L2/04-072

ISO/IEC JTC 1 N7335

2003-12-17

Replaces:

ISO/IEC JTC 1
Information Technology

Document Type: National Body Contribution

Document Title: Response to JTC1 Sophia Resolution #39: Development of a Solution for the Unambiguous Identification and Interworking of Codes Representing Countries, Languages, and Currencies (prepared on behalf of SC32/WG1)

Document Source: National Body of Canada

Project Number:

Document Status: This document is circulated to JTC 1 National Bodies for review and comment. Comments should be submitted to the JTC 1 Secretariat by the above due date.

Action ID: COM

Due Date: 2004-03-18

Distribution:

Medium:

Disk Serial No:

No. of Pages: 44

Secretariat, ISO/IEC JTC 1, American National Standards Institute, 25 West 43rd Street, New York, NY 10036; Telephone: 1 212 642 4932; Facsimile: 1 212 840 2298; Email: lrajchel@ansi.org

ISO/IEC JTC1 N---------

2003-10-10

Document Type:	Other Document (Open)
Document Title:	Response to JTC1 Sophia Resolution #39: Development of a Solution for the Unambiguous Identification and Interworking of Codes Representing Countries, Languages, and Currencies (prepared on behalf of SC32/WG1)
Document Source:	Dr. Jake V. Th. Knoppers and Mr. David Clemis (Canada) Contact: Tel: +1-613-234-3244; E-mail: <mpereira@istar.ca>
Project Number:	
Document Status:	JTC1 Sophia Resolution #39 "SC32/SC36 Topics of Internationalization and Localization" was directed to SC32/WG1. It in turn, through a SC32/WG1 resolution, requested the Co-Project Editors for the ISO/IEC 15944-5 "…Jurisdictional Domains…" project to prepare a response on behalf of SC32/WG1. This is the response to JTC1 Sophia Resolution #39 prepared by the Co-Project Editors for ISO/IEC 15944-5 as requested by SC32/WG1.. This document also serves as input to the follow-up work on JTC1 Sophia Resolution #22 "Cooperation and Coordination of Cross Cultural and Linguistic Matters"
Action ID:	For review and action by JTC1
Date:	2003-10-10
Distribution:	JTC1 P, O, and L members and JTC1 SCs
Medium:	
Disk Serial No.:	
No. of Pages:	41
Language Used:	English (eng)

Table of Contents

Table of Contents (Cont'd)

FOREWORD[1]

This contribution on behalf of SC32/WG1 responds to JTC1 Sophia Resolution #39 (and also in part to Resolution #22). {See JTC1 N6927}

This contribution identifies and summarizes the issues involved and provides a recommended solution for the unambiguous identification and systematic interworking of codes representing "countries", "languages", and "currencies"[2]. As such, it addresses not only the issues pertaining to "locale identifier" and "language identifier" {See JTC1 N6866}, but also current needs of e-business and jurisdictional domains.

This contribution and proposed solution focuses on the "semantic" aspects of the issues identified and not the delimiter aspects which are yet to be resolved.

This contribution recommends that JTC1 distribute this document to its SCs and others, {see Section 2.1 below} for comment and feedback so that the results can be integrated into "e-business" standards development work of JTC1/SC32/WG1 "Open-edi" and "metadata" standards development work of JTC1/SC32/WG2 "Metadata".

This contribution also requests JTC1 to take action focused on the development of a common pattern and use of delimiters (and one which is capable of supporting e-business requirements).

A number of short annexes are included in this document providing useful background information and examples of current problems from an IT-enablement perspective to which has been added the need for unambiguousness of data values used in "commitment" exchange of e-business transactions.

---

[1]The advice and assistance of M. Janice Pereira in the preparation of this document is appreciated and acknowledged.

[2]With respect to "country codes", "language codes" and "currency codes", it is recognized that:

< "country codes", i.e., ISO 3166-1, ISO 3166-2, ISO 3166-3, etc., that this multipart standard is the responsibility of ISO TC 46/WG2;

< "language codes", i.e., ISO 639-1, ISO 639-2, etc., that this multipart standard is the responsibility of ISO TC37/SC2; and,

< "currency codes", i.e., ISO 4217, that this is the responsibility of ISO TC68.

Note 1: The solution proposed for the unambiguous interworking of combinations of codes representing countries (and their administrative subdivisions), languages and/or currencies (in the context of a global economy, and jurisdictional requirements) does not require changes/modifications to existing code sets for these three standards.

Note 2: The three strategic directions of JTC1 for its standards development work are "portability", "interoperability", and "cultural adaptability". The proposed solution not only supports these three strategic directions of JTC1 for "information exchange" purposes but also addresses the requirements of "commitment" exchange of e-business.

Note 3: In a 30 September, 2003 Press Release (Ref: 871), the ISO reaffirmed its free-of-charge policy for use of its country, currency, and language codes. {See <http://www.iso.ch/iso/en/comcentre/pressreleases/2003/Ref871.htm>}

1.0 SHORT CHRONOLOGY OF EVENTS, DOCUMENTS AND RESOLUTIONS LEADING TO THIS CONTRIBUTION IN RESPONSE TO JTC1 SOPHIA RESOLUTION #39

Note: Those familiar with the documents and events resulting in JTC1 Sophia Resolution #39 may wish to skip this chronological overview. The purpose of this section is to provide a short chronology of events, documentation and resolutions leading up to the preparation of this contribution on behalf of JTC1/SC32/WG1. {See further Annex A}

The issues raised by JTC1/SC36 in its inquiry regarding cross-SC topics in internationalization and localization represent but the "tip of the iceberg" of other related issues which have arisen in the context of e-business requirements {See further Annex B}. A thorough systematic and pragmatic approach is needed to address the issues identified.

The inquiry by SC36 serves as the trigger for starting the process for resolving them.

1.1 In its work on e-business standards developments, i.e., via the multipart ISO/IEC 15944 standard, JTC1/SC32/WG1 identified the need for a common default convention for referencing "language codes" since ISO 639 now has three different "ISO" code sets. SC32/WG1 document N0178r, later SC32 document 32N0672, dated 2001-09-31, served as a liaison request to ISO TC37/SC2 the committee responsible for ISO 639.

The issues identified in SC 32 N0672 were later solved by both SC32/WG1 "Open-edi"[3] and SC32/WG2 "Metadata"[4] resolving to use the ISO 639-2/T(erminology) 3-alpha code set as the normative reference in their standards development work. {For a summary of the issues involved, see Annex C}.

1.2 SC32/WG1 in its standards development work pertaining to "jurisdictional domains" as part of new ISO/IEC 15944-5 continues to identify and address issues pertaining to the use of combinations of codes representing countries, languages and currencies from both e-business and metadata perspectives. {See further Annex A for some of the relevant SC32 documents on these topics}.

1.3 In its work on the development of e-learning standards, JTC1/SC36 through its liaison to SC22 prepared a paper (22 August, 2002) titled "Enquiry to SC22 Regarding Recommendations for Locale Specification for Internationalization and Localization" as JTC1/SC36 N0314 and also JTC1/SC22 N3491 (9 September, 2002).

1.4 In addition, the SC36 Chairman submitted a document to JTC1 titled "SC36 Inquiry Regarding Cross-SC Topics in Internationalization and Localization" {See document JTC1 N6866}.

1.5 The JTC1 N6866 document and related matters were discussed at the JTC1 Sophia-Antipolis, October, 20002 Plenary Meeting and resulted in Resolution #39 (found in JTC1 N6927), which reads as follows, and we quote:

---

[3] For SC32/WG1 in its 19544 multipart "e-business" standard on business agreement semantic descriptive techniques.

[4] For SC32/WG2 in its 11179 multipart "metadata registries" standard, i.e., the generic "metadata" standard.

Resolution 39 - SC32/SC36 Topics in Internationalization and Localization

JTC1 notes the request from SC36 as contained in document JTC1 N6866 concerning locale identifiers. JTC1 understands that SC32/WG1 is willing to assume this task subject to availability of resources.

JTC1 notes that ISO/IEC 15897 should be considered during this effort.

Unanimous.

This document responds to this JTC1 Resolution.

1.6 The reference in this JTC1 resolution to ISO/IEC 15897:1999 is that titled "Information technology - Procedures for the Registration of Cultural Elements". This standard is the responsibility of JTC1/SC22/WG20. It is currently undergoing revision as 2nd CD version "CD2 15897:2002 (E)" which is a SC22/WG20 N0978 document dated 2002-11-08.

1.7 JTC1 referred Resolution #39 to JTC1/SC32/WG1 "Open-edi". As the former SC30, it developed the generic e-business framework standard in the form of ISO/IEC 14662 "Information technology - Open-edi Reference Model"[5].

Based on this Framework Model, SC32/WG1 is developing the multipart ISO/IEC 15944 standard. This multipart standard focuses on the development of business agreement descriptive techniques for modelling business transactions through scenarios and scenario components, i.e, as identifiable, referenceable and reusable "business objects". This work includes incorporating the ability to support legal and jurisdictional requirements as well as those of a commercial, public policy, cross-sectorial and cultural adaptability nature[6].

As such, in its development of generic e-business standards, SC32/WG1 is already working on resolving the issues identified in the SC36 inquiry to JTC1 (as stated in JTC1 N6866), In addition, SC32/WG1 is placing these issues in the wider context of e-commerce requirements and those arising from jurisdictional domains (as already identified earlier by JTC1 Business Team on Electronic Commerce"). {See further Annex B}.

[5] ISO/IEC 14662 is a freely available standard at www.jtc1.org. It is undergoing its five-year revision and an FDIS for the 2003 version is to be issued before the end of the year.

