CAW N 21

Draft Version 2.0
December, 1997

IT-ENABLEMENT OF EXISTING STANDARDS USED IN COMMERCE

(Plus Localization and Multilingualism)

An ISO/IEC JTC1/BT-EC and JTC1/CAW Discussion Paper

Prepared for the:

JTC1 Cultural Adaptability Workshop (CAW)
JTC1 Business Team on Electronic Commerce (BT-EC)

Prepared by:

Dr. Jake V.Th. Knoppers,
Canaglobe® International Inc.

154 Slater St. #2. Ottawa Ontario, CANADA K1P 5H6
Tel: (613) 234-3244 Fax: (613) 234-3935 E-mail: mpereira@istar.ca

TABLE OF CONTENTS

IT-Enablement of Existing Standards Used in Commerce

An ISO/IEC JTC1/BT-EC and JTC1/CAW Discussion Paper

1.0 INTRODUCTORY NOTES

1.1 This is a personal contribution. The opinions expressed are those of the author and not necessarily those of his organization.

1.2 The purpose of this discussion paper is to assist the work of the ISO/IEC/JTC1/BT-EC (Business Team on Electronic Commerce) and in particular the work pertaining to horizontal issues.

1.3 This discussion paper is based on and benefits from:

- the discussion at the BT-EC Planning Meeting, 28/29 September, 1997 in Brussels, where four horizontal issues were identified and recognized as of being of general relevance for all scenarios involving electronic commerce. Therefore, they are to retrieve permanent attention in the future work of the BT-EC. These four horizontal issues are:

  - localization
  - cultural adaptability
  - cross-sectorial
  - IT-enablement

  {See further BT-EC N021}

- the preparatory work for and participation of the two authors in the Global Standards Conference (GSC) held 1-3 October, 1997, in Brussels. {See further BT-EC N028}. Here Dr. Jake Knoppers served as Regional Coordination (Canada) for Theme #1 Workshop on Electronic Commerce {See further BT-EC N020} and presented a paper titled "Global Electronic Commerce Through Localization and Multilingualism". {See further, http://www.ispo.cec.be/standards/conf97/abstract/abs4.html}

- work by the author on (1) transforming a Canadian national standard CGSB 171.3-94 **Directory Information Describing Digital Geo-referenced Data Sets** from its then current form as a "printed paper" standard into an information technology-enabled and bilingual (English/French) metadata standard; (2) development of a "paperless" implementation of reporting requirements of the Canadian Environmental Assessment Act {See URL http://www.ceaa.gc.ca}; (3) work completed on a bilingual Canadian federal government name and addressing information technology standard (covering individuals and organizations as well as physical and electronic locations); and, (4) participation in international standardization work in areas of IT, trade, geomatics, and EDI/EC.

- work by the author and David Clemis on the preparation of another discussion paper prepared for the BT-EC titled "Business Requirements: Localization and Cultural Adaptability - An ISO/IEC JTC1/BT-EC Discussion Paper" {See further BT-EC N029}

- the results of discussion on an earlier draft version of this contribution at the 13-14 November, 1997 Open Meeting of the BT-EC held in Atlanta (Georgia, USA). {See further JTC1/BT-EC N038}

1.4 The secretariat of the BT-EC works by electronic means. The BT-EC Nnnn document numbers in this discussion paper refer to those found on the BT-EC document server. For further information on accessing this Web site contact Dr. Ingo Wende, DIN, at e-mail <wende@ni.din.de> or fax: +49 30 26 01 1723.

1.5 This contribution is prepared jointly for the JTC1 Cultural Adaptability Workshop (CAW) and the JTC1 Business Team on Electronic Commerce (BT-EC). The primary purpose of this contribution is to bring to the fore electronic commerce-related perspectives on multilingualism and localization which go beyond "cultural adaptability" as "the special characteristics of natural languages and commonly accepted rules for their use (especially in written form) which are particular to a society or geographic area". {See further the (draft) definition of "cultural adaptability" as found in ISO/IEC JTC1 N4627 (re-issued as JTC1/CAW N04)}

"Cultural Adaptability

The special characteristics of natural languages and the commonly accepted rules for their use (especially in written form) which are particular to a society or geographic area. Examples are: national characters and associated elements (such as hyphens, dashes, and punctuation marks), correct transformation of characters, dates and measures, sorting and searching rules, coding of national entities (such as country and currency codes), presentation of telephone numbers, and keyboard layouts.[1]

The requirement for JTC 1 standards to enhance and ensure Interoperability is consistent with the approved JTC 1 Policy on Interoperability.

These characteristics are from the attributes identified in ISO/IEC JTC 1 GII N 123, ISO/IEC JTC 1 GII Roadmap: Guidelines for Evolution, Management and Development of GII Standards. While all the attributes in this document are important and should be taken into account as appropriate, we believe Interoperability, Portability and Cultural Adaptability are critical to the success of JTC 1 standards.

