L2/11-236

ISO TC 37/SC 2/WG 1 N204

Date: 2011-06-07

ISO 639-0

ISO TC 37/SC 2/WG 1

Secretariat: SCC

Code for the representation of names of languages — Part 0: Revision of linear text

Code pour la représentation des noms de langue — Partie 0

This document is the second Working Draft of the "next generation" ISO 639, which is intended to replace all linear text of all current parts of the ISO 639 series of International Standards, making this document the only linear (paper) document supporting ISO 639.

This Working Draft is made available to ISO/TC 37/SC 2/WG 1 and to ISO 639/RA-JAC for comments.

Document type: International Standard
Document subtype:
Document stage: (20) Working Draft
Document language: E

H:\Mine Dokumenter\ISO 639\TC037 SC02 WG01\N documents\N200-299\N204 WD 639-0 (2011-06-07).docx STD Version 2.1c

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 639 was prepared by Technical Committee ISO/TC 37, *Terminology and other language resources*, Subcommittee SC 2, *Terminographical and lexicographical working methods*.

This edition of ISO 639 cancels and replaces ISO 639-1:2002, ISO 639-2:1998, ISO 639-3:2007, ISO 639-4:2010, ISO 639-5:2008, and ISO 639-6:2009.

Introduction

ISO 639 provides codes for the identification and specification of individual languages, language variants, and language groups. The identifiers may be used in a variety of applications, including specification of the language used in a text, the language of terms or words in a dictionary or terminological database, the language used in a spoken presentation, language proficiency, language capabilities of software, localization, etc. Various parts of ISO 639 have already been implemented in a number of environments, as will this new version.

[There should possibly be a chapter of actual uses of the standard. Some sort of reference to IETF, W3C, Unicode, etc., although we cannot formally reference these documents. Possibly either as separate TS or informative annex. It should be an annex.]

Code for the representation of names of languages — Part 0: Revision of linear text

To the title: Agreement to retain “code” in singular.

What about “names of languages”? “Linguistic entities”? Not just “languages”.

Proposal: “Code for the representation of language”.

Avoid mention of “names of languages” in the text.

1 Scope

This International Standard sets out the general principles of language coding using the codes that are specified in the ISO 639 database and combinations with other codes. In addition, this International Standard lays down guidelines for the use of any combination of subsets of the ISO 639 code table.

[Some more work should be put into the scope.]

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this International Standard. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 3166-1:2006, *Codes for the representation of names of countries and their subdivisions — Part 1: Country codes.*

ISO 3166-2:2007, *Codes for the representation of names of countries and their subdivisions — Part 2: Country subdivision code.*

ISO 3166-3:1999, *Codes for the representation of names of countries and their subdivisions — Part 3: Code for formerly used names of countries.*

ISO 8601:2004, *Data elements and interchange formats — Information interchange — Representation of dates and times.*

ISO/IEC 11179-1:2004, *Information technology — Metadata registries (MDR) — Part 1: Framework.*

ISO/IEC 11179-2:2005, *Information technology — Metadata registries (MDR) — Part 2: Classification.*

ISO/IEC 11179-3:2003, *Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes.*

ISO/IEC 11179-4:2004, *Information technology — Metadata registries (MDR) — Part 4: Formulation of data definitions.*

ISO/IEC 11179-5:2005, *Information technology — Metadata registries (MDR) — Part 5: Naming and identification principles.*

ISO/IEC 11179-6:2005, *Information technology — Metadata registries (MDR) — Part 6: Registration.*

ISO 12620: —, *Computer applications in terminology — Data categories*[1].

ISO 15924:2004, *Information and documentation — Codes for the representation of names of scripts.*

ISO 16642:2003, *Computer applications in terminology — Terminological markup framework.*

ISO 19111:2007, *Geographic information — Spatial referencing by coordinates.*

ISO 19112:2003, *Geographic information — Spatial referencing by geographic identifiers.*

3 Terms and definitions

For the purposes of this International Standard, the following terms and definitions apply.

NOTE The definitions in this International Standard are intended for practical use within the context of ISO 639 and its applications. For various linguistic purposes there are needs for more detailed, and possibly deviating, definitions.

3.1
code
data transformed or represented in different forms according to a pre-established set of rules

NOTE The usage of the term “code” is not uniform in all standardized coding systems. According to the usage that is defined in this International Standard, a “code” to be understood as a **code table** (3.2) and the set of rules relating to the code table. Each individual row in a code table is a **code element** (3.4) (e.g. “de - German - allemand - Deutsch” in the alpha-2 code table of ISO 639), while the item “de” is the **language identifier** (3.5).

The WG is in agreement of this usage of core terminology.

3.2
code table
table of **code elements** (3.4) as part of a **code** (3.1)

3.3
code space
totality of possible values for a set of identifiers within a **code** (3.1)

EXAMPLE All sequences of two letters (a–z) form the code space of the alpha-2 language code.

3.4
code element
individual entry in a **code** (3.1)

NOTE In ISO 639, each code element consists of a language identifier and the names of the language.

3.5
language identifier
language symbol
string of characters assigned to a linguistic entity for the purpose of uniquely representing it

NOTE 1 In ISO 639, each language identifier is composed of two, three, or four letters.

NOTE 2 See 4.1.

[1] Currently undergoing revision.

3.6
language
systematic use of sounds, characters, symbols or signs to express or communicate meaning or a message between humans

NOTE 1 This definition is intended to serve as a working definition for the purpose of ISO 639, not as a universal definition of this concept.

NOTE 2 See also 4.1 and 4.2.

3.7
individual language
language (3.6) that is distinctly different from another language

NOTE See 4.2.

3.8
dialect
language variant (3.14) specific to a geographical region or a group of language users

NOTE See 4.5.

3.9
macrolanguage
language (3.6) that for some purpose may be subdivided into two or more **individual languages** (3.7)

NOTE See 4.3.

3.10
language group
two or more **individual languages** (3.7) that for a specific purpose may suitably be treated as a unit

NOTE See 4.6.