[6] See further ISO/IEC 15944-1:2002 *Information Technology - Business Agreement Semantic Descriptive Techniques - Part 1: Operational Aspects of Open-edi for Implementation*. This standard is also available for free via <www.jtc1.org>.

Consequently, in response to the JTC1 Sophia Resolution #39, SC32/WG1 passed the following resolution at its January, 2003 meeting and we quote:

> ***Resolution WG1/27: SC32/WG1 Response to JTC1 Sophia Resolution 39 "SC36/SC32 Topics in Internationalization and Localization" [Ref. JTC1 N6927 & SC32 N0901]***
>
> *In response to this resolution, SC32/WG1 notes its development work on ISO/IEC 15944-5 Information Technology - Business Agreement Semantic Descriptive Techniques - Part 5: Identification and Mapping of Categories of Jurisdictional Domains". [formerly ISO/IEC 18038]. As follow-up to this resolution, SC32/WG1 requests the Project Editors to take the draft Annex titled "Unambiguous Semantic Components and Jurisdictional Domains: Standard Default Convention for Identification, Interworking, and Referencing of Combinations of Codes Representing (Official) Countries, Languages and Currencies" and prepare this Annex as a separate document for wider distribution.*[7]

This document is based on the standards development work of JTC1/SC32/WG1 "Open-edi" on the multipart ISO/IEC 15944 "e-business" standard and specifically Part 5 which focuses on "jurisdictional domains" including these being a primary source of predefined sets of semantic components of a "Codes representing X.." nature , i.e. as "coded domains".

[7] The follow-up work by the Co-Project Editors for ISO/IEC 15944-5 resulted in the decision that simply presenting the mentioned draft Annex of ISO/IEC 15955-5 was not deemed appropriate or that useful. Instead this separate document has been prepared in response to JTC1 Sophia Resolution #39. It contains additional background information as well as Annexes which are not part of the draft Annex referenced. The Co-Project Editors for ISO/IEC 15944-5 are Dr. Jake V.Th. Knoppers and Mr. David Clemis (Canada). They can be contacted at tel: +1-613-234-3244 or e-mail <mpereira@istar.ca>.

2.0 REQUESTED ACTIONS OF JTC1

Two actions are requested of JTC1; namely:

< one pertaining to a proposed solution for the unambiguous identification and interworking of codes representing "countries", "languages", and "currencies" in terms of their ordering, identification, referencing, and use of specific code sets of those standards (in order to support semantic interoperability requirements of e-business); and,

< one pertaining to the need to develop a "common" set of delimiters for locale identifiers and language identifiers.

2.1 JTC1 is requested to distribute this document to its SCs and Liaison members as well as ISO TC37/SC2, ISO TC46/WG2, ISO TC68, and IETF for comment on the proposed solution, i.e., in the form of the two default conventions, for the unambiguous identification and interworking of codes representing countries (their sub-divisions), languages and currencies (in an e-business context and one which supports the requirements of "commitment" exchange).

Comments on the proposed solution should be directed to the Co-Project Editors for ISO/IEC 15994-5, Dr. Jake V. Knoppers and Mr. David Clemis at < mpereira@istar.com >

2.2 JTC1 is requested to instruct JTC1/SC22 in collaboration with the IETF to develop a "common" approach for the specification for the delimiters for locale identifiers, and one which also supports the unambiguous identification and interworking of codes representing countries, languages, and currencies (including support of e-business requirements).

[Note: An alternative here would be for JTC1 to direct this second request which focuses on a common delimiter type specification, to JTC1/SC32 {See document 32N0918}].

It is fully recognized and appreciated that multiple applications are currently in use based on the initial conventions for use of country code and language code based on ISO/IEC 9945 (POSIX) and IETF RFC 3066 (formerly RFC 1766). These were developed well over a decade ago and are increasingly inadequate in a world-wide context, (e.g., like "Y2K" but on a lesser scale).

However, the development of the 3-alpha code sets of ISO 639-2 was done in response to real world requirements (and inadequacies of ISO 639-1 which currently contains only 42% of the languages for which codes are now available via ISO 639-2).

The global requirements of e-business and commitment exchange demand a solution for the unambiguous identification and interworking of codes representing countries, languages and currencies. The solution proposed is systematic, pragmatic and does not require any changes to the existing code sets of ISO 639, 3166 and 4217.

Should the proposed solution be accepted, a migration strategy and timetable can be developed.

3.0 SUMMARY OF ISSUES

3.1 ISSUES IDENTIFIED BY JTC1/SC36

The text which follows is a paraphrase copy of that found in document SC36N0314 which was submitted as JTC1 N6866.

> *There is a need in the field of work of several JTC1 SCs, (e.g., SC32 "Data Management and Interchange", SC36 "Learning, Education and Training", etc.)., to be able to specify locale information for use in internationalization ("i18n") contexts. Further, a* ***locale***[8] *may be associated with or attached to some "content" such as the XML "**lang**" attribute. Furthermore, within certain IT environments, it is necessary to maintain multiple localized variants, (e.g., a multilingual /multicultural message list - for each kind of message, there is an array of* ***{string, locale identifier}*** *pairs, one for each locale).*
>
> *JTC1/SC32 N0918 identifies two relevant specifications which contain "patterns" for locale identifiers; namely:*
>
> - *ISO/IEC 9945-1 (POSIX, Part 1) specifies a locale identifier, using the pattern* ***{language code} + underscore character + {country code}.*** *UNIX and Java systems use this format.*
> - *IETF RFC 3066 (a revision of RFC 1766) specifies a language identifier, using the pattern* ***{language code} + hyphen character + {country code}****.*
>
> *[Note: A good number of developers of RFCs 1766 and 3066 also participate in SC22/WG20 and also participate in the SC22/WG15 (POSIX) development].*
>
> *There appears to be a variety of features one would want to incorporate into such a locale identifier, including sub-codes, (e.g., "en-US-philadelphia", or "US-NY", and personal or group locale-specific preferences.*

Imbedded in this (summary) statement of issues identified by JTC1/SC36 are several issues which need to be differentiated, identified and addressed/resolved on their own, i.e., as stand-alone lego-blocks. At the same time, the manner in which these issues pertaining to the use of country codes, language codes, and currency codes are resolved and the solutions provided need to interwork and interoperate in an integrated, pragmatic and semantically consistent manner.

In the sections which follow, these "imbedded" issues are identified and a pragmatic approach to their resolution is provided. Finally, such an IT-enabled approach must also address and integrate e-business requirements and that of "commitment" exchange.

---

[8] *ISO/IEC 15897:1999 defines "locale" as:*

> ***3.1***
> ***locale***
>
> *the definition of the subset of the environment of a user that depends on language and cultural convention (See 2.5 of the POSIX standard for the specification of the local file format).*

3.2 NEED TO DIFFERENTIATE BETWEEN (1) THE ISSUE OF A "COMMON" SPECIFICATION FOR DELIMITERS; AND, (2) UNAMBIGUOUS IDENTIFICATION OF COMBINATIONS OF COUNTRY CODES AND LANGUAGE CODES (AND CURRENCY CODES)

The **first** set of issues pertains to the need for ISO/IEC JTC1 and the IETF (as well as W3C, etc.), to establish a "common" set of delimiters. At present ISO/IEC POSIX and IETF 3066 both use "**braces**", i.e., "**{}**" as in **{country code}** and **{language code}**, as per SC36 provided examples.

However, POSIX (and UNIX and JAVA) use the **underscore character**, i.e., "**_**", as the delimiter while IETF RFC 3066 uses the "**hyphen character**", i.e., "**-**", in its pattern.

There is a need to establish a single "common" specification for delimiters here. This issue is outside the scope of SC32/WG1 and is one for ISO/IEC JTC1 and the IETF to resolve. Here the US National Body made a proposal for a new work item (NWI) to JTC1/SC32 on "Locale Identifiers" {See SC32 N0918}. No action has yet been taken on this matter by SC32.

This issue of the need for a "common" specification for delimiters for "locale identifiers" (POSIX) and "language identifiers" (IETF) remains to be resolved.

There is a side issue here of the need in the development of such a common pattern for delimiters to be able to support identification of ISO 3166-2 "administrative subdivisions" within a country code. In summary, it is often the local jurisdictional domain, (e.g., province, state, territory, länder, canton, etc.), within a nation-state, which specifies (official) language requirements. Thus, the common pattern "locale identifier" or "language identifier" for jurisdictional domains must be able to support combinations of the ISO 3166-1 and ISO 3166-2 code sets. {See further below Section 3.7}

The **second** set of issues pertains to the need for ISO/IEC JTC1 and the IETF (as well as W3C, etc.), to establish a "common" pattern for combinations of these three code sets not only from an IT functional services perspective, but also that of the business operational view perspective. Here, in addition to resolving the issue whether to use an **underscore** or **hyphen** character as a delimiter, it is equally important to be able to support combinations of ISO 3166-1 and ISO 3166-2 codes representing entities as jurisdictional domains.

Given the emergence of the Internet and the need to support e-commerce requirements, the order of the presentation of "country code" and "language code" should be that of providing that for the "country code", i.e., jurisdictional domain first and then that of the "language code". {See further below Section 3.6, 3.7 and 4.0}.

Thus the sequence or order of "language code" plus "country code" needs to be reversed to "country code" plus "language code".

The related set of issues pertains to the unambiguous identification and referencing of combinations of "country codes" and "language codes". The current approach here of POSIX with respect to "locale identifiers," and IETF RFC 3066, with respect to "language identifiers" is inadequate and ambiguous in the context of present global requirements especially those of an e-business nature involving exchanges of "commitments", (e.g., as in e-commerce, e-government, e-medicine, e-logistics, e-learning, etc.).