The applicability of these three requirements will vary across the technology directions. Work in each of the technology directions must properly take in account these requirements".

1.6 ISO/IEC JTC1 at its September, 1997 Plenary (Ottawa) also passed a Resolution #43 pertaining to the organization of a JTC1 Cultural Adaptability Workshop (CAW). Sponsored by JTC1/SC2 and hosted by the Standards Council of Canada, the CAW is scheduled to be held 20-22 January, 1998 in Ottawa.

[1] The text of this paragraph is the same as that found in ISO/IEC JTC 1 GII N 123 as the definition for the term "cultural elements".

1.7 The Secretariat for CAW is the Japanese member body. It also works by electronic means. CAW referenced documents are posted to the following web site: "http:www.itscj.ipsi.or.jp/caw"

1.8 In addition, the first version of this contribution was discussed at the 11 December, 1997 meeting of the Canadian Advisory Committee for JTC1/SC2 which is coordinating Canadian preparation for the JTC1 Cultural Adaptability Workshop (CAW) to be held 20-22 January, 1998 in Ottawa, CANADA. This Version 2.0 benefits from these discussions.

[Note: The Version 2.0 of this contribution now known as JTC1/BT-EC N046, has been prepared and submitted to JTC1/CAW in order to meet its 19 December, 1997 deadline for contributions].

2.0 OBJECTIVES

The objectives of this discussion paper are four-fold; namely:

- to serve as input of the work of ISO/IEC JTC1/BT-EC and in particular is work on "horizontal issues";
- to solicit additional feedback on the horizontal issues identified by the BT-EC especially from those BT-EC members not able to attend meetings but participating via the BT-EC document server, thereby furthering the work of the BT-EC in this area;
- to serve as a follow-up activity to the recommendations of the "ISO High-Level Steering Group on CALS: Report to Technical Management Board"; and,
- to serve as input of the discussions of the ISO/IEC JTC1 Workshop on Cultural Adaptability introducing perspectives and issues from an electronic commerce perspective including the examples of issues pertaining to examples such as country and currency codes. It is important that at this JTC1 Workshop, the perspectives of the BT-EC are presented (even if they are of a preliminary nature).

[Note: The BT-EC has scheduled its next (and final) Open Meeting 26-29 January, 1998 (Brussels) to be able to benefit from the results of the Cultural Adaptability Workshop].

3.0 NEED FOR PRAGMATIC ACTIONS

The key characteristic of global commerce is that, on the whole, it consists of business transactions which are:

(1) rule-based, i.e., mutually understood and accepted sets of business conventions, practices, procedures, etc; and,

(2) code/table-based representing choices, i.e., schemes of tables and lists of codes representing countries, currencies, stocks., manufacturers, geographic locations, tariffs, commodities, catalogues, specifications, prices, terms and conditions of trade, transport insurance, etc.

Each of these represents a discrete building block for global commerce. They also interwork. Further irrespective of how enterprises or chains of enterprises agree to interwork to bring goods and services to the physical and/or virtual marketplace, they all use these common building blocks consisting of discrete sets of data/codes of and associated rules, i.e., "dapplets". The challenge here facing enterprises and standardization bodies alike is that currently these data/rule sets, or dapplets, are not IT-enabled nor EDI and electronic commerce-facilitated. Both are still presented from a human perspective.

For example, anyone engaged in commerce globally as well as locally uses many different codes in a business transaction. These different codes represent a business "shorthand". In our daily business transactions, we know, i.e., have memorized, many of the different codes, others we have to look-up in a table, (e.g., for infrequently used codes especially if suggested by one's trading partner). We also know the business rules associated with each code, (e.g., "C.O.D." or "Cash-on-Delivery" means that the seller transports the goods to a delivery point specified by the buyer, and the buyer pays the full amount owed, (e.g., cash, cheque, credit card, as allowed by the seller) before the goods are transferred from the seller to the buyer.

Many of these code sets and associated rules are still presented from a human perspective. Even if the code values are captured in specific data elements in EDI-based transactions, the associated rules are not. Even worse some of the rules are not explicitly stated, they are implicit and require experts to explain them especially if one wants to capture and support these commerce "code/rule sets" in the IT system of the enterprise.

This costs time and money for each enterprise in essentially duplicating common non-competitive requirements.

Among trading partners their commercial relations can be characterized as early "loose couplings" and "late bindings". An example of a "loose coupling" is a decision by a trading partner to transport products by air (vis-à-vis rail, truck, marine). This brings into play automatically the rules and code tables of the air transport sector, i.e., those of ICAO and particularly IATA. This "loose coupling" is then transacted through a request for quote (and/or reservation) of space (NN) for transport of one ton (KG 1000) of horse meat (CMOD -nnn-0630) from Montreal/Mirabel (YMX) to Paris/Charles de Gaulle (CDG) with charges, insurance freight prepaid (CIF, PPD) and horse meat being a perishable commodity (PERC). Responses on availability of space, price, etc, are received. The shipper then makes the "late binding" decision, i.e., a confirmed reservation. The "(NN)", YMY", "CIF", PPD", CMOD, etc., are all codes taken from various code tables representing sets of permissible values, and as applicable, associated business rule sets. Similarly, the financial services sector is also highly organized.