3.11
language family
two or more **individual languages** (3.7) that are related to each other through having common ancestry

NOTE In exceptional cases a language family may have only one individual language as a member.

NOTE 2 This is for the purpose of ISO 639; not general linguistics. (Cf note to 3.6)

Could another term be used for this?

3.12
remainder group
language group (3.10) with the explicit exclusion of specified languages

NOTE See 4.6.

3.13
language variation
continuous variation within and between **individual languages** (3.7)

NOTE Language variation is seen and may be described as variation over time, space, cultural affiliation, etc.

3.14
language variant
variant of an **individual language** (3.7) that may be identified and named

3.15
standard variant
language variant (3.14) with a high degree of status and normalization

Merknad [HHj1]: The best word? “acceptance” instead.

NOTE A standard variant of a language may typically be used in official or public communication and in communication between users of different language variants.

3.16
writing system
system for writing a **language** (3.6), including the **script** (3.17) and character set used

NOTE See also 4.9.

Is “character set” defined somewhere? Needed here?

3.17
script
set of graphic characters used for the written form of one or more **languages** (3.6)

[ISO 15924 and ISO/IEC 10646]

NOTE See also 4.9.

3.18
orthography
set of rules for accepted spelling of words and text in one or more **languages** (3.6)

3.19
transcription
system for representing text in a different **script** (3.17) than the text originally was represented in

NOTE The resulting text is also referred to as a “transcription”.

Add reference?

3.20
transliteration
transcription (3.19) that enables the reconstruction of the original **script** (3.17) without any loss of information about graphic characters

NOTE The resulting text is also referred to as a “transliteration”.

3.21
written language
individual language (3.7) or **language variant** (3.14) that is commonly represented in writing with a relatively normalized **orthography** (3.18)

3.22
spoken language
individual language (3.7) or **language variant** (3.14) that is represented in spoken form

NOTE Any spoken language may be represented in writing using a phonetic writing system, where characters represent sounds (phones or phonemes) directly.

3.23
living language
individual language (3.7) or **language variant** (3.14) in present-day use, in particular as a **spoken language** (3.22)

3.24
extinct language
individual language (3.7) or **language variant** (3.14) that is no longer in use and that has no present-day descendant

NOTE See 4.7.

3.25
ancient language
extinct language (3.24) with a distinct literature and special status in the scholarly community

NOTE See 4.7.

Does an “ancient language” have to be “extinct”?

3.26
historical language
known earlier historical stage of a **living language** (3.23) or an **extinct language** (3.24)

EXAMPLE “Old English” and “Middle English” as historical stages of “English”.

NOTE See 4.7.

3.27
natural language
language (3.6) for human communication that is not an **artificial language** (3.28)

3.28
artificial language
language (3.6) for human communication that has been artificially devised

NOTE See also 4.8.

3.29
machine language
Define: NOT language

4 Fundamental concepts of language coding

Remove references to “parts” of 639.

4.1 Language identifiers and languages

Language identifiers are composed 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.

A language identifier represents a language, which may also be represented by one or more language names. The objects of identification are languages themselves; language names are the means by which the languages denoted by language identifiers are designated.

Languages are not static objects every instantiation of which is identical to every other. Every language corresponds to some range of variation in linguistic expression. In ISO 639 a language identifier denotes some range of language variation. The range of variation that is denoted can have three different scopes: individual language, macrolanguage or language group. Also, languages that are represented can be of various types~~:~~, e.g., living languages, ancient languages, artificially constructed languages~~, etc~~. The following provides further explanation regarding assignment of identifiers for different scopes or to different types of languages in ISO 639.

4.2 Individual languages

Identifiers in ISO 639 are assumed to denote individual languages, unless there is explicit reference in the language name or otherwise to a language group.

There is no one definition of a “language” that is agreed upon by all and appropriate for all purposes. As a result, there can be disagreement, even among speakers of the language or experts in linguistics, as to whether two variants represent dialects of a single language or two distinct languages. For ISO 639, judgments regarding when two variants are considered to be the same or different languages are based on a number of factors, including linguistic similarity, intelligibility, a common literature, the views of speakers concerning the relationship between language and identity, and other factors. The following basic criteria are followed:

— Two related variants are normally considered variants of the same language if speakers of each variant have inherent understanding of the other variant (that is, can understand based on knowledge of their own variant without needing to learn the other variant) at a functional level.

— Where spoken (or signing) intelligibility between variants is marginal, the existence of a common literature or of a common ethnolinguistic identity with a central variant that both understand can be strong indicators that they should nevertheless be considered variants of the same language.

— Where there is enough intelligibility between variants to enable communication, the existence of well-established distinct ethnolinguistic identities can be a strong indicator that they should nevertheless be considered to be different languages.

[How does sign language get into this? How are variations and “distinct differences” in sign languages realized?]

Some of the distinctions made on this basis may not be considered appropriate by some users or for certain applications. However, these basic criteria are thought to best fit the intended range of applications.

4.3 Macrolanguages

ISO 639 includes identifiers that correspond in a one-to-many manner with individual language identifiers. For instance, part 3 of ISO 639 contains over 30 identifiers designated as individual language identifiers for distinct variants of Arabic, while parts 1 and 2 each contain only one identifier for Arabic, “ar” and “ara” respectively, which are designated as individual language identifiers in those parts of ISO 639. It is assumed here that the single identifiers for Arabic in parts 1 and 2 of ISO 639 correspond to the many identifiers collectively for distinct variants of Arabic in part 3 of ISO 639.

In this example, it may appear that the single identifiers in parts 1 and 2 of ISO 639 should be designated as collective language identifiers. That is not assumed, however. In various parts of the world, there are clusters of closely-related language variants that, based on the criteria discussed in 4.2, can be considered individual languages, yet in certain usage contexts a single language identity for all is needed. Typical situations in which this need can occur include the following:

— There is one variant that is more developed and that tends to be used for wider communication by speakers of various closely-related languages; as a result, there is a perceived common linguistic identity across these languages. For instance, there are several distinct spoken Arabic languages, but Standard Arabic is generally used in business and media across all of these communities, and is also an important aspect of a shared ethno-religious unity. As a result, a perceived common linguistic identity exists.