3.3 ADDED REQUIREMENTS OF e-BUSINESS AND "COMMITMENT" EXCHANGE

A key and distinguishing characteristic of any business transaction is that it involves "commitment" exchange among Persons in addition to information exchange among their IT systems[9].

To date, the primary focus of IT-related standards for functional services, (e.g., in the areas of information technology, telecommunication services, security services, programming languages, etc.), has been on information exchange and interoperability among the technical components of the IT systems of the participating parties. The POSIX and IETF common patterns reflect such an approach. The use of the Internet in support of business transactions adds the new requirement of the ability to be able to support "commitment" exchange.

ISO/IEC 15944-1:2002 defines "commitment" as:

> ***commitment:*** *the making or accepting of a right, obligation, liability or responsibility by a Person that is capable of enforcement in the jurisdiction in which the commitment is made.* [ISO/IEC 15944-1:2002 (3.9)]

A key requirement in commitment exchange is that of unambiguity in the semantics, i.e., meaning, of the contents data values on the electronic data interchanges among the parties involved in a business transaction. **Unambiguity in the semantics of the contents is vital since the actual data values interchanged drive (automated) decision-making processes of the parties involved as "Persons".** The definition of "Person" is:

> ***Person:*** *an entity, i.e., a natural or legal person, recognized by law as having legal rights and duties, able to make* ***commitment****(s), assume and fulfil resulting obligation(s), and able of being held accountable for its action(s).*
>
> *NOTE 1 Synonyms for "legal person" include "artificial person", "body corporate", etc., depending on the terminology used in competent jurisdictions.*
>
> *NOTE 2 Person is capitalized to indicate that it is being utilized as formally defined in the standards and to differentiate it from its day-to-day use.*
>
> *NOTE 3 Minimum and common external constraints applicable to a business transaction often require one to differentiate among three common subtypes of Person, namely "individual", "organization", and "public administration".* [ISO/IEC 15944-1:2002 (3.47)]

Here the ISO/IEC 15944 multipart standard provides the methodology, concepts and definitions, rules, templates, models of formal description techniques, etc., for ensuring the required level of certainty through its focus on "business semantic descriptive techniques". It defines "unambiguous" as:

> ***unambiguous:*** *the level of certainty and explicitness required in the completeness of the semantics of the recorded information interchanged appropriate to the goal of a business transaction.* [ISO/IEC 15944-1:2002 (3.66)][10]

[9] See further Clause 6.1.3 "Business Transaction: Commitment exchange added to information exchange", in ISO/IEC 15944-1:2002, pp. 16-20.

As such, from an e-business needs perspective, as well as those of jurisdictional domains, interoperability is no longer just a technical requirement from an IT systems perspective, i.e., "IT interoperability", but also now from a semantic perspective, i.e., "semantic interoperability," in support of commitment exchange. Here the focus is ensuring that level of unambiguity in the semantics of the recorded information interchanged is to be able to support commitment exchange and automated decision-making processes among the Persons involved.

The sections which follow, focus on identifying existing ambiguities in the use of the sets of codes representing countries, languages, and currencies in support of e-business and resulting semantic interoperability requirements.

3.4 NEED TO SWITCH TO USE OF THE 3-ALPHA LANGUAGE CODE, I.E, ISO 639-2/T

The original ISO 639 standard contained a 2-alpha code. The SC36 N6866 inquiry document re: POSIX and IETF contains examples of the use of this 2-alpha code set. Other examples of W3C, Dublin Core, etc., in referencing language codes provide examples based on the original ISO 639:1988 2-alpha code set standard. This 2-alpha code set proved insufficient in meeting world-wide requirements[11].

In the late 1990s, ISO TC37/SC2, addressed this problem by making ISO 639 a multipart standard, with ISO 639-1 containing the 2-alpha code set for representation of names of languages and ISO 639-2 containing two 3-alpha code sets as ISO 639-2/T and ISO 639-2/B. {See further Annex C}.

ISO 639-2 has two 3-alpha code sets for the representation of names of languages, namely:

- one for terminology applications, i.e., ISO 639-2/T; and,
- one for bibliographic applications, i.e., ISO 639-2/B.

They are the same except for twenty-five languages that have a variant code. The problem here is that these include codes for major languages such as Chinese, French and German. {See further Annex B} One therefore needs a "default" convention for referencing codes representing "languages".

One result is that both ISO/IEC 9945 multipart POSIX standard in referencing language codes as part of locale identifier and elsewhere should reference the 3-alpha ISO 639-2/T code set instead of the 2-alpha

---

[10] In addition, ISO/IEC 15944-1:2002 has two informative Annexes which focus on the need for unambiguity in the semantics of the contents of EDI in support of e-business transactions; namely:

- Annex C (Informative) Unambiguous identification of entities in (electronic) business transactions; and,
- Annex D (Informative) Existing standards for the unambiguous identification of persons in business transactions (organizations and individuals) and some common policy and implementation considerations.

[11] For example, neither Canada nor the USA adopted the original ISO 639:1988 standard as a "national standard" in their respective countries. A primary reason here was that this ISO standard did not meet the needs of both countries in that it did not include codes representing native/aboriginal languages in use in their jurisdictional domains. Other countries had similar requirements.

code set. The same holds true for W3C, Dublin Core, etc.,[12] a supporting migration strategy and timetable can be developed.

The issue of use of the ISO 639-1 2-alpha code sets versus ISO 639-2 3-alpha code sets is recognized in the IETF's 3066:2001. It stated the following under "clause 2.3 Choice of language tag[13]", and we quote:

> *"One may occasionally be faced with several possible tags for the same body of text. Interoperablity is best served if all users send the same tag, and use the same tag for the same language for all documents. If an application has requirements that make the rules here inapplicable, the application protocol specification MUST specify how the procedure varies from the one given here.*
>
> *The text below is based on the set of tags known to the tagging entity.*
>
> *1. Use the most precise tagging known to the sender that can be ascertained and is useful within the application context.*
>
> *2. When a language has both an ISO 639-1 character code and an ISO 639-2 3-character code, you MUST use the tag derived from the ISO 639-1 2-character code.*
>
> *3. When a language has no ISO 639-1 2-character code, and the ISO 639-2/T (Terminology) code and the ISO 639-2/B (Bibliographic) code differ, you MUST use the Terminology code. NOTE: At present, all languages for which there is a difference have 2-character codes, and the displeasure of developers about the existence of 2 code sets has been adequately communicated to ISO. So this situation will hopefully not arise.*

The information provided on the ISO's identified website where the language codes are publicly available[14] currently has 195 entries in ISO 639-1 for the 2-alpha code set, and 466 entries in the ISO 639-2 3-alpha code set.

Thus, **ISO 639-2 already contains 271 3-alpha codes for which there are no entries in ISO 639-1.** As such, the ISO 639-1 2-alpha codes currently represent only 42% of the available set of "codes for the representation of Names of Languages".

The number of ISO 639-2 3-alpha codes continues to grow. A major contributing factor here is the introduction and increasingly widespread use of ISO/IEC 10646-1:2000 *"Information Technology -- Universal Multiple-Octet Coded Character Set (USC) -- Part 1: Architecture and Basis Multilingual Plane (a.k.a "Unicode") and its amendments up to Amendment 11:1997 Unified Canadian Aboriginal Syllabics, and the Unicode Standard Version 3.0*.

---

[12] For a concordance table for the 3-alpha language codes equivalents from the 2-alpha codes, visit <http:www/loc.gov/standards/ISO639-2/langcodes.html>. It also identifies those languages for which a 3-alpha code exists but not a 2-alpha code.

[13] See <http://www.faqs.org/rfcs/rfc3066.html>.

[14] See <http://www.loc.gov/standards/iso639-2/langcodes.html>.

This IT standard represents but one of many ways in which IT technology and standards development responds to the world-wide requirements of multicultural and pluralistic societies, the increased recognition and accepted need for IT to serve and adapt to the needs of human beings.

Further, from an e-commerce perspective, being able to deliver goods and services is a world-wide challenge. Tailoring one's product to the user environment of the local market including jurisdictional levels, demands attention to detail, adaptability and flexibility. Being successful in a market requires not only understanding the needs of clients but also being able to communicate with tem in their language. Thus, multilingualism and ability to meet local requirements is good for business and "good business".

The default chosen by JTC1/SC32/WG1 "Open-edi" (or e-business) and JTC1/SC32/WG2 "Metadata" in their standards development work is "ISO 639-2/T, i.e., the 3-alpha code for terminology applications[15].

It is <u>recommended</u> that ISO/IEC JTC1 as a whole and its SCs also adopt the "ISO 639-2/T" 3-alpha code as its <u>default convention</u> for referencing codes representing languages.

The original 2-alpha ISO 639 code set reflects a "Eurocentred" linguistic approach. The ISO 639-2/T alpha code set represents a wider global requirements perspective and one which is able to encompass/expand to meet world-wide requirements of users.

3.5 VARIANCES IN REPRESENTATIONS OF COMBINATIONS OF LANGUAGE CODES, COUNTRY CODES, CURRENCY CODES, ETC., AND CURRENT NEEDS

The approach taken in the examples by ISO/IEC 9945-1 and IETF RFC 3066 presents the language code as a "2-alpha lower case" and then "country code" as a "2-alpha UPPER CASE". Other Internet oriented applications including Dublin Core also use combinations of 2-alpha codes of ISO 639 for "languages" and ISO 3166-1 for "countries".

However, these sets of 2-alpha code sets are not, nor ever intended to be, unique and mutually exclusive. The same 2-alpha code can be and is used to designate both an ISO 639 "language" and an ISO 3166-1 "country". {See further Annex D}.

