".


## 4.  CCTAGs

ASCII Country Code Tags (cctags) are the strict equivalent to their
own locally authoritative version in the script and language that
they refer to.  This memo documents cctags in English and the ASCII
compact format.  Following RFC 4260 [RFC 4260] considerations, this

format is alphanumeric.  They could be equally validly documented in
any other language, script, and syntax.

Their different forms for every language and script, strictly have
the same meaning and value.  This is why cctags are said to be
"polynym".

Four classes of cctags are considered:

## 4.1.  Basic cctags

These cctags are only formed by the ISO 3166 country code elements,
mode, and language that are implied by the context, in which they can
be used for pragmatics needs or for application compatibility
reasons.

## 4.2.  Regular cctags

These cctags are formed only from ASCII elements that are obtained
from ISO 3166-1:2006 country and language alpha2 codes and user
extensions that have been reported to the ISO 3166/MA, separated by
the numeric of the mode.  They are five alphanumeric long (up to
seven in case of some specific modes.

## 4.3.  Extended cctags

There are several ASCII cctag types that are formed along the same
syntax as regular cctags in using other tables and standards.  Their
purpose is to extend the coverage and use of cctags.  They can be up
to eight alphanumeric long, and even to ten in case of some specific
modes.

## 4.4.  Numeric cctags

Numeric cctags are a universal (language and mode independent) form
of the cctags.  They can be formed by using ISO 3166 and other
standards' numeric tables, or any algorithm to transform them in a
numeric grid or an IPv6 sub-addressing.  This memo does not cover
numeric cctags.


## 5.  Syntax of the cctags

The syntax of the regular cctags concatenates country, mode, and
language code elements into five (or more) alphanumeric formatted
strings.

## 5.1.  Country code

Country and territory code elements come from the ISO 3166-1 alpha2
list and from Appendix A (Geographic Aggregations).  Extended cctags
can be builtby using ISO 3166-1 alpha3 [ISO 3166] and ISO 3166-3 [ISO
3166] alpha4 lists.

## 5.2.  Language information

Language code elements come from ISO 3166 (administrative languages)
and Appendix B (alpha2 code elements of ISO 3166 administrative
languages that are missing in ISO 639-1).  Extended cctags can be
formed with ISO 639-2 and 639-3 code elements.

## 5.3.  Mode number/information

Languages can be used in different modes.  The mode code elements can
be one, two, or more characters long.  The first character is always
numeric, whereas the other characters, if any, are always alpha.  The
mode code element list is documented in Appendix C (UNIMODE table).

The default mode ("0") corresponds to a "trihexadecimal" character
set, i.e. a numbering from 0 to Z, which corresponds to the current
English ASCII version of ISO 3166.  Trihexadecimal does not change
the current usage, but it can help interoperability with other
standards and non-ASCII tables.  It is noted as "Trih" script.

## 5.4.  Examples

In using the documented components shown above, one can form cctags
such as:

"FR0FR" - French from France in pure ASCII (without diacritics)

"FR1FR" - the same but with diacritics

"CN0CN" - ASCII Romanized Chinese from China

"SG1CN" - Han written Chinese from Singapore


## 6.  Different uses of cctags

The polynym nature of cctags enables its use as:

## 6.1.  labels

They can be used in order to name a local item that is specific of a
mode and language in a file, object, document, application, etc.  For
example, a French airport information audiogram for Chinese
passengers can be labeled FR1CN.

## 6.2.  multilingual and multimode headers

They enable the tagging and sorting of multilingual and multimode
information.  The cctag sort is made in the sorting order of the
cctag script and the content sort is made in the sorting order of its
own script as documented by the cctag.

## 6.3.  roots of information space metastructures

The cctags enables the creation of country code oriented name
(taxonomy) and information spaces that can be used in ISO 11179
conformant registries [ISO 11179], Ontology building, or Dynamic
Delegation Discovery Systems (DDDS) [RFC 3401].

This will enable the mapping of unique cctag headed multi-taxa (class
and attribute labels) strings to data stored within DDDS databases,
or to resolve the absolute or relative network address of their
distributed description instances.

## 6.4.  Integration to the DDDS network architecture

DDDSs are a proven part of the Internet architecture (the DNS is a
DDDS) that might particularly match the distributed (multi-zone) and
multilingual network constraints well.  The "first well known rule"
that will be applied to the cctag headed application "unique strings"
will be the cctag polynym table itself (the list of its own language
and mode versions).  It will "rewrite" the cctag based metastructure
portion of the unique string into the application mode and language.

## 6.5.  Digital Ecosystem cctag taxonomy

A digital ecosystem oriented cctag taxonomy is considered by the
MULTINIC project.  A Draft taxonomy is introduced in Appendix D.