— There is a common written form used for multiple closely-related languages. For instance, multiple Chinese languages share a common written form.

— There is a transitional sociolinguistic situation in which sub-communities of a single language community are diverging, creating a need for some purposes to recognise distinct languages while, for other purposes, a single common identity is still valid. For instance, in some business contexts it is necessary to make a distinction between the languages Bosnian, Croatian, and Serbian; yet there are other contexts in which these distinctions are not discernable in language resources that are in use.

Merknad [HHj2]: Check!

Where such situations exist, an identifier for the single, common language identity is considered to be a macrolanguage identifier.

Macrolanguages are distinguished from language groups in that the individual languages that correspond to a macrolanguage must be very closely related, and there must be some domain in which only a single language identity is recognized.

4.4 Linguistic norm

[Remove most of the “etc.”s]

Some linguistic forms are “normalized” or “standardized” by official or private bodies like academies or language councils. This normalization may be applied to any elements: orthography, morphology, syntax, semantics, phonology, etc. The degree of normalization varies greatly from one language to another.

Languages or forms of languages may be considered to have some sort of official status within countries or regions. Language status may be established through acts of parliament or through other formal procedures, giving a language status as “national language”, “official language”, “regional language”, etc.

4.5 Dialects

The linguistic variants denoted by each of the identifiers in parts 1, 2, and 3 of ISO 639 are assumed to be distinct languages and not dialects of other languages, even though for some purposes some users may consider a variant listed in part 1 or 2, or in particular part 3 of ISO 639 to be a “dialect” rather than a “language” (see 4.2 and 4.3). In this International Standard, the term *dialect* is used as in the field of linguistics where it identifies any sub-variant of a language such as might be based on geographic region, age, gender, social class, time period, etc.

The dialects of a language are included within the denotation represented by the identifier for that language. Thus, each language identifier represents the complete range of all the spoken or written variants of that language, including any standardized form.

For applications in which it is necessary to identify dialects, a separate standard may be developed that provides identifiers for dialects, or that combines identifiers from ISO 639 with other distinguishing identificational qualifiers.

4.6 Collective language code elements and language groups

Merknad [HHj3]: Clarify the notions of “groups” and “families” and “collections”. Think about this!

Part 2 of ISO 639 includes alpha-3 identifiers for collections of languages, and part 5 is especially dedicated to language groups and language families. Parts 1 and 3 provide identifiers for individual languages and macrolanguages only.

Some of the code elements in part 2 specify remainder groups. These items have the word “other” in their English names (and “autres” in their French names). The same alpha-3 identifiers are included in part 5 of ISO 639 denoting the language group or family as a whole. For applications using part 2 only, the remainder group identifiers shall be used for languages that belong to the language group or family in question, but that do not have an individual-language identifier in part 2. Applications using parts 2 *and* 5 and applications using parts 2 *and* 3 *and* 5 shall use the collective language code elements in the sense specified in part 5, allowing for hierarchies of language identifiers.

4.7 Extinct, ancient and historic languages

ISO 639 includes identifiers that denote extinct languages as well as living languages. In order to qualify for inclusion in ISO 639, the language must have an attested literature or be well-documented as a language known to have been spoken by some particular community at some point in history; it may not be a reconstructed language inferred from historical-comparative analysis. The code also includes identifiers that denote historic languages that are considered to be distinct from any modern languages that may be descended from them; for instance, Old English and Middle English. Here, too, the criterion is that the language has a literature that is treated distinctly by the scholarly community.

4.8 Artificial languages

ISO 639 includes identifiers that denote artificial (or constructed) languages that meet the following criteria:

— the language has a literature; and

— the language is designed for the purpose of human communication.

Specifically excluded from ISO 639 are reconstructed languages, computer programming languages, and mark-up languages.

4.9 Writing systems and scripts

A single language identifier is provided for a language even though the language may be written in more than one writing system. ISO 639 language identifiers may be combined with script identifiers from ISO 15924 (see 8.3).

5 Description of the ISO 639 database and code

5.1 The ISO 639 database

The ISO 639 database is an integral part of the ISO Concept Database, which is accessible at http://cdb.iso.org/.

[Further description needed.]

5.2 ISO 639 as one code space

All language identifiers that are specified in this International Standard share a single code space. This has the implication that one language identifier always denotes one specific item.

All alpha-2 language identifiers of this International Standard, furthermore, denote exactly the same items as corresponding alpha-3 language identifiers with the same reference.

5.3 Subsets of ISO 639

The ISO 639 code table is one coordinated set. [The following description needs to be revised.]

[Give a simple example of subsets that can be defined, e.g. “European languages”. We need to merge 5.3, 5.4 and 5.6. The last sentence of 5.6 should be the first sentence of the merged 5.3&5.6&5.4.

— Alpha-2 identifiers are assigned to a limited number of languages (currently 186), in particular languages with a long-standing scientific literature and developed terminology and lexicography.

— Part 2 assigns alpha-3 identifiers to a somewhat larger number of languages (currently 484), in particular languages with a significant body of literature in central libraries and documentation holdings.

— Part 3 assigns alpha-3 identifiers to most individual languages of the world (currently about 7000).

— Part 5 assigns alpha-3 identifiers to language groups and language families.

— Part 6, which is currently under development, enables the encoding of items on a more detailed level than “individual language”.

5.4 Principles

The following principles apply to each of the parts of ISO 639 and their interrelation: [The following description needs to be revised.]