Many of these sets of agreed upon rules used in business world-wide and their associated lists of tables/codes are "de jure" and "de facto" standards. They are, however, "paper-based" and human

understandable (although at times a level of expertise and experience is required). They are not however, IT-enabled, i.e., "computer processable", (e.g., as callable data objects or "dapplets").

The JTC1/BT-EC Planning Meeting (29-30 September, 1997) identified "IT-enablement" as one of the four horizontal issues of general relevance for all scenarios for modelling various ways for doing electronic commerce and described it as follows:

> *"IT-Enablement: transition of currently accepted standards of identification, process, etc., from a manual to a computational perspective*
>
> - *harmonization of existing standards of identification and identification of computational transition; and,*
> - *identification of new standards and the incorporation of transition to computational perspective.*
>
> *Impediments*
>
> - *ambiguous identification*
> - *multiple definition of objects*
> - *"Versioning" and other administrative issues*
>
> *Scope*
>
> - *transition between existing manual systems to automated systems".*

For historical reasons, standards at national and international levels have evolved with the scope of the business communities, meeting identified needs and application approaches.

Electronic commerce introduces a requirement for standards that are prepared, structured and made available for unambiguous usage within and among information systems. This requirement can be expressed as "**computational integrity**", in particular:

> *"the expression of standards in a form that allows precise description of behaviour and semantics in a manner that allows for automated processing to occur, and the managed evolution of such standards in a way that enables dynamic introduction by the next generation of information systems".*[2]

Numerous international standards are already in use in support of commerce world-wide. The problem is that they are paper-based and lack a computer process-able version. Key examples here are ISO 3166 Country Codes, ISO 4217 Currency Codes, and ISO 639 Language Codes. Even if distributed in electronic form, these ISO standards used in commerce world-wide consist of tens of "printed" pages available in hard or soft form. They cannot be "plugged-in" for use in electronic

[2] Source of draft definition is Stephen McConnel, OMG. This is an important concept/term/definition, comments are especially welcome here.

commerce. Much of intelligence in these (and other) international standards is human understandable explicitly or implicitly. From an entity-relationship or object-oriented perspective, these and other ISO (and non ISO) "code set" standards are "a mess". They have not formally been described using Formal Description Techniques (FDTs), i.e., in their present form they do not support "computational integrity". Consequently, each enterprise using these code sets has to spent considerable time and effort to: (1) figure them out and interpret them; (2) build applications; and, (3) pray that their interpretation is interoperable when they interwork with other networks or enterprises let alone supporting changes and updates.

Human beings like to name "objects". But the approach of using "names" is not very IT friendly, cost-efficient or efficient. Here for global electronic commerce, we must note the ISO/IEC 1087 definition *"**name**: designation of an object by a linguistic expression".*

Depending on the interplay of multilingual and localization requirements, it is a fact that in electronic commerce, a singular product or service being offered for sale will have multiple names and differing names even in the "same" language. Thus if we wish to ensure rapid and widespread use of electronic commerce globally, we must on the one hand identify "objects", i.e., products or services being offered for sale, in an unambiguous, linguistically neutral, and IT-process-able and EC-facilitated manner, and, on the other hand, present the same via a range of linguistic names (and associated character sets) from a point-of-sale perspective, i.e., human readable, as required by the "local" marketplace.

In this discussion paper, we present, four (4) examples to start work on IT-enablement:

- the ISO 3166 standard - "Country Codes"
- ISO 639 - "Language Codes"
- the Harmonized System - "Potato"
- ISO 4217 - "Currency Codes.

The paper ends with a recommendation in support of mirroring one of the key results from the Global Standards Conference (GSC), i.e., "that standards must focus on interface not the implementation".

4.0 EXAMPLES

4.1 EXAMPLE #1 ¾ ISO 3166 - COUNTRY CODES

The title of ISO 3166 is "Codes for the representation of names of countries and their subdivisions".

Within ISO 3166, there are now three parts; namely:

- Part 1: Country Codes;
- Part 2: Country Subdivision codes; and,
- Part 3: Code for formerly used names of countries.

Here ISO 3166-1 "established codes that represent the names of countries, dependencies, and other areas of particular geopolitical interest, on the basis of lists of country names obtained from the United Nations". Currently, each entry (or "record" of a permitted instance) contains:

(1) a three-digit numeric code
(2) a two letter alpha code
(3) a three letter alpha code
(4) a short name - English
(5) a long, i.e., formal name - English
(6) a short name - French
(7) a long, i.e., formal - French

Of these, the three digit numeric code is the most stable and tends to change only when the physical boundaries change. Names short and long do change and at times the accompanying two and three-letter alpha codes as well. ISO 3166 also has a note field.