## 6.6.  Examples

Using that taxonomy in rather simple semantic cctag headed unique
strings that can be used to define

o   a national ccTLD: "FR0FR"."ccTLD" : will be rewritten as ".fr"

o   a country name: "CI0FR"."country name" : will be rewritten as
    "Cote d'Ivoire"

o   a national TLD in a registry: "nationalTLD.cn1cn.reg.org" : will
    rewritten as ".zhangguo.reg.org"


## 6.7.  BCP 47 langtags

   BCP 47 permits the use of cctags as private agreement subtags in the
   "x-" private use area.  This can enable easy RFC 4646
   interoperability for national administrative languages variants and
   extensions.  In order to avoid confusion with other private use
   subtags, the syntax to be "x-cc-" followed by a cctag.  Subsequent
   "-" can be used as further information separators.

   Example: "x-cc-cn1fr" can tag a legal Chinese document writen in
   Latin script French.


## 7.  Polynym ccTAGs

   Polynym ccTAGs (also called pcTAGs in case of possible confusion) are
   ccTAGs built after a locally authoritative ISO 3166 polynym table.
   They are strictly equivalent.  They can be handled by applications in
   two different ways:

## 7.1.  pcTAG local usage

   When pcTAG use a single script, they can be directly used if the
   application supports this script.

## 7.2.  pcTAG global usage

   In other cases they should be interlinked ("polynym interlink") to
   the trihexadecimal version of the country and language table (the
   mode stays in hexadecimal).

## 7.3.  mixed ccTAGs

   In the case of ccTAGs using different scripts for the country and
   language code elements polynym interlinking should be first applied.

8.  **Further work**

   This section is not a part of the cctag current state of the art
   description.  It only aims at providing Internet community Members
   with information on the possible extensions of the cctag
   considerations, in order to assist the interoperability of their own
   projects or for them to participate in the corresponding work.

8.1.  **Mode definition**

   Modes are understood as belonging to a congruent set of content
   expression characteristics.  They may range from vocality to specific
   scripts, or particular orthographies.

8.2.  **Other legitimate cctags**

   Other cctags can be privately formed or researched, which might use a
   digital format that is based on number lists, or other language code
   elements from: ISO 639 series, or widely approved, established, or
   open language emergence tables, codes, or directories of language
   referents.

   Emergences are the common properties that arise from syneergy within
   a concept cluster.

   Referents can be any form of a dictionary, online language tables,
   etc. that authors or interlocutors may refer to.  This memo does not
   cover language referents.

8.3.  **Taxonomy related work**

   The number of taxonomies that can be associated or that can take
   advantage of cctags is quasi unlimited.  Documenting them should be
   carried out in cooperation with other endeavors.

   Some of the works that could possible be considered are:

   o  taxonomy garden management (taxonomies not belonging to the same
      class)

   o  integration of date and epistemological elements

   o  redundancy reduction: to avoid the same information being entered
      twice.

o   polynimy - multilingual thesauri

o   polylingualisation - the support of several languages in the same
    content.


8.4.  pragmatics related work

In linguistics, syntax is the literal organization, semantics is the
literal meaning, and pragmatics is the implied meaning of an idea.
Access to referent registries that determine the sentence meaning and
access to context and personal registries (that influences the
speaker's meaning and the reader's interpretation) are essential to
the inter-relational process.

Work on networking pragmatic competence assistance (also called
"extended services") should extensively benefit from cctags since
most of the norms, standards, cultural, local, and personal models,
and relational contexts are country related and expressed in their
administrative language or in the administrative language of its
visitors or interlocutors.

o   multilingual distributed referent systems

o   non-linear DDDS and interlink systems?


8.5.  langtags local variants and extensions

Pragmatics mainly sort the utterances by language, author/reader,
date, or location.  RFC 4646 sorts them by language and regalian
location considered as geographic, and cctags by regalian location
and languages.  Work on the use of the cctag metastructure, in an RFC
4646 consistent manner, could help the preservation of
interintelligibility, and even conditional interoperability, between
these two schemes.


9.  MULTINIC Registry

This memo does not cover the MULTINIC repository.  Its purpose is to
host cctag based and multilaterally sourced multilingual information
ontologies and registries.  They should encompass data of common good
necessary to navigate the real (norms), technical (standards),
virtual (models), and relational (contexts) worlds and to master
their multiple technologies in the world digital ecosystem as well as
their permanent evolution.