— The set of languages included in part 1 is a subset of the set of languages included in part 2. The extension of any single item in part 1 is exactly the same as the extension of the corresponding item in part 2. An alpha-3 identifier in part 2 and the corresponding alpha-2 identifier in part 1 shall be considered synonyms. In cases where the “B table” and the “T table” of part 2 have different alpha-3 identifiers, these identifiers shall be considered synonyms. Examples: (1) The language identifiers “en” and “eng” designate exactly the same language. (2) The language identifiers “fra” and “fre” (and “fr”) designate exactly the same language. Implementations should, whenever possible, allow free choice between such synonyms.

— Some items are included in part 2 and in part 5 with slightly different designations. Example: “gem” designates “Germanic (other)” in part 2 and “Germanic languages” in part 5. In the context of part 2, “gem” shall be used to identify the set of languages that fall into the group “Germanic languages”, and that has no individual language identifier in part 2. In other contexts “gem” may be used to identify the language group “Germanic languages” as a linguistic entity.

— The alpha-3 code of parts 2, 3, and 5 forms one single code space. No alpha-3 identifier assigned in any one of these parts has been assigned or will be assigned to another item.

— Code elements of part 3 may also be included in part 2 following the procedures of the Joint Advisory Committee (see 6.2).

5.5 Applications

ISO 639 and its parts may be implemented in a variety of applications. It is expected that few or none of these applications will utilize the totality of the language codes of all parts of this International Standard. As part of the implementation process it may be needed to specify usage relating to some of the issues that are listed in 7.2 and 7.3.

Some types of applications are listed below. The numbers of this list are referenced in 7.2 and 7.3 in the format “7.1: 1”. These references are intended for guidance only.

[Go into more detail as to “recommendations relating to the application of ISO 639”.]

Language identifiers from the various parts of ISO 639 may be used in connection with, e.g.:

[Input to this list of use cases is very welcome. The list should be moved to the Introduction.]

1) multilingual terminological or lexicographical databases to identify the language of an individual piece of information, e.g. a word, a term, a context, or a definition;

2) a text document or a set of text documents to identify the language of the entire text or of text segments, e.g. quotations;

3) bibliographical documents or databases (in general) to identify the language(s) of the bibliographic entries or the language(s) of the referenced documents;

4) linguistics and bibliographical documents or databases of linguistic or lexicographical documents to identify the language(s) that are the object of description in the documents (e.g. “source language”, “target language”, “description language”, “described language”, etc.);

5) translated documents to identify the source language for the translation;

6) translation and interpretation services to identify languages covered by the service;

7) notes or minutes of meetings to identify language(s) actually used during the meeting;

8) registries of individuals or organizations to identify language proficiencies or preferences;

9) software to identify language capabilities of e.g. character set handling, built-in grammar control, and dictionaries; and

10) localization in general, comprising a number of the types of applications above, but being in itself a major user of ISO 639; see also 8.

A well defined and much used implementation of the code tables of ISO 639 is designed by the Internet Engineering Task Force (IETF, see http://www.ietf.org/). Its RFC 4646 (which has replaced RFC 3066 and RFC 1766) defines the use of ISO 639 alpha-2 and alpha-3 language identifiers in combination with other information elements to identify the language of documents and text segments. [Add mention of W3C and Unicode.]

5.6 Subsets of the code table

The specification of any implementation of ISO 639 shall include information about which subset of the total ISO 639 code tables that is used. Some of the recommended options are: [Revise the description of this.]

— part 1 only (7.1: 1–9, if all languages in question are included in part 1);

— part 1 and part 2, using part 2 only for items not included in part 1 (or a user-defined subset of part 1);

— part 2 only (7.1: 3);

— part 2 and part 3;

— part 2 and part 5, using items included in both parts in the sense specified in part 5 (see also 7.3) (7.1: 4);

— part 1, part 2, and part 5, using items included in both parts 2 and 5 in the sense specified in part 5 (see also 7.3) (7.1: 4);

— part 2, part 3, and part 5, using items included in both parts 2 and 5 in the sense specified in part 5 (see also 7.3); and

— a user-defined subset of any of the parts or any combination of parts.

Part 5 is expected to be used only in combination with other parts of ISO 639.

It is expected that a mechanism will be developed in the future for naming and registering defined subsets of the totality of the ISO 639 language code.

5.7 Language groups [This needs to be moved to 4.6]

As discussed in 5.3, second list item, the simultaneous application of part 2 and part 5 of ISO 639 will require implementation-level specification.

There are currently 64 items that are included both in part 2 and in part 5 (listed in annex A of part 5). Of these items, 29 items are identical in the two parts (e.g. “alg – Algonquian languages”). The remaining 35 items are intended to cover remainder groups in part 2 and entire language groups in part 5 (e.g. “afa – Afro-Asiatic (Other)” in part 2 and “afa – Afro-Asiatic languages” in part 5).

According to the principles of part 2, the identifier “afa” will be assigned only to a document or information in (or about) an Afro-Asiatic language that does not have an individual-language identifier in part 2, and that does not fall into the remainder groups “ber – Berber (Other)”, “cus – Cushitic (Other)”, or “sem – Semitic (Other)”, all of which are Afro-Asiatic language groups.

According to the principles of part 5, the identifier “afa” may be assigned to a document or information in (or about) any Afro-Asiatic language. The use of “afa”, “sem”, or “ara” in a concrete case relating to Arabic, depends on the purpose of the encoding, as specified in the implementation.

The use of identifiers from part 5 will depend on the purpose of the application. It is expected that user-defined subsets of the items in part 5 will frequently be used in combination with, e.g., the totality or defined subsets of part 2 or part 3.

6 Language description format (LDF)

[Needs to be in line with 12620]

6.1 Compatibilities between the ISO 639 model and ISO 12620

6.1.1 General

The model for ISO 639 has been developed to be compatible with models being developed by other groups within ISO/TC 37. ISO/TC 37 standards for computational use of terminology, specifically ISO 16642 and its combination with ISO 12620, emphasise the use of a metamodel in combination with metadata identifiers, referred to as data categories. These data categories may be referred to also as administered items, in accordance with ISO/IEC 11179.