Further, this approach, via the Internet, while sufficient for sharing information among human beings, who may be able to filter out semantic inconsistencies and ambiguities is <u>neither</u>:

(1) conductive to ensuring global portability, interoperability, and/or cultural adaptability of metadata and data through the use of information communication technologies (ICT); nor,

---

[15] For example, in the new ISO/IEC 11179-3 standard *"Information technology - Metadata Registries (MDR) - Part 3: Registry Metamodel and basic attributes,"* the generic "metamodel" standard has defined "language identifer" as follows in Clause 3.3.79 as:

> ***language identifier:*** *information in a* ***Terminological Entry*** *which indicates the* ***name*** *of a* ***language****.*
>
> *NOTE 1 Use the three character alphabetic codes and names from ISO 639-2/Terminology, with extensions if required.*
>
> *NOTE 2 Metamodel constructs is: Attribute of Language Identification.*
> [ISO/IEC 11179-3:2003(E) 3.3.79].

(2) useful with respect to metadata and (content) data as semantic components in support of e-business, i.e., any application involving "commitment" exchange, (e.g., as any activity involving a "fee" (whether on a for-profit or not-for-profit basis), exchange or recognition of rights, obligations, liabilities, accountabilities, etc.

In addition, the various conventions of current W3C, IETF, Dublin Core, etc., used for the identification and referencing of combinations of "languages and countries" and/or "countries and languages" using combinations of the 2-alpha code sets of ISO 639-2 and ISO 3166-1 is both semantically ambiguous and inadequate for international requirements generally and particularly in support of e-business requirements.

There are also the currency codes of ISO 4127:2001 which are upper case 3-alpha codes. They are not mutually exhaustive of the 3-alpha language codes and/or the 3-alpha country codes.

Together the code sets of ISO 639, ISO 3166, and ISO 4217, are the most widely used globally especially in applications of an e-business transaction nature. Of these, both ISO 639 and ISO 3166 have more than one applicable code set (and more than one representation) that can be utilized. ISO 4217 has one code set and one form of representation.

In addition, various sectors use different code sets, combinations and/or representations of these code sets, (e.g., lower case or UPPER CASE). Until recently, each industry sector used its own convention and had very little need for interworking and interoperability requirements with other sectors with respect to the code sets of these three standards let alone doing so in an IT-enabled manner.

A standards-based and systematic approach is needed for the unambiguous interworking of the code sets of these three standards. Further, the solution found must be one which also provides semantic completeness and unambiguousness.

3.6 MULTIPLE SETS OF "SIMILAR" CODES REPRESENTING "COUNTRY", "LANGUAGE", AND "CURRENCY"

ISO/IEC 9945-1 (POSIX) and IETF RFC 3066 contain examples of combinations of "language code" and "country code" as well as "administrative subdivisions" within a country, i.e., states, provinces, etc. {See further 2.1 above}. These examples are no longer representative of current user requirements in a global context let alone from an e-business needs perspective.

3.6.1 ISO 3166-1 "COUNTRY CODES"[16]

At present, three (3) "equivalent" different code sets are in use for the identification of names of "countries" and other geopolitical entities, etc.; namely:

- a 3-digit numeric code. [Source = United Nations]
- a 3-alpha code. [Source = United Nations]
- a 2-alpha code. [Source = ISO 3166-1].

Here the alpha codes are often represented in both lower and UPPER case although ISO 3166-1 specifies UPPER CASE[17].
Of these three code sets, the 3-digit numeric code is the most stable. The 2-alpha and 3-alpha codes of ISO 3166-1 can and do change at the request of the country concerned either because a country decided to change its name designation(s) or prefers to use another alpha code designation.[18] Further, the 3-digit

[16] It is recognized that ISO 3166-1 contains codes for entities other than "countries", i.e., "UN member states". The complete title of ISO 3166-1 reflects this fact, the complete title being *Codes for the representation of names of countries and their subdivisions - Part 1: Country codes/Codes pour la représentations des noms de pays et de leur subdivisions - Partie 1: Codes pays*.

It should also be noted as ISO 3166-1 itself states in its Introduction, and we quote:

> *"The three parts of ISO 3166 do not express any opinion whatsoever concerning the legal status of any country, dependency or other area named herein, or concerning its frontiers or boundaries".*

Nearly 20% of the entities listed in ISO 3166-1 are not "countries", i.e, UN member states. {See further JTC1/SC32 document N1063 titled "(2nd Draft) *"Annex C (Normative) Codes Representing UN Member States and Their Official (or de facto) Languages"* forming part of draft CD Document (version 2.0) of *ISO/IEC 15944-5 Information technology - Business agreement semantic descriptive techniques - Part 5: Identification and mapping of various categories of jurisdictional domains as external constraints""*.

[17] ISO 3166-1:1997 (E/F) states in Clause 5.2 "Construction of the alpha-2 code", and we quote:

> *"The alpha-2 code uses combinations in upper case of two letters of the 26-character Roman alphabet (ignoring diacritic signs) from the range AA to ZZ".*

The most widespread use of the ISO 3166-1 2-alpha code in lower case is as part of the top level domain in Internet (ICANN) domain names.

Clause 5.3 "Construction of the alpha 3-Code" states, and again we quote:

> *"This part of ISO 3166 also provides an alphabetic 3-character (alpha 3) code, based on the alpha-2 code, and using combinations, in upper case, of three letters of the 26-character Roman alphabet (ignoring diacritic signs) from the range AAA to ZZZ, for use in cases where a specific need has been identified.*
>
> *NOTE 3 - Attention is drawn to the fact that other 3-letter codes exist".*

[18] For UN member states, changes in names must be approved by the Security Council to be recognized. {See further document JTC1/SC32 N0535 "Approach to Development of the new ISO/IEC 18038 "Identification and Mapping of Various Categories of Jurisdictional Domains". [Note: "18038" is now "15944-5"]. Also, N0535 contains an Annex B titled "Identification and Mapping of "Countries" as Jurisdictions on a Peer-to-Peer Basis"].

numeric code has other advantages such as providing language independence[19]. For example, multiple equivalent valid (official) name representations, i.e., human interface equivalents. {see further Annex E below for some examples}.

Neither ISO 639 nor ISO 4217 have a "numeric" code set. In addition, the financial services sector already uses the 3-digit numeric code for countries in financial transactions.[20]

ISO 3166-1 states, in Clause 5.5 "Specification for use" that:

> *"When applying this part of ISO 3166, users should clearly state which of the three codes they are using. If a code element from this part of ISO 3166 is used in combination with other characters for special purposes, it is strongly recommended that the choice and function of any such additional character be specified".*

Consequently, it is <u>recommended</u> that ISO/IEC JTC1 and its SCs adopt the ISO 3166-1, 3-digit numeric as its default convention for referencing codes representing countries.

3.6.2 ISO 639-2 "LANGUAGE CODES"

With respect to "language codes," the 2-alpha code set is no longer adequate to meet global requirements. Consequently, in 1998, ISO TC37/SC2, the ISO committee responsible, introduced ISO 639-2:1998 *Codes for the representations of names of languages - Part 2: Alpha-3 code/Codes pour la représentation des noms de langue - Partie 2: Code alpha-3*.

ISO 639-2 has two 3-alpha code sets for the representation of names of languages, namely:

- one for terminology applications, i.e., ISO 639-2/T; and,
- one for bibliographic applications, i.e., ISO 639-2/B.

They are the same except for twenty-five languages that have a variant code. The problem here is that these include codes for major languages such as Chinese, French and German. {See further Annex C} As a result ISO 639 has three sets of codes for the representation of names of languages; namely:

- a 2-alpha code [Source = ISO 639-1]]
- a 3-alpha code - bibliographic [Source = ISO 639-2/B]
- a 3-alpha - terminology [Source = ISO 639-2/T]

Codes representing names of languages are presented in lower case[21] (although conventions for some user applications use UPPER case).

---

[19] Here ISO 3166-1 states in Clause 5.4 "Construction of the numeric-3 code", and we quote:

> *"Recognizing that a numeric code for country names is of advantage (e.g. to provide language independence), a three-digit numeric (numeric-3) code from the range 000 to 899 is also provided in this part of ISO 3166. It is made available by the United Nations Statistics Division".*

[20] For this and other reasons in financial transactions, the 3-digit ISO 3166-1 code set is used. See further ISO 8583-1:2003 "Financial transaction card originated messages - Interchange message specifications – Part 1: Messages, data elements and code values".

One therefore needs a "default" convention for referencing codes representing "languages".

The default chosen by JTC1/SC32/WG1 "Open-edi" (or e-business) and JTC1/SC32/WG2 "Metadata" in their standards development work is "ISO 639-2/T, i.e., the 3-alpha code for terminology applications.

It is <u>recommended</u> that ISO/IEC JTC1 and its SCs also adopt the "ISO 639-2/T" 3-alpha code lower case as its default convention for referencing codes representing languages.

3.6.3 ISO 4217 "CURRENCY CODES"

ISO 4217 has two sets of codes for the representation for currencies and funds; namely:

- 3-alpha; and,
- 3-numeric.

The 3-alpha codes are represented in UPPER case only in ISO 4217:2001. The numeric currency code is derived, where possible, from the United Nations standard Country or Area Code. Additional codes to meet special requirements are allocated as necessary from within the user-assigned range of codes 950 to 998. Thus for many countries their 3-digit currency code is not the same. The introduction of the "euro" (EUR) has greatly increased this number[22].

Consequently, the 3-alpha code set UPPER case is the most widely used and avoids confusion with country codes.