The 3-digit numeric code is linguistically neutral and unambiguous. Each of these 3-digit numeric codes has in ISO 3166 associated with it six (6) alphabetic linguistic expressions, two of which also serve as "human understandable" (and computer-process-able codes).

ISO 3166-1 thus has seven (7) "standardized" representations for each unique entity or object, three (3) of which are codes. ISO 3166-1 allows any one of the seven to be utilized although in practice and especially in IT systems one usually utilizes one of the three codes.

In addition, we should note the fact that in their own locale and language, countries have their own short and long (or formal) name. For example 528 Netherlands = "Nederland" and "Koninkrijk der Nederlanden"; 578 Norway = "Norge" and "Kongerike Norge; 280 Germany = "Deutschland" and "Bundesrepublik Deutschland"; 246 Finland = "Suomi" and "Suomen tasavalta"; 792 Turkey = "Turkiye" and "Turkiye Cumhuriyeti"; 484 Mexico = "Mexico" and "Estados Unidos Mexacanos". Further, countries which being bilingual (multilingual) have two (or more) sets of local short and long/formal names namely: 058 Belgium = "Belgie" and "Koninkrijk van Belgie", and "Belgique" and "Le Royaume de Belgique", or 246 Finland = "Suomi" and "Suomen tasavalta", (e.g, Switzerland).

Further, there is the fact that many countries use not Latin Alphabet-based character sets. This means that one also has the original country language character script as "alphas" plus their "latinized" equivalents.

To this is added the fact that in each locale and language, countries have their "own names". For example, a person in France uses "Allemagne" not "Germany" or "Deutschland" as the linguistic equivalent for "3166:280". All this is common, non-competitive information.

For this ISO 3166 standard, we currently do not have a "standard" for the interface. The 3-digit numeric code, the 2-alpha code and the 3-alpha code are all used in interchanges. However, while name changes do occur, the 3-digit numeric code remains the same for a country unless the physical entity changes, (e.g., Burma to Myanmar, Zaire to the Democratic Republic of the Congo, etc.).

The use of Alpha-3 code tags for ISO 3166 country name causes overlap confusion with ISO 4217 currency and funds codes which are represented as Alpha-3 codes (upper case).

In addition, ISO 3166 Alpha-2 and Alpha-3 codes are not that stable, i.e., whenever a country changes its name it often also changes, its alpha-2 and alpha-3 codes. The 3-digit numeric code is much more stable. It changes only when the actual physical boundaries of the countries change, i.e., the entity being identified and referenced is no longer the same. For example, the alphabetic (written language) equivalents to "3166:180" recently under went the following changes:

	Former	**New**
Alpha-2	ZR	CD
Alpha-3	ZAR	COD
Short Name (en)	Zaire	Congo, Democratic Republic of the
Long Name (en)	Republic of Zaire	the Democratic Republic of Congo
Short Name (fr)	Zaïre	Congo, la République démocratique du
Long Name (fr)	République de Zaïre	La République démocratique du Congo

From an interoperability perspective, i.e., both that of commerce and IT, the "3166" identifying rule set and the 3-digit numeric code together form an unambiguous global identifier for the entity being referenced. The alpha codes and names should simply be considered linguistic equivalent expressions.

Further, **it should be noted that ISO 3166-1 contains many instances and codes for entities which are <u>not</u> "countries**", i.e., they are dependencies of other entities, (e.g., France, Great Britain, USA, etc.). Human beings "filter" and easily make these distinctions. To make things even worse many of these 3166-1 "sub-entities" have a different code in 3166-2. [Note: One would have thought that when ISO 3166 was split into its three parts, all the sub-entities in ISO 3166 would no longer be found in 3166-1 but be moved to 3166-2].

It suffices to state that all the rules and intelligence implicit in ISO 3166-1 (as well as 3166-2 and 3166-3) has not yet been captured explicitly in an IT-enabled and EC-facilitated manner, (e.g., as a "normalized" (callable) database).

We conclude this discussion, with Exhibit #1 where, in matrix form, we present in the left column titled "IT-Needs", the suggested "interface" requirement for ISO 3166 and in the right-side columns some of the linguistic equivalents required to support "Localization and Multilingual" requirements for particular implementations and supporting IT applications.

<u>Notes on Exhibit #1</u>

[1] Normally the eight (8) "fields" under "Localization and multiculturalism" would be separate (sets of) "columns" in a database schema all forming part of the "row". it is noted that the physical presentation here does not reflect this.

[2] The 2-letter language codes, (e.g., en, fi, fr, nl, sv, tr), are taken from ISO 639.

[3] The "->" entries are not part of ISO 3166-1. Although only Latin alphabet character set is utilized here, it is understood that "local" means exist which use non-Latin alphabet-based character sets.