ISO 639 uses a specific model for language identification/documentation and a list of metadata identifiers can be associated with this model. The model for ISO 639 has been developed to be in conformity with the ISO/IEC 11179 series of standards. As such, language information is:

— specified according to ISO/IEC 11179-3;

— defined according to ISO/IEC 11179-4;

— named according to ISO/IEC 11179-5; and

— registered according to ISO/IEC 11179-6.

It is intended to be fully compatible with the metadata registry specified in ISO 12620. It is also intended to be fully compatible with the Data Category Interchange Format (DCIF) defined in ISO 12620. Some variation between the metamodel of the metadata registry and the specific model described in this International Standard has been unavoidable, but the core of the model shall be as consistent with ISO 12620 as possible.

The identifiers and associated data shall be managed within a metadata registry conforming with ISO 11179-6.

The metamodel of ISO 12620 is applied within the scope of ISO 639.

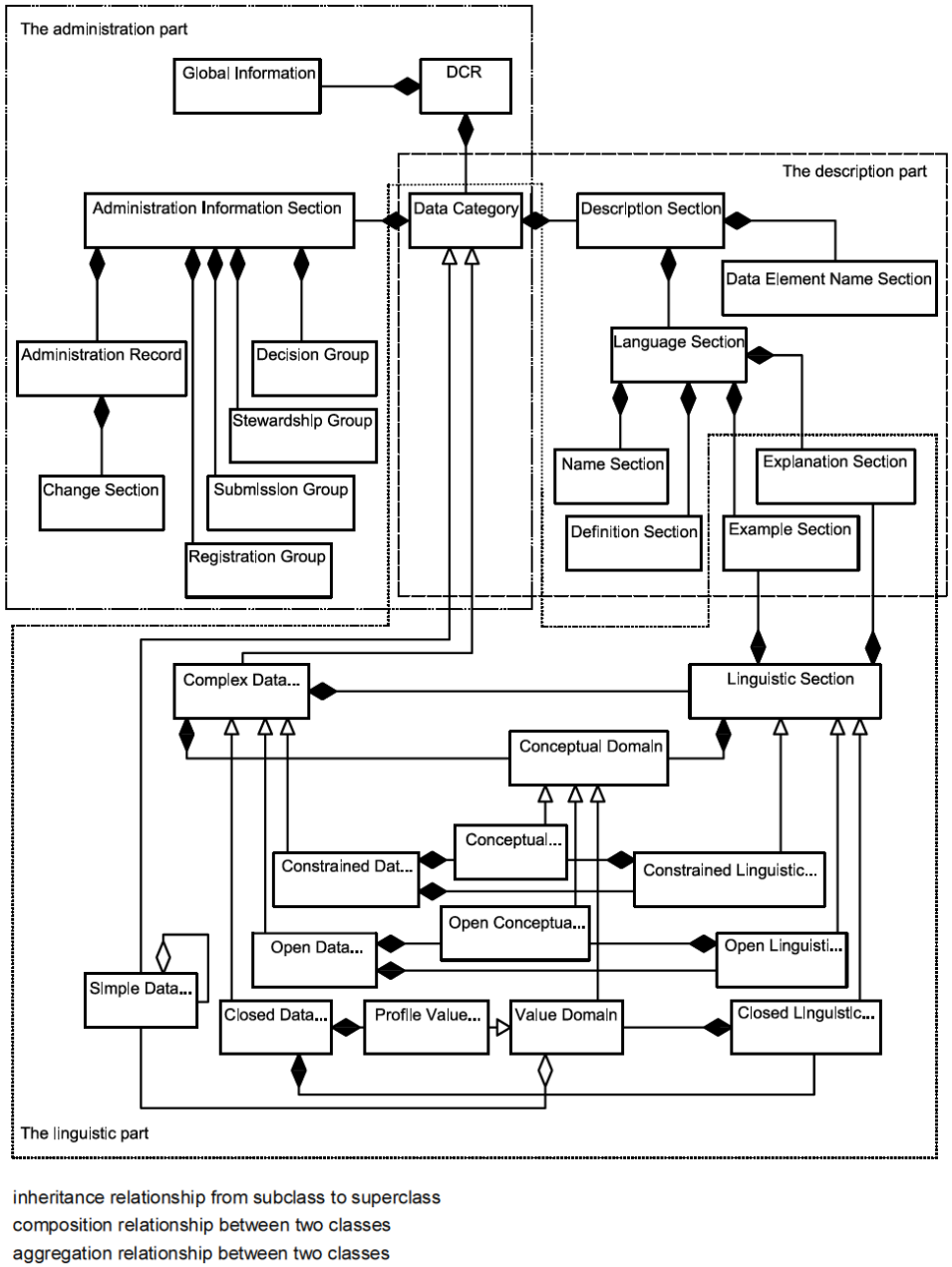

Figure 1 — Overview of the metamodel of ISO 12620 underlying the DCIF

6.1.2 Identification

Each item within the language code is provided with one reference name. This reference name is intended for use as the unique data identifier (DI), to be used in combination with a version identifier (VI) and registration

authority identifier (RAI) for composing the international registration data identifier (IRDI), in accordance with ISO/IEC 11179. The combination created by the IRDI shall provide a unique identifier.

According to ISO 12620, identification occurs within the **Administration Record** section of the **Administration Identification** section.

Identification, according to this scheme, should be an implemented form of:

/identifier/ = Western Apache
/registration authority/ = SIL
/version/ = 1

/identifier/ uniquely identifies the data category in the registry. This field is annotated by information about the registration authority (/registration authority/) and the version of the identifier (/version/). Uniqueness is a condition within the combination; multiple identifiers using the same /identifier/ value are possible if either the registration authority or the version differ, e.g.

/identifier/ = Western Apache
/registration authority/ = SIL
/version/ = 2

and

/identifier/ = Western Apache
/registration authority/ = LOC
/version/ = 1

are both different from the first example. This prevents the potential for collision of differently described identifiers upon import.