---

[21] With respect to the 2-alpha code, ISO 639-1:2001 states in Clause 4.4 "Form of the language identifiers", and we quote:

> *"The language identifiers consist of the following 26 letters of the Latin alphabet in lower case: a, b, c, d, e, f, g, h,i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x y, z. No diacritical marks or modified characters are used. Implementors should be aware that these identifiers are not intended to be an abbreviation for the language, but to serve as a device to identify a given language. The language identifiers are derived from the language name. Each identifier is based on the indigenous name of the language or the preference of the communities using the language".*

And with respect to the 3-alpha codes, ISO 639-2:1998 states in Clause 4.1 "Form of the language codes", and we quote:

> *"The language codes consist of three Latin-alphabet characters in lowercase [sic]. No diacritical marks or modified characters are used. Implementors should be aware that these codes are not intended to be an abbreviation for the language, but to serve as a device to identify a given language or group of languages. The language codes are derived from the language name."*

[22] See further Table A.1 Currency and Funds Code lit in ISO 4217:2001 (E/F).

3.7 NEED TO BE ABLE TO SPECIFY "LOCALES" AND JURISDICTIONAL DOMAINS AT THE ISO 3166-2 LEVEL

Note: Part 2 of ISO 3166 focuses on "Country subdivision codes". On the whole these are the administrative subdivisions of those countries listed in ISO 3166-1 which are nation-states, i..e, member states of the United Nations[23]. The alpha code sets used in ISO 3166-1 and ISO 3166-2 are not mutually exclusive. For example, "NL" in ISO 3166-1 represents the "Netherlands" while "NL" in the ISO 3166-2 country subdivision code set for Canada represents the province of "Newfoundland and Labrador".

It is up to each nation-state to decide on its administrative subdivision(s), whether these are homogeneous or heterogeneous, and the codes to be assigned. Thus with ISO 3166-2 code assignment is unique only within the code set of the administrative subdivision for each UN member state.

However, the combinations of Part 1 Country Code and Part 2 Country subdivision codes do yield unique values, i.e., as composite identifiers.

The need to be able to specify "locales" and jurisdictional domains at the country subdivision level arises from the fact that:

(1) as a jurisdictional domain they may have linguistic requirements different from or in addition to those of the country of which they are part thus impacting the user environment;

(2) in the context of "localization" requirements, i.e., "locale" + "jurisdictional domain", it may well be that there are legal and public policy requirements which a user environment may be required to support.

In section 3.8 examples of requirements of this nature are presented.

3.8 UNAMBIGUOUS IDENTIFICATION AND INTERWORKING OF COMBINATIONS OF CODES REPRESENTING COUNTRIES, LANGUAGES, AND CURRENCIES

Many combinations of the code sets for names representing "countries", "languages", and "currencies" are possible and many combinations and representations are currently in use in various sectors and applications. To date most have not been required to be able to interwork and interoperate especially from semantic and IT-enabled perspectives.

The cross-sectorial nature of the "e-world" coupled with the requirements of e-business and jurisdictional domains have brought to the fore the need for a common horizontal approach which is systematic, unambiguous, pragmatic, reflects jurisdictional requirements and also ensures semantic interoperability from both the IT interface and human interface perspectives.

---

[23] *Note: The code sets for geopolitical entities in ISO/IEC 3166-1 and 3166-2 are not mutually exclusive. The same geopolitical entity can be a member of both Part 1 and Part 2 with the same name representation but with different codes, (e.g., St. Pierre and Michelon, Bermuda, etc.).*

Key factors here include:

(1) the same natural language is utilized in many countries but use of that same natural language may vary in each country in choice of words, terms, spelling, capitalization, hyphenation, etc.

These variations in use of natural languages are significant enough to require one to specify the country in which the language is used.

Canada - English (=124:eng)
United States - English (=840:eng)
United Kingdom - English (=826:eng)
Australia - English (=036:eng)

or

Canada - French (=124:fra)
France - French (=250:fra)
Senegal - French (=686:fra)

etc., for use of Spanish, Portuguese, German, etc., in different countries.

(2) a country, as a jurisdictional domain, may have more than one official language. For example, Canada at the nation-state level has two, i.e., English and French, (124:eng, and 1245:fra). South Africa has eleven (11) official languages.

(3) a jurisdictional domain as an ISO 3166-2 "administrative sub-division", (e.g., state, province, länder, canton, etc.), may have its own official language(s).

For example, the Territory of Nunavut in Canada has three official languages, English, French, and Inuktitut. The ISO 3166-2 code for Nunavut as an administrative sub-division of Canada is "NU". Thus the valid locale identifiers for combinations of country codes and language codes here are:

124-nu:eng
124-nu:fra
124-nu:iku.

(4) A country (or jurisdictional domain) may have more than one (official) currency.

A primary example here are those Member States of the European Union who adopted the single currency in accordance with the Treaty establishing the European Community, i.e., the euro (= EUR) plus the previous currency.

In addition, several countries do have more than one currency. Examples include East Timor which has the Timor Escudo and Rupia:

East Timor: Timor Escuda = 626:TPE
East Timor: Rupia = 626: IDR.

Haiti which has the Gourde and US Dollar, i.e.,

Haiti Gourde = 332:MTG
Haiti US Dollar = 332:USD

Panama which has the Balbon and the US Dollar, i.e.,

Panama Balboa = 590:PAB
Panama US Dollar = 590:USD.

The cross-sectorial nature of the "e-world" has not only created the need for a systematic common horizontal and IT-enabled approach to these three code sets. The requirements of e-business and "commitment" exchange also require a high level of unambiguity in the interworking of these three code sets.

A common default convention is required. In the section which follows a solution is proposed which meets these requirements. It has the added advantage of not requiring any changes to the existing standards referenced, i.e, ISO 639-2/T, ISO 3166-1 (and ISO 3166-2) and ISO 4217.

4.0 PROPOSED SOLUTION

In order to minimize ambiguity and maximize interoperability from IT and semantic perspectives as well as supporting commitment exchange requirements of e-business the following solution is proposed.

4.1 DEFAULT CONVENTION #1 FOR UNAMBIGUOUS IDENTIFICATION AND REFERENCING OF COMBINATIONS OF CODES REPRESENTING COUNTRIES, LANGUAGES, AND CURRENCIES

The proposed common default convention is to identify and reference:

- **countries via their ISO 3166-1 3-digit numeric code, i.e., as "nnn";**
- **languages via their ISO 639-2/T 3-alpha code using the lower case, i.e., as "aaa"; and,**
- **currencies via their ISO 4217 3-alpha code using upper case, i.e., as "AAA".**

This standard default convention for the unambiguous identification and references of codes representing the names of countries, languages, and currencies allows any combination of these three code sets to be specified in a semantically complete and interoperable manner.

It is recognized that associated with each of these codes are multiple equally valid (if not official) name representations. {see further Annex E below). The adoption of a common default convention as proposed here will not only provide a systematic, pragmatic, cost-effective and efficient approach in support of "cultural adaptability", but will also support the global requirements of e-business and commitment exchange.

4.2 DEFAULT CONVENTION #2 FOR THE ORDERING OF CODES REPRESENTING COUNTRIES, LANGUAGES AND CURRENCIES

In the context of being able to specify user requirements, i.e., as "locales" and the requirements of e-business and jurisdictional domains, use of a natural language is conditioned by the country or jurisdictional domain in which it is used.

Thus, the default convention where both country code and language code are used should be that of:

- if both country code and language code are used their order is:

  {country code} + {language code};

- where the geopolitical entity or jurisdictional domain of the user environment of an administrative subdivision of a country the applicable ISO 3166-2 code set is to be utilized in conjunction with the country code as:

  {country code-applicable ISO 3166-2 code set} + {language code}.

Default convention #2 as applied to combinations of country code and currency code are likewise to be ordered as:

{country code} + {currency code}.

4.3 APPLICATION OF DEFAULT CONVENTION #1 FOR IDENTIFYING CODES REPRESENTING COUNTRIES, LANGUAGES AND CURRENCIES WITH DEFAULT CONVENTION #2 FOR ORDERING THEM

The application of the proposed default conventions #1 and #2 provide for unique combinations which also support interoperability requirements from both IT and semantic perspectives as well as those of e-business and jurisdictional domains..

Thus a unique combination of:

(1) country and (official) language is identified and referenced as "**nnn:aaa**"[24];

(2) country and currency as "**nnn:AAA**".

Should one wish not to use the default ordering convention #2 then: language code + country code would be "aaa:nnn".

However, a common, horizontal generic default convention or standard for the unambiguous ordering and identification of codes representing countries, languages and currencies will not only:

(1) ensure semantic completeness and interoperability required in support of e-business and jurisdictional domains; but also,

(2) serve as a standard "pivot" from which one can map to legacy systems, local usage conventions, sectorial applications, etc., use of various syntaxes (including ASN.1, UN/EDIFACT, HTML, XML, etc.).

P.S. A variant on default convention #1, i.e., as convention #1a, would be the use of the 2-alpha code UPPER case for country codes instead of the 3-digit numeric. However, here one must take into consideration the fact that:

(1) the 2-alpha and 3-alpha country codes can and do change whenever a country changes its name and the UN Security Council agrees to such a change in name. However, the 3-digit numeric code remains stable and does not change; and,

(2) the 2-alpha and 3-alpha country codes are "reused" after a 10 year period of non-use, (e.g., the recent case of reuse of "CS" for Serbia and Montenegro where "CS" code was previously used to represent "Czechoslovakia"). The 3-digit numeric codes are not "re-used".

[24] In this document we use the colon "(:)" as a stakeholder delimiter awaiting resolution of the POSIX use of the underscore delimiter versus the IETF use of the hyphen delimiter issue.