EXHIBIT #1 - TOWARDS AN IT-ENABLED STANDARD FOR ISO 3166 - COUNTRY CODES

IT-Needs	Localization and Multilingual [1]	
3166:246	Alpha-2: FI	Alpha-3: FIN
	Short Name (en) [2]: Finland	Long Name (en): Republic of Finland
	Short Name (fr): Finlande	Long Name (fr): République de Finlande
[3] ->	Local Short Name (fi): Suomi	Local Long Name (fi): Suomen tasavalta
->	Local Short Name (sv): Finland	Local Long Name (sv): Republiken av Finland
3166:056	Alpha-2: BE	Alpha-3: BEL
	Short Name (en): Belgium	Long Name (en): Kingdom of Belgium
	Short Name (fr): Belgique	Long Name (fr): Royaume de Belgique
->	Local Short Name (nl): Belgie	Local Long Name (nl): Koninkrijk van Belgie
->	Local Short Name (fr): Belgique	Local Long Name (fr): Royaume de Belgique
3166:792	Alpha-2: TR	Alpha-3: TUR
	Short Name (en): Turkey	Long Name (en): Republic of Turkey
	Short Name (fr): Turquie	Long Name (fr): République turque
->	Local Short Name (tr): Turkiye	Local Long Name (tr): Turkiye Cumhuriyeti

4.2 EXAMPLE #2 ¾ ISO 637 - LANGUAGE CODES

ISO 639 is the standard for codes representing the names of languages. Each entry (or permitted instance) in ISO 639 consists of:

- a language symbol, in the form of a two-letter code;
- the language name - English
- the language name - French
- the original language name (as written in the Latin-1 alphabet).

Here two, initial observations must be made. First is that Canada (and the United States) has not adopted ISO 639 as a "national standard" due primarily to its current lack of inclusion of North American aboriginal and native languages. Secondly, the LANG attribute is important in SGML (ISO/IEC 8879) in the recently proposed New Work Item (ISO/IEC JTC1 N4742) for "Standard HTML". Here the LANG attribute

> *"identifies a natural language spoken, sung, written or otherwise used by human beings for communication between people. Computer languages are explicitly excluded. The value of the LANG attribute is referred to as the "language tag"... The name space of language tags is administered by IANA. Example tags include: en, en-US, en-cockney, i-cherokee and x-pig-latin.*
>
> *Two letter primary tags are reserved for ISO 639 language abbreviations. This Committee Draft does not specify three-letter primary tags, however their description may be found in the "Ethnologue" {Gri92}. Any two-letter initial sub-tag is an ISO 3166 country name..."*

Serious reflection and more systematic thinking is required here especially if one wishes to use SGML → HTML → XML in the GIS/GII in general and in electronic commerce specifically as well as ensuring interoperability not only with the use of other syntaxes but among various consumer markets, industry sectors, etc.

First of all, "i" and "x" are single characters; and they do not exist in ISO 639. Secondly, "cherokee" and "pig-latin" are not ISO 639 languages. Thirdly, for "en-us", it is not clear at all, given the other examples whether this represents English language as used in the United States or something else.

Fourthly, use of Alpha-2 code tags for ISO 3166 country name is confusing since these are not mutually exclusive vis-à-vis ISO 639 language codes. They overlap, i.e., many two letter combinations serve both as country codes and language codes. This at times is confusing for humans (and even more so for "dumb" computers). Fifthly, in many sort algorithms and search/retrieval engines upper and lower case letters are treated the same. This causes even more confusion in IT-enabled processing of these code sets if two letter alphas are used as codes for both countries and languages.

With respect to sets of codes representing "countries" or "languages", the ISO is not the only organization to issue and maintain standards used world-wide.

One set of other coding schema for country codes and language codes is that of the US Library of Congress. Its primary application is in the bibliographic/information sciences domain. It should be noted that these coding schemas pre-date those of the ISO. For country codes the Library of Congress uses two or three character lower case alphabetic codes. These represent existing national entities, provinces and territories of Canada, states of the United States, divisions of the United Kingdom, and internationally recognized dependencies. It is known as the USMARC[3] Code List for Languages and is maintained by the Library of Congress. Similarly the Library of Congress maintains a USMARC Code List for Languages. This code list consists of three letter mnemonic representing only written languages of the modern and ancient world. "*Where one spoken language is written in two different sets of characters, each set of characters is assigned a specific code. For example, Serbian and Croatian are the same spoken language but the former is written in the Cyrillic alphabet and the latter in the Roman alphabet*" ("Roman" known within ISO as "Latin" character set).

Further, the Universal Decimal Classification (UDC) scheme also used in the bibliographic/information science work (primarily in Europe) also has language codes

Human beings can recognize and filter these differences, computers cannot unless explicitly instructed. Keeping in mind that the scope and definitions of these different coding schemes also differ for what are generally the same business needs, one can bridge such differences through construction of concordance tables. This allows one to maximize insofar possible interoperability from across sectorial perspective as well as identifying "non-interoperability" instances.