For each registered item, the /registration status/ data category shall be provided with a value. The list of possible values for this identifier differs between ISO/IEC 11179-6 and ISO 12620.

From ISO 12620, the listed registration status values are:

— /standard/: the Registration Authority confirms that the administered item is of sufficient quality and of broad interest for use in the Registry community *(2)*;

— /candidate/: it has been proposed for progression up the Registry registration levels; note: the registration status of a data category is set to /candidate/ until its administration status is finally determined as /accepted/ (in which case the registration status becomes /standard/) *(6)*;

— /deprecated/: the Registration Authority has approved the administered item as no longer recommended for use in the registry community and this item should no longer be used;

— /superseded/: the Registration Authority has approved the administered item as no longer recommended for use in the registry community but the successor administered item is the preference for users *(8)*.

From ISO/IEC 11179-6, further values for registration status are:

— /preferred standard/: the Registration Authority confirms that the administered item is preferred for use within the Registry community *(1)*;

— /qualified/: the Registration Authority has confirmed that the mandatory metadata attributes are complete and conform to applicable quality requirements *(3)*;

— /recorded/: the Registration Authority has confirmed that all mandatory metadata attributes have been completed *(4)*;

— /incomplete/: the submitter wishes to make the community that uses this Registry aware of the existence of this item *(6)*;

— /retired/: the Registration Authority has determined that this item is no longer recommended for use in the community that uses this Registry and the item should no longer be used *(7)*.

NOTE Numbers in *( )* are the order in 11179-6. There is obvious progression between some of these.

/deprecated/ appears in the list from ISO 12620 but does not appear in the list from ISO 11179-6; it should be replaced with /retired/.

The following identifiers should also be populated, though it is the choice of each implementing system whether these fields are filled manually or automatically:

— /creation date/: the date when the data category was first created (for instance in an expert's working space or private area);

— /effective date/: "the date an administered item became/becomes available to registry users" (ISO/IEC 11179-3); the date on which a value for registration status was assigned;

— /last change date/: the date of the last modification of information about a data category (interdependent with change description);

— /change description/: free text description of the modification undergone by the data category (e.g. "definition updated …");

— /explanatory comment/: descriptive comments about the data category;

— /origin/: source (document, project, discipline or model) for the data category;

— /unresolved issue/: problem that remains unresolved regarding proper documentation of the data category;

— /until date/: the date a data category is no longer effective in the registry; note: this information is set when the registration status of the data category changes to /retired/ or /superseded/.

The values within source should be produced in conformity with an appropriate standard. Free text fields such as explanatory comment and unresolved issue require standardization consideration. All date information shall be provided in conformity with ISO 8601.

Further documentary information may be provided by each Registration Authority for a data category within the Registration Group, Submission Group, Stewardship Group and Decision Group sections. It is the responsibility of the specific Registration Authority to document the information provided within these sections. As ISO 12620 does not impose any constraints on these sections, data interoperability and consistency of description between registration authorities that involve these sections is not guaranteed. In general, these sections will be considered as empty for the purposes of interchange.

6.1.3 Description of an ISO 639 language identifier

The Description section of a data category (ISO 12620) provides the capability for a single item to have multiple names, with each name organized according to the language within which it is used (see Figure 2). One or more names may be given in one or more languages; however, there must be at least one name. In ISO 12620 the analogy is drawn with the terminological metamodel (ISO 16642) where a concept has multiple terms organized by the language in which they are used.

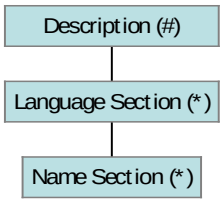

Figure 2 — Metamodel of ISO 12620 underlying the DCIF

A simple example of this multiplicity is provided for data from ISO 639-1/-2. For information about languages identified by ISO 639-2 **gla** or ISO 639-1 **gd**, within **Description**, we have:

[**Language Section**] /language/ = /eng/
[**Name Section**] /name/ = Gaelic
[**Name Section**] /name/ = Scottish Gaelic

[**Language Section**] /language/ = /fra/
[**Name Section**] /name/ = gaélique
[**Name Section**] /name/ = gaélique écossais

For the purposes of this standard, the item being described, i.e. the language, is named in one or more languages but does not need to be named in its own language.

/broader data category/ can be used within **Description** to associate language data categories with more broadly descriptive language data categories. An example is taken from ISO 639-5 for part of the expansion of the Indo-European languages involving the West Germanic language family.

Alpha-3	Parent alpha-3	English	French
ine		Indo-European languages	indo-européennes, langues
gem	ine	Germanic languages	germaniques, langues
eng	gem	West Germanic	germanique occidental

The data is expanded to the description shown below in which each section is identified in square brackets in bold:

[**Data Category**]
[**Administration Identification**]
[**Administration Record**] /identifier/ = West Germanic
[**Description**] /broader data category = /Germanic languages/
[**Language Section**] /language/ = /eng/
[**Name Section**] /name/ = West Germanic
[**Language Section**] /language/ = /fra/
[**Name Section**] /name/ = germanique occidental

[**Data Category**]
[**Administration Identification**]
[**Administration Record**] /identifier/ = Germanic languages
[**Description**] /broader data category = /Indo-European languages/
[**Language Section**] /language/ = /eng/
[**Name Section**] /name/ = Germanic languages
[**Language Section**] /language/ = /fra/
[**Name Section**] /name/ = germaniques, langues

[**Data Category**]
[**Administration Identification**]
[**Administration Record**] /identifier/ = Indo-European languages
[**Description**]
[**Language Section**] /language/ = /eng/
[**Name Section**] /name/ = Indo-European languages
[**Language Section**] /language/ = /fra/
[**Name Section**] /name/ = indo-européennes, langues

This description can be implemented within the DCIF. Further descriptive information should be added to the Description section including:

— /definition/: should be used to provide a definition in the data category registry. As far as possible, the definition should be language and theory neutral. This information is mandatory for each DEC. It may be repeated to provide translations of the definition in other working languages. When necessary, /definition/ may be refined by a /source/ and a /note/;