ANNEXES

ANNEX A - REFERENCED AND/OR RELEVANT DOCUMENTS

A.1 EXPLANATORY NOTE

In Annex A we provide the complete title and reference for the documents cited in this contribution as well as those considered relevant to the issues identified.

All the documents cited are freely available via the ISO/IEC JTC1 website <http://www.jtc1.org>. Those pertaining to a specific sub-committee (SC) of JTC1 can also be accessed using this JTC1 website.

With respect to the ISO/IEC and ISO standards referenced, these can be obtained via the normal channels. One notes that the following standards have been made freely available by ISO/IEC; namely:

- ISO/IEC 11179-3:2003;
- ISO/IEC 14662:1997; and,
- ISO/IEC 15944-1:2002.

This Annex A is composed of the following six (6) sections; namely:

- ISO/IEC and ISO Standards;
- ISO/IEC JTC1 Documents;
- ISO/IEC JTC1/SC32 Documents;
- ISO/IEC JTC1/SC36 Documents;
- ISO/IEC JTC1/SC22/WG20 documents; and,
- Other Documents.

The standards and documents identified in this Annex A are meant to be "informative" and not "exhaustive".

A.2 ISO/IEC AND ISO STANDARDS

Note: The ISO/IEC and ISO standards are presented in ascending numerical order.

Ref. No./Date ISO/IEC, or ISO	Title
ISO 639-1:2001	Codes for the representation of names of languages - Part 1: Alpha-2 code/Codes pour la représentation de noms de langues - Partie 1: Code alpha-2
ISO 639-2:1998	Codes for the representations of names of languages - Part 2: Alpha-3 code/Codes pour la représentation des noms de langue - Partie 2: Code alpha-3
ISO 2788:1986	Documentation - Guidelines for the establishment and development of monolingual thesauri/Documentation - Principes directeurs pour l'établissement et le développement de thesaurus monolingues
ISO 3166-1:1997	Codes for the representation of names of countries and their subdivisions - Part 1: Country codes/Codes pour la représentations des noms de pays et de leur subdivisions - Partie 1: Codes pays
ISO 3166-2:1998	Codes for the representation of countries and their subdivisions - Part 2: Country subdivision code/Codes pour la représentation des noms de pays et de leurs subdivisions - Partie 2: Code pour les subdivisions de pays
ISO 4217:2001	Codes for the representation of currencies and funds/Codes pour la représentation des monnaies et types de fonds
ISO 5218:2003	Information technology - Codes for the representation of human sexes/Technologies de l'information - Codes pour la représentation des sexes humains (at FDIS stage)
ISO 5964:1985	Documentation - Guidelines for the establishment and development of multilingual thesauri/Documentation - Principes directeurs pour l'établissement et le développement de thesaurus multilingues
ISO 8583-1:2003	Financial transaction card originated messages - Interchange message specifications - Part 1: Messages, data elements and code values
ISO/IEC 9945-1:2003	Information technology - Portable Operating System Interface (POSIX) - Part 1: Base Definitions
ISO/IEC 9945-2:2003	Information technology - Portable Operating System Interface (POSIX) - Part 2: System Interfaces (available in English only)
ISO/IEC 10646-1:2000	Information Technology -- Universal Multiple-Octet Coded Character Set (USC) -- Part 1: Architecture and Basis Multilingual Plane (a.k.a "Unicode") and its amendments up to Amendment 11:1997 Unified Canadian Aboriginal Syllabics, and the Unicode Standard Version 3.0
ISO/IEC TR 11017:1998	Framework for internationalization
ISO/IEC 11179-3:2003	Information technology - Metadata Registries (MDR) - Part 3: Registry Metamodel and basic attributes
ISO/IEC 14651:2000	International String Ordering

Ref. No./Date ISO/IEC, or ISO	Title
ISO/IEC 14662:2003	Information technology - Open-edi Reference Model/Technologies de l'information - Modèle de référence EDI-ouvert
ISO 15836:2003	The Dublin Core Metadata Element Set
ISO/IEC 15897:1999	Information technology - Procedures for the registration of cultural elements. [Note: Currently under revision as 2nd CD. See further "CD2 15897:2002 (E)" which is a SC22/WG20 N0987 document dated 2002-11-08].
ISO/IEC 15944-1:2002	Information Technology - Business Agreement Semantic Descriptive Techniques - Part 1: Operational Aspects of Open-edi for Implementation
ISO/IEC 15944-2:200n	Information Technology - Business Agreement Semantic Descriptive Techniques - Part 2: Registration of Scenarios and their Components as Business Objects (at 2nd CD stage)
ISO/IEC 15944-5:200n	Information Technology - Business Agreement Semantic Descriptive Techniques - Part 5: Identification and Mapping of Various Categories of Jurisdictional Domains (formerly ISO/IEC 18038) (at 2nd WD/draft CD stage)

A.3 ISO/IEC JTC1 DOCUMENTS

Note: The ISO/IEC JTC1 documents are presented in descending numerical order. They can be accessed via the JTC1 website at <www/jtc1.org>.

ISO/IEC JTC1 Doc. No.	Title	JTC1/SCs X- Ref.
6927	Resolutions Adopted at the 17th Meeting of ISO/IEC JTC1, 21-25 October 2002 in Sophia Antipolis, France	
6866	SC 36 Inquiry Regarding Cross-SC Topics in Internationalization and Localization	36 N0314
5296	JTC1 Business Team on Electronic Commerce: Report to JTC1: Work on Electronic Commerce Standardization to be Initiated	

A.4 ISO/IEC JTC1/SC32 DOCUMENTS

Note 1: The ISO/IEC JTC1/SC32 documents are presented in descending numerical order.

Note 2: ISO/IEC JTC1/SC32 does have its own website and document register accessible via <http://www.jtc1.org>. SC32/WG1 has its own website and document register accessible via the JTC1/SC32 website.

ISO/IEC JTC1/SC32 Doc. No.	Title	SC32/WG1 Doc. No. X-Ref.
1063	(2nd Draft) *"Annex C (Normative) Codes Representing UN Member States and Their Official (or de facto) Languages"* forming part of draft CD Document (version 2.0) of *ISO/IEC 15944-5 Information technology - Business agreement semantic descriptive techniques - Part 5: Identification and mapping of various categories of jurisdictional domains as external constraints"*	
0978	Revised - JTC1/SC32 Plenary Resolutions 2003-01-31, Santa Fe, NM, USA	
0949	(Draft) Applying an ISO/IEC 18022 Model Approach to Dublin Core as an IT-enabled Coded Domain with Multilingual Equivalency: A Business Case Study - Canadian Federal Government	SC36/N0433
0930	What is a "Coded Domain"?	SC32/WG1/N252r
0918	US Contribution in reference JTC1/Sophia-Antipolis Resolution #39 on Internationalization and Localization	
0917	US Contribution in Reference to JTC1 Sophia-Antipolis Resolution #39 on Internationalization and Localization	
0901	Resolutions Adopted at the 17th Meeting of ISO/IEC JTC1, 21-25 October, 2002 in Sophia-Antipolis, France. (1N6927)	
0696	ISO/IEC WD 18038 Identification and Mapping of Various Categories of Jurisdictional Domains	
0695	ISO/IEC WD 18022:200x - IT-Enablement for Widely-Used Coded Domains	
0672	Need for a standard “default” convention for referencing ISO 639-2: “Codes for the representation of names of languages” in Open-edi business transactions and e-commerce, e-business, etc.	SC32/WG1 N178r
0535	Approach to Development of the New ISO/IEC 18038 Identification and Mapping of Various Categories of Jurisdictional Domains	
0462	Making Standards Work in Electronic Commerce and Among Jurisdictions: IT-Enablement of Data Element-based Standards - Presentation at the Open Forum on Metadata Registries in Santa Fe	
	Languages and Jurisdiction: "Natural", "Special", "**Official**", "Artificial", "Indexing", "Programming," etc.	SC32/WG1 N210R

A.5 ISO/IEC JTC1/SC36 DOCUMENTS

Note 1: The ISO/IEC JTC1/SC36 documents are presented in descending numeric order.

Note 2: ISO/IEC JTC1/SC36 does have its own website and document register at <http://www.jtc1sc36.org>.

ISO/IEC JTC1/SC36 Doc. No.	Title
0463	Languages and Jurisdiction: "Natural", "special", "de Jure", "National", "Artificial", "Indexing", "Programming", etc.
0421	Proposal for NWI on "Information Technology for Learning, Education and Training - Description of Language Capabilities".
0397	Progress Report on JTC1 Activities in Culture, Language, and Functional Accommodation
0371	SC36 Chair's Remarks on JTC1 Sophia Antipolis Resolution #39 Concerning SC32/SC36 Topics in Internationalization and Localization
0314	Inquiry Regarding Cross-SC Topics in Internationalization and Localization
0307	Maximizing Bilingual/Multilingual e-Learning Capability through a Bilingual/Multilingual Approach

A.6 ISO/IEC JTC1/SC22/WG20 DOCUMENTS

Note 1: The ISO/IEC JTC1/SC22/WG20 documents are presented here in descending numeric order.
Note 2: ISO/IEC JTC1/SC22/WG20 and its document register can be accessed via the JTC1/SC22 website.