The registry files documented in Appendixes are:

o   Appendix A - Geographic Aggregation table
    This registry documents ISO 3166 conformant aggregation as per ISO
    3166-1 8.1.2. and 8.1.3. clauses.  The ISO 3166 MA has been copied
    to the present Draft.

o   Appendix B - Element of information missing in ISO 3166
    This registry documents the information that is missing in ISO
    3166 to fully support cctags.

o   Appendix C - The UNIMODE table
    This registry documents the mode related information.  Its syntax
    is one numeric, which is optionally followed by one or several
    alphas.

o   Appendix D - a cctag taxonomy
    These are initially suggested taxa matching Internet needs.  New
    taxa can be added as they become necessary, and the online version
    of this list can be subsequently extended.

o   Appendix E - the CCTAG table
    The ISO 3166 CCTAG resulting ASCII code, documenting the
    corresponding country, script, and administrative language for
    each CCTAG.

    These files will be maintained by the MULTINIC and should be
    distributed through MDRS.


10.  Security Considerations

   This memo does not introduce any security related issue.  However, it
   is to be considered that in strictly abiding by RFC 1591 and the
   definition of e-countries from ISO 3166, and by the positions of the
   World Submit on Information Society [WSIS] regarding national naming
   sovereignty, it may conflict with alternative projects.  The ICANN
   IDN working group investigates such an alternative project, which may
   want to consider that the territories covered by IDNs could differ
   from the territories Internet Communities named by the ASCII ccTLDs,
   and thereby, departing from regular cctags.  However, nothing opposes
   extended cctags being used, based on ISO 3166-2 or ISO 3166-2
   aggregations.

## 11.  IANA Considerations

   This memo does not call for any IANA involvement.  However, the IANA
   may freely decide to host the MULTINIC's files.  Currently, the
   MULTINIC update service to the MDRS is not documented.

Appendix A.  **Geographic Aggregations table**

   Geographic Aggregations table Documents ISO 3166 conformant
   aggregation as per ISO 3166-1 8.1.2. and 8.1.3. clauses: ISO 3166 MA
   Members has been copied the present Draft.

            ISO 3166 customized codes for geographic aggregations

```
     +------+--------+--------+-----+--------------------------------+
     | CONT | alpha2 | alpha3 | NR  | Description                    |
     +------+--------+--------+-----+--------------------------------+
     | QT   | QT     | ZQT    | 001 | World                          |
     | QM   | QM     | ZQM    | 002 | Africa                         |
     | QM   | XA     | ZXA    | 014 | Eastern Africa                 |
     | QM   | XB     | ZXB    | 017 | Middle Africa                  |
     | QM   | XC     | ZXC    | 015 | Northern Africa                |
     | QM   | XD     | ZXD    | 018 | Southern Africa                |
     | QM   | XE     | ZXE    | 011 | Western Africa                 |
     | QN   | QN     | ZQN    | 019 | Americas                       |
     | QN   | QO     | ZQO    | 419 | Latin America and the Caribbean |
     | QN   | QP     | ZQP    | 021 | Northern America b/            |
     | QN   | XF     | ZXF    | 029 | Caribbean                      |
     | QN   | XG     | ZXG    | 013 | Central America                |
     | QN   | XH     | ZXH    | 005 | South America                  |
     | QO   | QO     | ZQO    | 419 | Latin America and the Caribbean |
     | QP   | QP     | ZQP    | 021 | Northern America               |
     | QQ   | QQ     | ZQQ    | 142 | Asia                           |
     | QQ   | XI     | ZXI    | 143 | Central Asia                   |
     | QQ   | XJ     | ZXJ    | 030 | Eastern Asia                   |
     | QQ   | XK     | ZXK    | 034 | Southern Asia                  |
     | QQ   | XL     | ZXL    | 035 | South-Eastern Asia             |
     | QQ   | XM     | ZXM    | 145 | Western Asia                   |
     | QR   | QR     | ZQR    | 150 | Europe                         |
     | QR   | XN     | ZXN    | 151 | Eastern Europe                 |
     | QR   | XO     | ZXO    | 154 | Northern Europe                |
     | QR   | XP     | ZXP    | 039 | Southern Europe                |
     | QR   | XQ     | ZXQ    | 155 | Western Europe                 |
     | QS   | QS     | ZQS    | 009 | Oceania                        |
     | QS   | XR     | ZXR    | 053 | Australia and New Zealand      |
     | QS   | XS     | ZXS    | 054 | Melanesia                      |
     | QS   | XT     | ZXT    | 057 | Micronesia                     |
     | QS   | XU     | ZXU    | 061 | Polynesia                      |
     +------+--------+--------+-----+--------------------------------+
```

                   Table 1: Addition to 3166 - (sub)continents

ISO 3166 customized codes for geographic aggregations

```
+------+--------+--------+-----+-----------------------------------+
| CONT | alpha2 | alpha3 | NR  | Description                       |
+------+--------+--------+-----+-----------------------------------+
| QU   | QU     | ZQU    | 990 | Martime part of the World (5      |
|      |        |        |     | oceans cluster)                   |
| QV   | QV     | ZQV    | 991 | Atlantic Ocean                    |
| QW   | QW     | ZQW    | 992 | Indian Ocean                      |
| QX   | QX     | ZQX    | 993 | Pacific Ocean                     |
| QY   | QY     | ZQY    | 994 | Arctic Ocean                      |
| QZ   | QZ     | ZQZ    | 995 | Antartic Ocean                    |
+------+--------+--------+-----+-----------------------------------+
```