— /explanation/: can be used to provide additional information about the data category that would not be relevant for a definition (e.g. more precise linguistic background for the use of the data category);

— /example/: the use of examples should be limited to those that illustrate the data category in general, excluding language specific usages, which should be documented at Object language level;

— /source/: may refine /definition/, /explanation/, or /example/ to indicate the source from which the corresponding text has been borrowed or adapted. When a definition is compiled from more then one source, this field can be repeated. The /source/ field should not be used alone in the Description section;

— /profile/: shall be identified as Language description;

— /conceptual domain/: since language identifiers are considered to be simple data categories, there are no possible values for conceptual domain;

— /note/: additional information associated with the Description section, excluding technical information that would normally be described within /explanation/;

— /broader data category/: shall be used to refer to a more encompassing language identifier. This mechanism can be used to cross-refer between identifiers across arbitrary boundaries, for example to make the familial link from **English** (**eng** or **en**) to **West Germanic** (**gmw**).

The last description above, with integration across administrative boundaries, presents one minor difficulty: cross-reference across systems indicates use of the unique identifier from the combination of RA:ID:Ver. This gives 2 alternatives:

[**Data Category**]
[**Administration Identification**]
[**Administration Record**] /identifier/ = English
[**Description**] /broader data category = /**West Germanic**/
[**Language Section**] /language/ = /eng/
[**Name Section**] /name/ = West Germanic
[**Language Section**] /language/ = /fra/
[**Name Section**] /name/ = germanique occidental

or

[**Data Category**]
[**Administration Identification**]
[**Administration Record**] /identifier/ = English
[**Description**] /broader data category = /**639-5.West Germanic.1**/
[**Language Section**] /language/ = /eng/
[**Name Section**] /name/ = West Germanic
[**Language Section**] /language/ = /fra/
[**Name Section**] /name/ = germanique occidental

[**Data Category**]
[**Administration Identification**]
[**Administration Record**] /identifier/ = West Germanic
[**Description**] /broader data category = /Germanic languages/
[**Language Section**] /language/ = /eng/
[**Name Section**] /name/ = West Germanic
[**Language Section**] /language/ = /fra/
[**Name Section**] /name/ = germanique occidental

This description can be implemented within the DCIF.

In a **Description section**, the Language section level is likely to contain repeated entries as demonstrated above. The following data categories will be used in this section:

— /language/: shall be used to identify the language being described (i.e. *object language*, as defined in ISO 16642). Values for this data category shall be those of ISO 639 such that an identifier could be described potentially using any kind of media;

— /definition/: to define the data category when it occurs in a specific system within a language, so that it impacts on the accuracy of the reference definition;

— /example/: provides an example of how the data category is used for the current object language;

— /explanation/: additional explanation specific to the use of the data category in the object language;

— /source/: see Description section;

— /conceptual domain/: since language identifiers are considered to be simple data categories, there are no possible values for conceptual domain;

— /note/: additional information associated with the Object language level, excluding technical information that would normally be described within /explanation/.

The **Name section** shall be used to record an appellation for the data category in the object language elicited at Language section level. The Name section may be repeated within a Language section as demonstrated above. The following descriptive elements are associated with the Name section level:

— /name/: one word or multi-word unit used to refer to the data category for the corresponding object language as expressed in the encompassing language section. Names given to a data category shall not be used for the purpose of identifying a data category (see /identifier/);

— /name status/: with the following conceptual domain: {/standardized name/, /preferred name/, /admitted name/, /deprecated name/, /superseded name/} (taken as such from ISO/IEC 11179).

6.2 Extensions to ISO 12620 for ISO 639 LDF

6.2.1 Representation

Within ISO 12620 there is no provision for non-linguistic identification/representation, which is essential for LDF. The ISO 639 language identifiers shall be considered as representations in line with ISO/IEC 11179. The

first expansion allows for the alpha-2 and alpha-3 representations and is associated with the Description section. Representations are the permitted values a data category may use. Names are language-dependent representations; ISO 639 provides language independent representations also, and specific description of these does not appear to be catered for fully in either ISO/IEC 11179 or ISO 12620.

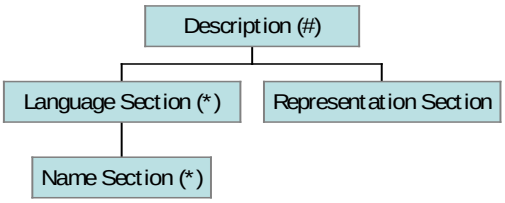

Figure 3 — Naming and representation for ISO 639

Multiple representations are available in ISO 639, including the terminological and bibliographical varieties of ISO 639-2 alpha-3s and, for a number of the ISO 639-2 alpha-3s there are equivalent ISO 639-1 alpha-2s. For ISO 639, each of these varieties of representation are identified by the following data categories:

— /representation term/: alpha-2 or alpha-3 language identifier;

— /representation source/: refinement on /representation term/ that contains one of the values /iso639-1/, /iso639-2/, /iso639-3/, /iso639-5/.

[**Data Category**]
[**Administration Identification**]
[**Administration Record**] /identifier/ = English
[**Description**] /broader data category = /**West Germanic**/
[**Language Section**] /language/ = /en/
[**Name Section**] /name/ = West Germanic
[**Language Section**] /language/ = /fr/
[**Name Section**] /name/ = germanique occidental
[**Representation Section**] /representation term/ = en; /representation source/ = /iso639-1/
Representation Section] /representation term/ = eng; /representation source/ = /iso639-2/

From the representations, further ISO/IEC 11179 metadata identification is possible, for example the enumeration of value domains in the conceptual domain:

/conceptual domain name/ = Languages of the world
/conceptual domain definition/ = Lists of languages of the world represented as names or codes

/value domain name/ = language codes – 2 character alpha/
/permissible values/ = /en/, /fr/, ….