ISO/IEC JTC1/SC22 /WG20 Doc. No.	Title
1049	Business Plan and Convenor's Report: ISO/IEC JTC1 SC22/WG20 - Internationalization
1003	ISO/IEC JTC1/SC22/WG20 Resolutions, Busan meeting 2003-02-11/13
0978	Enquiry to SC22 Regarding Recommendations for Locale Specification for Internationalization and Localization

A.7 OTHER DOCUMENTS

Source	Title
IETF RFC 3066:2001	Tags for the Identification of Languages (replaces RFC 1766)
IETF RFC 2234:1997	Augmented BNF for Syntax Specifications: ABNF
IETF RFC 1766:1995	Tags for the Identification of Languages (replaced by RFC 3066)
Computer Standards & Interface:1998	Knoppers, J.V.Th. "Global electronic commerce through localization and multilingualism". Computer Standards & Interfaces. 20 (1998):101-109.

ANNEX B - JTC1 BUSINESS TEAM ON ELECTRONIC COMMERCE (BT-EC) PERSPECTIVE ON "LOCALIZATION", “IT-ENABLEMENT”, "MULTILINGUALISM", "CULTURAL ADAPTABILITY", ETC.

There are a multitude of issues pertaining to cultural adaptability. The 36N0314 document, which resulted in the JTC1 Sophia Resolution #39 focused on "locale identifiers". However, in this SC36 contribution are embedded other related issues and a wider context which were identified by the JTC1 Business Team on Electronic Commerce (BT-EC). It is in this wider e-business and operational context that some key issues are identified and a solution proposed.

The "Report of the JTC1 Business Team on Electronic Commerce (BT-EC)" took a wider horizontal perspective and one which remains valid. {See document JTC1 N5296}

The BT-EC Report (Clause 6) identified four horizontal issues of being of general relevance for all scenarios involving Electronic Commerce and gave these horizontal issues some prominent attention in its work.

These issues are:

- information technology (IT)-enablement;
- localization including multilingualism;
- cross-sectorial aspects; and,
- cultural adaptability.

The JTC1 BT-EC ordered these horizontal issues on the basis of:

- the need to go from the simpler to the more complex challenges;
- placing priority on the "do-able" and immediately most useful in the context of increasing resource constraints in standardization work; and,
- promotion and visibility in ISO/IEC JTC1 work within the ISO, IEC and ITU and especially outside these standardization communities.

The BT-EC noted that from a user perspective, these four horizontal issues need to be addressed in a harmonized manner. The BT-EC also reviewed existing terms and definitions of **"locale"** {see ISO/IEC 15897) focused on aspects related to character sets associated with a natural language (including collating and ordering), date/time formats, monetary formats, etc., a.k.a. "cultural elements".

The JTC1 BT-EC introduced "IT-enablement" as the term used to identify the need to transform currently accepted standards used in commerce world-wide from a manual to a computational perspective. E-business introduces a requirement for standards that are prepared, structured and made available for unambiguous usage within and among information systems.

The objective of IT-enablement is to capture explicitly and in a computer-processable manner, and one which maximizes interoperability, the implicit rules and relations, i.e., those known to "experts, of the code sets found in standards used in commerce world-wide, i.e., capture and state from an entity relationship and/or object technology perspective, using Formal Descriptive Techniques. Also, issues arising from change management in "code tables", i.e., synchronization, backwards compatibility, migration, etc., need to be addressed.

The BT-EC identified four additional sets of parameters of requirements which also needed to be addressed; namely:

- jurisdictional requirements;
- consumer requirements;
- supplier requirements; and,
- human rights related requirements, (e.g., handicapped/disabled, privacy, etc.).

Consequently, BT-EC differentiated between "locale" (as per JTC1/SC22 and ISO/IEC 15897) and "localization" and defined the latter as:

> ***localization***: *pertaining to or concerned with anything that is not global and is bound through specified sets of constraints of:*
>
> *(a) a linguistic nature including natural and special languages and associated multilingual requirements;*
> *(b) jurisdictional nature, i.e., legal, regulatory, geopolitical, etc.;*
> *(c) a sectorial nature, i.e., industry sector, scientific, professional, etc.;*
> *(d) a human rights nature, i.e., privacy, disabled/handicapped persons, etc.; and/or,*
> *(e) consumer behaviour requirements*[25].

This contribution is based on this wider JTC1 perspective and definition of "localization" in the context of e-business requirements.

As such, it takes the current concept and definition of "locale" as defined in Clause 3.1 ISO/IEC 15897:1999 (and placed in a POSIX context) and combines it with a wider set of requirements as well as the need to meet the requirements of jurisdictional domains.

"Locale" is the definition of a sub-set of a user environment that depends on language and cultural conventions. To these are added those of a jurisdictional nature, i.e., as sets of external constraints of jurisdictional domains, including "official languages", public policies (privacy, disabled/handicapped, consumer, etc.).

[25] Here JTC1/SC37 has recently added "(f) health and safety" as an additional set of constraints.

"Localization" is thus a perspective on specifying the sub-set of the environment of a user that supports:

- locale identifiers {country, language, and cultural conventions}; plus,
- jurisdictional domain requirements {country and sub-divisions) + (official) language, + currency, + public policy, (e.g., privacy, consumer, disabled/handicapped) + sectorial (especially those which are regulated)}[26].

[26]It is this integrated perspective and associated set of requirements which is being addressed in the standards development work on ISO/IEC 15944-5.

ANNEX C - NEED FOR A COMMON DEFAULT CONVENTION FOR REFERENCING CODES REPRESENTING "LANGUAGES"

C.1 SUMMARY OF ISSUES

From a suppliers perspective tailoring one's product or service to the user environment of the local market including jurisdictional levels, demands attention to detail, adaptability and flexibility. Being successful in a market requires not only understanding the needs of clients but also being able to communicate with them in their language. Thus multilingualism is good business, is good for business and those developing products and services should think multilingual from the start.

Unambiguous identification of (natural) language to be utilized or required at the human interface is an e-business transaction requirement. Each jurisdictional domain has its own "official language" or "de jure" language requirements for language(s) to be utilized, i.e., like that for product packaging and labelling[27].

The source of the ISO 639 multipart standard for "Codes representing languages" is ISO TC37/SC2. Several code sets exist. SC32/WG1 "Open-edi" which is developing the generic e-business standards, i.e., ISO/IEC 14662 and the multipart ISO/IEC 15944 series for business semantic descriptive techniques therefore prepared a liaison request to TC37/SC2 pertaining to a standard "default" convention for referencing ISO 639-2 code sets. {See document 32 N0672}

The issues are summarized as follows:

< the original ISO 639:1988 (E/F) (First Edition) *Code for the representation of names of languages/Code pour la représentation des noms de langue* and now the revised and updated ISO 639-1:2001 *Codes for the representation of names of languages - Part 1: Alpha-2 code/Codes pour la représentation de noms de langues - Partie 1: Code alpha-2* has a 2-alpha code set. Canada, the USA and other countries did not adopt this ISO 639:1998 (E/F) standard as a national standard in their countries because it lacked codes for recognized languages in their countries, (e.g., native and aboriginal languages).

< ISO 639 was divided into two parts; namely:

- Part 1: *Alpha-2 code* / Partie 1: *Code alpha-2*; and,
- Part 2: *Alpha-3 code* / Partie 2: *Code alpha-3*.

< ISO 639-2 was developed and issued jointly with ISO TC46/SC32. However, it contains two 3-alpha code sets as stated in Clause 1:

Current situation of ISO 639-2 as stated in Clause 1 "Scope"/«Domaine d'application» ISO 639-2:

[27] See further the ISO/IEC JTC1/SC32 N1063 document titled *"(2nd Draft) Annex C (Normative) Codes Representing UN Member States and Their Official (or de facto) Languages" forming part of draft CD Document (version 2.0) of ISO/IEC 15944-5 Information technology - Business agreement semantic descriptive techniques - Part 5: Identification and mapping of various categories of jurisdictional domains as external constraints"*. This earlier draft of document is also available as SC32/WG1 N0261 and SC36N0603.

"...provides two sets of three-letter alphabetic codes for the representation of names of languages, one for the terminology applications and the other for bibliographic applications. The code sets are the same except for twenty-five languages that have a variant language code because the criteria used for formulating them (See 4.1)..."

Further, the relevant text of Clause 4.1
The two code sets are identified as:

"ISO 639-2/B" for bibliographic applications, i.e., "Code set B",

"ISO 639-2/T" for terminology applications, i.e., "Code set T".

Quoting Clause 4.1 the:

"criteria for selecting the form of a language code for code set B were:

- *preference of the countries using the language;*
- *established usage of codes in national and international bibliographic databases; and,*
- *the vernacular or English form of the language.*

Code set T was based on:

- *the vernacular form of the language; or,*
- *preference of the countries using the language".*

Further, ISO 639-2 states in Clause 4.1:

"The bibliographic or terminology code set must be used in its entirety, and the choice of the set used must be made clear by exchanging partners prior to information interchange. Users shall refer to ISO 63902/B for the code set for bibliographic applications and ISO 639-2/T for the code set for terminology applications".

From a global world-wide Open-edi and e-business, e-commerce, e-government, e-logistics, etc., business transaction perspective, one needs to be able to identify, reference and re-use code sets. There is an interoperability requirement for a single default set of codes representing names of languages to be utilized at the IT interface (and thus to be able to generate their human interface "linguistic" equivalents).

ISO 639-2 contains two candidate code sets. Unfortunately, from an Open-edi perspective (and e-commerce/e-business perspective), there are twenty-three (23) language names that have variant codes assigned depending on the 3-alpha code set chosen. Some of these variant codes apply to languages which have extensive use world-wide, (e.g., Chinese, French, German, etc.). {See further Annex C.3 below}

This causes significant confusion and ambiguities for use of ISO 639-2 in support of Open-edi. For example:

- for referencing the French language in business transaction, does one use the code "fra" as found in ISO 639-2/T or the code "fre" as found in ISO 639-2/B?; or,

- for referencing the German language in business transactions does one use the code "deu" as found in ISO 639-2/T; or the code "ger" as found in ISO 639-2/B?