In Exhibit #2 below, we present a sample concordance table for ISO 3166 + ISO 639 with the Library of Congress (LC) country and language code sets and the UDC language code set.

Notes on Exhibit # 2

[1] As found in ISO 639:1988

[2] For human representation, we have included the ISO 3166-1 "Short Name - English".

[3] For human representation, we have included the ISO 639 English name of the language. There is also the French name and of course the actual "name" of the language in the language itself. ISO 639 captures this "Original" name in its Latin alphabet equivalent version.

[4] One notes LC country codes are not the same as ISO 3166-1 for the same entities.

[3] The acronym "MARC" stands for "Machine Readable Cataloguing". The preceding characters represent the country who utilize the "MARC" format, have amended it for their specific cataloguing needs, and have an infrastructure at the national level for addressing these national needs. There are primarily 3 countries namely the US, Canada, and the UK Hence the designation of, USMARC, CANMARC, UKMARC.

[5] Added here to indicate that in Canada under the Nunavut Act, a new "territory" will be established 1 April, 1999 from the existing Northwest Territories. In Nunavut, in addition to English and French, Inuktitut will become a recognized "official" language. The language code "ik" is the one that has been reserved for Inuktitut.

[6] In ISO 639 the "ik" represents "Inupiak" which is grouped/classified as an "Eskimo language".

[7] The LC codes place "Inuktitut" under the Eskimo family of languages.

[8] One notes that all LC language codes are not the same as ISO 639. At times even the first letter is not the same, (e.g., "nl" versus "dut").

EXHIBIT #2 ¾ SAMPLE COUNTRY AND LANGUAGE CODES CONCORDANCE: ISO, LC AND UDC

ISO				Library of Congress		UDC [1]
3166-1		639				
Numeric Code	Short Name (E) [2]	Applicable Languages (E) [3]	Applicable Language Codes	Country Codes	Language Codes	Language Codes
3166-1:124	Canada	English	en	xxc [4]	eng	= 20
		French	fr		fre	= 40
		Inuktitut [5]	ik [6]		esk [7]	= 947.51
3166-1:056	Belgium	French	fr	be	fre	= 40
		Dutch	nl [8]		dut	= 393.1
3166-1:246	Finland	Finnish	fi	fi	fin	= 945.41
		Swedish	sv		swe	= 398
3166-1:792	Turkey	Turkish	tu	tu	tur	= 943.5
3166-1:840	United States	English	en	xxu	eng	= 20
3166-1:826	United Kingdom	English	en	xxk	eng	= 20
		Scots Gaelic	gd		gae	= 916.3
		Welsh	cy		wel	= 916.6

4.3 EXAMPLE #3 ¾ HARMONIZED SYSTEM - COMMODITY CODES

In combining multilingualism and localization requirements, one must recognize the fact that associated with use of a language, (e.g., English, French German, Spanish, Portuguese, etc.), there are many locale". The same object may well be and is often used and known by different terms in the same language in different local usage conventions. The Universal Product Code (UPC) and European Article Numbers (EAN) systems recognize this as they have multilingual terms associated with each code for "local" packages/labelling purposes. This has implications for Electronic Commerce and particularly that via the Internet. For example English as a language is in use in many countries or "locales", (e.g., Australia, Britain, Canada, India, Ireland, Jamaica, New Zealand, USA, etc.). Similar examples exist for other languages.

In this context, let's take the example of an enterprise wishing to sell potatoes world-wide. This is a simple example yet representative of the interplay of three of the four horizontal issues "localization", "cultural adaptability", and "IT-enablement". This means that these goods have to pass through customs for export/import into various countries. The custom authorities world-wide have an organization that sets common rules and procedures, i.e., the World Customs Organization (WCO), formerly the Cooperative Customs Council (CCC)[4]. The WCO has established a classification scheme for goods traded called the Harmonized System (HS). It was formerly known as the Brussels Tariff Nomenclature (BTN).

Within the Harmonized System (HS) of the WCO, the general code for potato is "0701". This linguistically neutral code "0701" is a data item or data element instance in the HS permitted value domain. Here the German German equivalent name of "potato" is "kartoffel", but the Austrian German equivalent is "Erdapfel". Similarly, the Spanish Spanish equivalent name is "patata" while the Mexican Spanish equivalent name is "papa", and the Dutch equivalent is "aardappel", etc. In French the dictionary term is "pomme de terre", with "patate" as a "local" specific, i.e., Canada/Quebec term (and one which is not slang). The equivalent names noted above are thus cultural adapted equivalent linguistic expressions associated with "0701". Depending on the "locale," the appropriate human oriented names or linguistic expressions can be systematically/automatically generated from the linguistically neutral numeric code for human understanding, labelling, reporting, filing, etc., and, where required, in multiple languages.