Table 2: Addition to 3166 - Oceans

Appendix B.  Elements of information missing in ISO 3166

   Elements of information missing in ISO 3166 Documents the information
   that is missing in ISO 3166 in order to fully support cctags.

```
          +--------+--------+-----------------------------+
          | alpha2 | alpha3 | language                    |
          +--------+--------+-----------------------------+
          | ns     | nso    | Northern Sotho, Pedi, Sepedi |
          | pu     | pau    | Palau                       |
          | sy     | cpt    | Selsewa creole              |
          | tp     | tpi    | Tok Pisin                   |
          | tu     | tet    | Tetum                       |
          | zz     | zzz    | no defined language         |
          +--------+--------+-----------------------------+
```

                    Table 3: Addition to 3166

## Appendix C.  UNIMODE table

UNIMODE table Documents one numeric, plus one or two optional alpha, code for mode information.

Value          Description

0              trihexadecimal (0-Z) numbering is equivalent to the
               ASCII character set, but adds some computing, sorting,
               etc. possibilities.

1 to 4         lists the value of other scripts associated with
               languages.

5              voice.

6 to 7         reserved for further use.

8 to 9         escape sequence to the general alpha second character
               from A to Z in order to document special modes.

**Appendix D**.   **CCTAG namespace Taxonomy**

   CCTAG namespace Taxonomy These taxa are suggested to form an initial
   Internet cctag taxonomy.  New taxa can be added as they become
   cecessary, and the online version of this list can be subsequently
   extended.  This is an embryonic taxonomy project for initiating a
   debate on the organisation and the support of this namespace.

```
    +-------+---------+----------------------------------------+
    | Class | Taxa    | Description                            |
    +-------+---------+----------------------------------------+
    | MA    |         | Maintenance Agency taxonomy            |
    | MA    | IT      | Maintenance Agency ITU information     |
    | MA    | IN      | Maintenance Agency Internet information |
    | IT    |         | ITU taxonomy                           |
    | IT    | MA      | Maintenance Agency                     |
    | IT    | E164    | E.164 code for the country             |
    | IT    | DCC     | X.121 Data Country Code.               |
    | IN    |         | Internet taxonomy                      |
    | IN    | IANA.IP | IP address of the IANA repository      |
    | IN    | IANA.DN | iana.org - IANA Repositiry Domain Name |
    +-------+---------+----------------------------------------+
```

          Table 4: Proposed cctag taxonomy for the World Digital Ecosystem
                                   (WDE)

Appendix E.  BASIC and REGULAR CCTAG Table


12.  References

12.1.  Normative References

   [RFC1591]  Postel, J., "Domain Name System Structure and Delegation",
              RFC 1591, March 1994.

   [RFC2026]  Bradner, S., "The Internet Standards Process -- Revision
              3", BCP 9, RFC 2026, October 1996.

   [RFC2119]  Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119, March 1997.

   [RFC3401]  Mealling, M., "Dynamic Delegation Discovery System (DDDS)
              Part One: The Comprehensive DDDS", RFC 3401, October 2002.

   [RFC3935]  Alvestrand, H., "A Mission Statement for the IETF",
              BCP 95, RFC 3935, October 2004.

   [RFC4646]  Phillips, A. and M. Davis, "Tags for Identifying
              Languages", BCP 47, RFC 4646, September 2006.

12.2.  Informative References

   [ICP-1]    ICANN, "ICANN/ICP-1 Internet Domain Name System Structure
              and Delegation (ccTLD Administration and Delegation)",
              1999, <http://www.icann.org/icp/icp-1.htm>.

   [ISO3166]  ISO, "ISO 3166 Codes for the representation of names of
              countries and their subdivisions", Month 2006,
              <http://iso.org>.

   [ISO3166-2]
              ISO, "ISO 3166-2 Codes for the representation of names of
              countries and their subdivisions", Month 2007,
              <http://iso.org>.


Appendix F.  Acknowledgments

   Any list of contributors is bound to be incomplete.  This one would
   have to go back far in the past and quote so many people unknown who
   contributed greatly at their level to the world stability.  I wish
   however to thank in particular the co-founders of the MLTF for the
   work achieved together and those who have made the ISO 3166-1:2006

administrative language section possible.

Author's Address

    Jean-Francois C. Morfin
    INTLNET
    23 rue Saint Honore
    Versailles  78000
    France

    Phone: (33.1) 39 50 05 10
    Email: jefsey@jefsey.com
    URI:   http://intlnet.org