Ensuring interoperability in Open-edi, i.e., e-business, requires one to chose one of these two code sets as the default standard and then use the set chosen in its entirety.

C.2 SOLUTION/APPROACH ADOPTED BY SC32/WG1 "OPEN-EDI", (E.G., "E-BUSINESS"), AND SC32/WG2 "METADATA" - STANDARD DEFAULT IS ISO 639-2/T

SC32/WG1 "Open-edi" and SC32/WG2 "Metadata" recognize that both the ISO 639-2/B and ISO 639-2/T alpha-3 code sets re in use world-wide.

The ISO 639-2/B alpha-3 code set reflects its USA Library of Congress origin and is used extensively in the bibliographic world. The 3-alpha codes are derived from the names of the languages as represented using the English language. The subset of Dublin Core Version 1.1 which was fast-tracked via ISO TC46 into the new ISO 15836 used examples of codes for names of languages based on ISO 639-2/B.

However, the 3-alpha codes ISO 639-2/T reflect the names of the languages in that language (using the Latin-1 character set). For example, French speaking use environments use "fra" for "français", German speaking environments use "deu" for "Deutsch", etc.

Faced with the issues presented above and based on the need to minimize ambiguity in information sharing and maximize interoperability of semantic components in commitment exchange (and e-business), ISO/IEC JTC1/SC32 WG1 "Open-edi" (the e-business standardization group including "business descriptive semantic techniques) decided that "ISO 639-2/T" would be its common default code set standard for the identification and referencing of natural languages.

Similarly, the ISO/IEC JTC1/SC32 WG2 "Metadata" also decided that for the revised/new version of ISO/IEC 11179 Data Element and Metadata multi-part series of standards that "ISO 639-2/T" would be its common default code set standard for the identification and referencing of natural languages.

Further, we note that RFC 3066 under Clause 2.3 "Choice of language tag" contains the following statement, and we quote:

> *"When a language has no ISO 639 2-character code, and the ISO 639-2/T (Terminology) code and the ISO 639-2/B (bibliographic) code differ, you MUST use the Terminology code. NOTE: At present, all languages for which there is a difference have 2-character codes, and the displeasure of developers about the existence of 2 code sets has been adequately communicated to ISO. So this situation will hopefully not arise".*

C.3 LIST OF NATURAL LANGUAGES HAVING DIFFERENT ISO 639-2 ALPHA-3 CODES

This list has been developed based on ISO 639-2 "Table 1 - Alpha-3 codes arranged alphabetically by English name of language"/«Tableau 1 - Code alpha-3 classé alphabétiquement par le nom de la langue en anglais»

Table B-1: Examples of Variation - ISO 639-2 "2/B" and "2/T" 3-Alpha Codes Representing the Same Natural Language

English name Nom anglais	French Name Nom français	ISO 639-2/B Bibliographic code Codet bibliographique	ISO 639-2/T Terminology code Codet terminologique
Albanian	albanais	alb	sqi
Armenian	arménien	arm	hye
Basque	basque	baq	eus
Burmese	birman	bur	mya
Chinese	chinois	chi	zho
Croatian	croate	scr	hrv
Czech	tchèque	cze	ces
Dutch	néerlandais	dut	nld
French	français	fre	fra
Georgian	géorgien	geo	kat
German	allemand	ger	deu
Greek, modern (post 1453)	grec moderne (après 1453)	gre	ell
Icelandic	islandais	ice	isl
Macedonian	macédonien	mac	mkd
Malay	malais	may	msa
Maori	maori	mao	mri
Persian	persan	per	fas
Romanian	roumain	rum	ron
Serbian	serbe	scc	srp
Slovak	slovaque	slo	slk
Spanish	espagnol	spa	spa (esp)
Tibetan	tibétain	tib	bod
Welsh	gallois	wel	cym

Spanish is included here because the Note "6)" in ISO 639-2,Table 1 states "After a period of five years from the publication of this standard, **esp** may be used as the ISO 639-2/T (terminology code) for Spanish". «Après une période de cinq ans suivant la publication de cette norme, **esp** pourra être utilisé autant que l'ISO 639-2/T (codet terminologique) pour l'espagnol.» That is, in 2003, the ISO 639-2/T code for Spanish is scheduled to become "esp".

C.4 NOTES ON COVERAGE OF ISO 639-2

Note 1: The current ISO 639-2 alpha 3 code sets pertain only to 466 natural languages. Some of these are no longer in use.

Note 2: At present there are literally thousands of (living) natural languages spoken and/or written in use by "communities of people", somewhere in the world. For example, Nigeria alone is known to have over 500 languages in use. {See for example, Grimes, B.F. Ethnologue. Dallas, Texas: SIL International, 2000. (14th ed., 2 vols)}. As such ISO 639-2 represents only those natural languages for which ISO TC37 has assigned a unique 2 or 3 alpha code.

Note 3: ISO TC37 as part of its development of ISO 639-2 has streamlined procedures and established clear criteria for adding additional "codes" representing the names of languages. It is expected that many are/will be added during the next few years.

Note 4: "Invented" languages and codes for them are not included in ISO 639-2, (e.g., "klingon", "pig-latin", "elfin", etc.).

ANNEX D - LIST OF OVERLAPPING 2-ALPHA COUNTRY CODES AND 2-ALPHA LANGUAGE CODES

The following table provides some examples of the same 2-alpha codes having different meanings and associated representations in ISO 639-1 and ISO 3166-1.

Table D-1: Examples of the same 2-alpha codes being used to designate both an ISO 639-1 language and an ISO 3166-1 "country"

2-Alpha Code	ISO 639-1 Language Representation	ISO 3166-1 "Country" Representation
af	Africaans	Afghanistan
ar	Arabic	Argentina
ca	Catalan	Canada
cr	Cree	Costa Rica
lb	Luxembourgish	Lebanon
ne	Nepali	Niger
sa	Sanskrit	Saudi Arabia
sl	Slovenian	Sierre Leone
ve	Venda	Venezuela
vi	Vietnamese	Virgin Islands (US)

Similar tables can be prepared providing examples of the same 3-alpha codes having different meanings among:

- ISO 639-2 and ISO 3166-1;
- ISO 4217 and ISO 639-2; and,
- ISO 3166-1 and ISO 4217.

ANNEX E - MULTIPLE HUMAN INTERFACE EQUIVALENTS (LINGUISTIC) FOR "CODES REPRESENTING...": COUNTRY CODE EXAMPLES[28]

Associated with any ID code are one or more "human interface equivalents", i.e., a representation of the semantics in linguistic, non-linguistic, (e.g., symbolic) manner suitable for communication to and understanding by human beings.

This Annex E provides an example of the most common human interface equivalents of a linguistic nature for three ISO 3166-1 countries where:

< 246 = Finland;
< 056 = Belgium; and,
< 792 = Turkey.

For each of these are provided in this Annex E, six different representations as valid human interface equivalents (linguistic) for ID codes of these ISO 3166-1 member entities; namely:

< the ISO (UN) short name - English;
< the ISO (UN) short name - French;
< the ISO (UN) long name - English;
< the ISO (UN) long name - French;
< the local, i.e., official, short name(s) in the official language(s)[29] of that country; and,
< the local, i.e., official, long name(s) in the official language(S) of that country.

There are many other human interface equivalent representations in use for the name representations for those three (and other) countries. For example, those from a German language perspective, a Russian language perspective, a Spanish language perspective, etc.

The point here is that agreement on the use of the specific code sets of ISO 639, ISO 3166, and ISO 4217, i.e., as ID codes, allows one to generate any equivalent human interface equivalent name representation as may be required by the (human) user at the user interface.

[28]The country code examples provided here are taken from an earlier SC32N0462 document titled *"Making Standards Work in Electronic Commerce and Among Jurisdictions: IT-Enablement of Data Element-based Standards - Presentation at the Open Forum on Metadata Registries in Santa Fe"*.

[29]For the complete set of ISO 3166-1 entities which are "countries, i.e., UN member states, and their official (or de jure) language(s), see JTC1/SC32 document N1063.

Common IT Interface	Human Interface Equivalents (Linguistic)					
3166-1:246	Alpha-2: FI			Alpha-3:FIN		
3166-1:246	Short Name	(en):	Finland	Long Name	(eng):	Republic of Finland
3166-1:246	Short Name	(fr):	Finlande	Long Name	(fra):	République de Finlande
3166-1:246	Local Short Name	(fi):	Suomi	Local Long Name	(fin):	Suomen tasavalta
3166-1:246	Local Short Name	(sv):	Finland	Local Long Name	(swe):	Republiken av Finland
3166-1:056	Alpha-2: BE			Alpha-3: BEL		
3166-1:056	Short Name	(en):	Belgium	Long Name	(eng):	Kingdom of Belgium
3166-1:056	Short Name	(fr:)	Belgique	Long Name	(fra):	Royaume de Belgique
3166-1:056	Local Short Name	(nl):	Belgie	Local Long Name	(nld):	Koninkrijk van Belgie
3166-1:056	Local Short Name	(fr):	Belgique	Local Long Name	(fra):	Royaume de Belgique
3166-1:792	Alpha-2: TR			Alpha-3: TUR		
3166-1:792	Short Name	(en):	Turkey	Long Name	(eng:)	Republic of Turkey
3166-1:792	Short Name	(fr):	Turquie	Long Name	(fra):	République turque
3166-1:792	Local Short Name	(tr):	Turkiye	Local Long Name	(tur):	Turkiye Cumhuriyeti