In Exhibit #3, we present the "potato" from an IT-enabled and EC-facilitated perspective. Again on the left-hand side of the matrix, under "IT-Needs" we identify the schema ID, i.e, "HS" along with the permitted value, i.e., in this case 0701 for potato. In the middle column, we present examples of countries that import potatoes, i.e., through their ISO 3166 country code and short name (English), while on the right-hand side, we present linguistic equivalents required to support local and multilingual requirements from both a jurisdictional and consumer perspective.

[4] The WCO is but one example of "coordinated autonomy" among autonomous organizations. The degree to which autonomous organizations achieve interoperability from a business operational perspective sets the limit to the extent of interoperability of supporting IT-based functional services.

Exhibit #3 also demonstrates two key aspects of the interworking of "localization", "cultural", and "multilingual" requirements; namely:

(1) that within a jurisdiction, (e.g., a country, a province, canton, etc.), there can be more than one natural language of use; and,

(2) that localization needs can result in a product, i.e., entity or object, having more than one equivalent "name" within a particular natural language.

EXHIBIT #3 - TRADE IN GOODS - POTATO (CULTURAL ADAPTABLE) [1]

IT-Needs	Country Code - Short Name (en) [2]	Localization and Multilingual Needs [3]
HS: 0701	124 CANADA	(en): potato (fr): pomme de terre (ik): patiti [4]
	464 MEXICO	(es): papa
	724 SPAIN	(es): patata
	040 AUSTRIA	(de): erdapfel
	276 GERMANY	(de): kartoffel
	056 BELGIUM	(fr): pomme de terre (nl): aardappel
	246 FINLAND	(fi): peruna (sv): potatis

Notes on Exhibit #3

[1] Example #3 focuses on human understandable representation of an IT-enabled global standard for trade in goods based on the existing Harmonized System (HS) of the World Customs Organization (WCO). The example here is "potato" where, under the HS, "0701" is the primary code for "potato", and ".01" is for seed potato, while ".09" is for "other potatoes".

[2] The country Code and short name are taken form ISO 3166-1.

[3] The 2-letter language codes, (e.g., de, en, es, fi, fr, ik, nl, sv), are taken from ISO 639.

[4] In 1999, Nunavut will become a new territory with Inuktitut as an added "official" language to English and French. In Inuktitut "potato" is "patiti" (Latin character set equivalent) or "_ _ _" in Inuktitut character set.

4.4 Example #4 ¾ ISO 4217 - CURRENCY CODES

ISO 4217 is the standard for codes representing currencies and funds. The principles for inclusion in the code lists of ISO 4217 is that (1) they must be/represent currencies and funds used within the entities described by ISO 3166; and, (2) the codes listed are intended to reflect current status, at the date of publication.

ISO 4217 also has a number of features and anomalies which although human understandable need to be identified and explicitly captured in an IT-enabled manner.

The standard includes funds and their codes in addition to currencies and their codes. Here we note that there are in ISO 4217, ISO 3166 entities where:

- the three digit country code not being the same as the three digit ISO 4217 3-digit code, (e.g., due to the creation/utilization of ISO 4217 of ISO 3166 "User Extensions"). For example, one can readily identify in ISO 4217 twenty-five (25) instances for ISO 3166 entries where the ISO 3166 Country Codes 3-digit numeric differs from the ISO 4217 "Code Name" 3-digit numeric. Nor is there any relation between many of the ISO 3166 and ISO 4217 alpha codes.

- a country (or dependency) has no currency of its own and utilizes the currency of another country;

- a country has more than one currency, i.e., its own and that of another country;

- countries having both a currency code and a funds code;

- a set of countries collectively sharing and using a currency which has no "issuing country", (e.g., SDR, XDR, XOF, and XAF). Here to set, one need to add the "euro" as currency (in addition to the "ecu", i.e., XEU;

- special fund types;

- "currency" not linked to any country or organization, (e.g., precious metals such as gold - 959, alpha = XAU, special settlement currencies, etc.); and,

- "currencies" having no numeric code but only a 3-alpha code, (e.g., XFO = Gold Franc).

Again, here one also needs to develop a common approach for explicitly capturing the rules, code tables, variations, etc., in a systematic manner, (e.g., as an "intelligent" normalized database).

Experience in the financial services/banking sector indicate that on the Internet those engaged in electronic commerce as well as in general text, e-mail, etc., business transactions need to be made aware of standard notation for currencies. For example, in actual practices the Canadian dollar is being represented as "CDN", "CAN", "CA",, etc. Further, the 3 alpha codes of ISO 3166-1 for countries often is confused with the ISO 4217 currency code.

5.0 NEXT STEPS? BUSINESS AND USER PERSPECTIVE

5.1 STANDARDIZE AT THE INTERFACE

On 1-3 October, 1997 in Brussels a global standards conference was held. Sponsored jointly by the ISO, IEC, and ITU its title was "Building the Global Information Society (GIS) for the 21st Century: New Applications and Business Opportunities, Coherent Standards and Regulations".[5] The Conference had four (4) major themes developed through parallel workshops. Key results of the Theme #1 - Electronic Commerce" workshop applicable to this contribution (paraphrasing the Conference Rapporteur) are:

(1) standards must focus on the interface (as opposed to implementation) as the best means of arriving at globally harmonized solutions for interoperability of ICT products and services.

(2) local requirements can be accommodated. Multilingualism must be considered. The expansion of open, multilingual standards could significantly increase the volume and value of world-wide GIS. These standards should act as a catalyst to use and industry-led innovations that better serve the globe's rich and diverse cultures.

(3) standard interfaces must be technology neutral accommodating advances in technology to the extent possible.

(4) in order to empower users and consumers, standards should be adaptable to cultural and linguistic needs on a national and regional levels, while ensuring full transparency on available market solutions to the consumer.

In this discussion paper, we have identified three (3) "de jure", i.e., ISO standards, and one (1) "de facto", i.e., WCO standard. They are currently used for commerce world-wide as well as in many other areas. These four (4) standards are currently not cast in an IT-enabled and EC-facilitated form. They are, however, being used and implemented by enterprises and their IT systems in a wide variety of different ways. This does not promote global interoperability including that from a cross-sectoral perspective which is another key BT-EC horizontal issue.

The results of the analysis leads one to the following conclusions:

- the urgent need to update ISO 639 to include North American aboriginal and native languages, (and if necessary provide for user extensions for "cockney", "pig latin" "klingon", etc.),

- clean up the confusion arising from the (1) use of two-letter alpha codes for both country codes and language codes; and, (2) use of three-letter alpha codes for both country codes

[5] {See further JTC1/BT-EC N028 "Report on the Global Standards Conference". In this Report the URLs of the host site for the Conference and "homepage" are provided as well as those for the Workshops including that on Electronic Commerce}.

and currency codes by taking a much more systematic and IT-enabled approach by adopting the following rules from a standard interface perspective, namely:

(1) For language code, use ISO 639 Alpha-2 code (lower case)

(2) For currencies and funds, use ISO 4217 Alpha-3 codes (upper case)

(3) For countries, use ISO 3166 3-digit numeric codes.

(4) For (1), (2), (3), any additional codes required by users these should become registered user extensions with ISO 639, 4217, and 3166 respectively.

(5) For human interface, one can present the appropriate linguistic expression for each of these codes in the language of the user as shown in the following example,

Entity	Code	Linguistic Expressions*		
Language	tr	Turkish	turc	Türkçe
Currency	TRL	Turkish lire	Lire turque	?
Country	792	Turkey	Turquie	Turkiye

Notes:

(1) * There are many other linguistic equivalent expressions here as well as such as long names and for bilingual $\rightarrow$ tri-lingual^{+} countries even more. In addition, each country/language combinations has its own linguistic expressions, (e.g., people in France send mail to "Allemagne", people in the Netherlands to "Duitsland", people in the United States to "Germany", etc., while people in that country call it "Deutschland". The same applies to "local" use of names of languages or currencies, etc.

(2) In addition, there are other equivalent expressions for these ISO codes. For example, in the Universal Decimal Classification (UDC)[6] schema the language code for "tr" is "943.5".

One result and significant benefit of adopting the above approach from a standard interface perspective is unambiguous combinations of these three ISO standards, i.e., country/language "NNNaa" or "aaNNN"; country/currency "NNNAAA" or AAANNN", NNNaaAAA, etc., for electronic data interchanges electronic commerce purposes.

5.2 IT-ENABLEMENT

[6] The maintenance agency for this scheme is the International Federation for Information and Documentation (FID) = Féderation Internationale d'Information et de Documentation (FID), in the Hague.

One needs to address the IT-enablement of ISO 639, 3166 and 4217 in an integrated manner, i.e., to ensure the interoperability required among these three in support of global electronic commerce.

One can state quite bluntly that irrespective of advances in information technologies, reduced per unit costs in its acquisition and/or use, etc., it is difficult to envisage (cross-sectorial) interoperability required for global electronic commerce, if one cannot sort out and rectify current problems with elementary lego-block standards such as the three ISO standard examples provided here.

Standardization development work in JTC1 should be of use here, (e.g., that of JTC1/SC31 Data Management Services). However, an even greater challenge is that of JTC1 interworking with ISO TCs to IT-enable their existing standards used in commerce as well as those of other *de jure* and *de facto* standards, business conventions, etc.

5.3 MULTILINGUAL AND LOCALIZATION ADAPTABILITY

This contribution provided some examples of adaptability of existing standards to meet multilingual and localization requirements.

This area requires further analysis and work. Hopefully, the results of the JTC1 Cultural Adaptability Workshop will provide inputs and further insights on these issues.

c:\wp51\doc\swg\swg-m245.doc 97.12.19 JVK/mjp