/value domain name/ = language codes – 3 character alpha/
/permissible values/ = /eng/, /fre/, ….

The variety of information about names and representations enables the computational construction of metadata hierarchies from more generic to more specific language identifiers across the series of the ISO 639 standards. The instances of /broader data category/ can be followed, for example, to construct a (fragment) sub-tree of identifiers related to Manx.

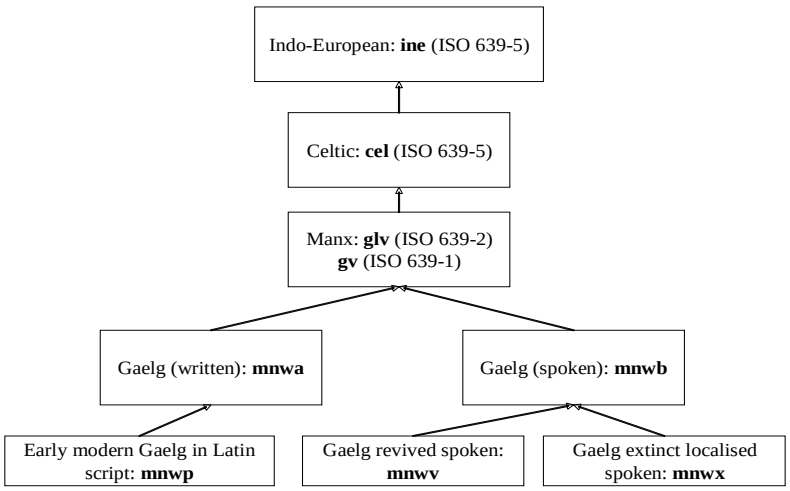

Figure 4 — A (fragment) sub-tree of identifiers related to Manx

6.2.2 Documentation

Beyond the documentation of names and representations, ISO 639 Registration Authorities and other language documentation projects should provide further documentary information for each language data category. Since this information is not catered for in ISO 12620 and for ease of interchange, a Documentation section in LDF extends the Description section:

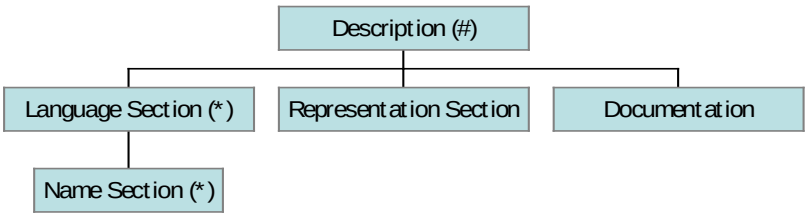

Figure 5 — Documentation section (1)

The Documentation section is subdivided into further sections:

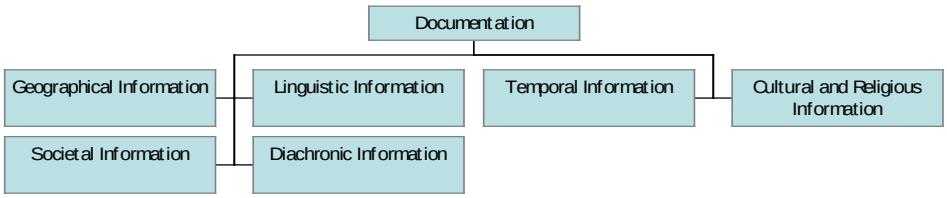

Figure 6 — Documentation section (2)

The following lists comprise metadata descriptors used for documenting languages. These descriptors are intended for capturing information that assists in the identification, provenance and monitoring of quality in a language metadata registry. These descriptors will be used to help avoid unnecessary variations when describing highly similar objects within the registry. The development of the metadata registry may result in the addition of further meta-descriptors, and the registration authorities for ISO 639 shall document such additions where it is essential for interoperability within the ISO 639 series and with respect to the ISO 12620 DCR.

Geographical information includes:

— placeholder for link to GPS information, e.g. metadata as described with ISO 19115 using lat/long coordinate system;

— toponym;

— UN region or ISO 3166 identifier.

Linguistic information includes:

— mode of communication, e.g. spoken or written or signed;

— writing system;

— script (including ISO 15924 identifiers, where available);

— bilingual (in “language A” and “language B”);

— status, e.g. standard or historic or extinct, etc.;

— language family.

Temporal information includes:

— historical note;

— modern events and changes;

— date.

Diachronic information for recording changes between versions includes:

— historic class;

— historical classification.

Cultural and religious information includes:

— community;

— religious culture.

Societal information includes:

— population size;

— social status;

— legal status;

— speaker identification;

— migration;

— census data;

— age information;

— health information.

6.3 Language information

Language information that can be computed, includes:

— number of living languages;

— number of languages that have become extinct since a certain date;

— number of languages that are used only as a second language;

— number of speakers by language family;

— number of speakers by country;

— language counts, e.g. number of languages indigenous to a country.

Language naming information includes:

— language name, which includes all the language names, e.g. alternate names, dialect names, and alternate dialect names that appear in the language entries;

— language name status, e.g. preferred, admitted, offensive, deprecated.

Country information includes:

— official country name, i.e. name used by the country in its official documents;

— country population, i.e. approximate number of people in a country;

— geographic reference;

— national language, i.e. language spoken by a large portion of the population of a nation;

— official language, i.e. language designated by an official body;

— country literacy rates, i.e. estimate of the percentage of the population in the country that is literate in some language;

— non-indigenous language: language spoken in one country by a community that has migrated from another country where there is no significant dialect difference between the two locations.

“Country-dependent” language information includes:

— country of origin of a language, normally its primary country;

— country where most of the speakers of a language are located;

— non-primary country in relation to the use of a language;

— country speaker population, i.e. estimated number of first-language speakers in a country;

— non-speaker population;

— total speaker population;

— monolingual population, i.e. number of people who are monolingual in a language;

— blind population, i.e. population of a country that are reportedly blind;

— deaf population, i.e. population of a country that are reportedly deaf;

— population remarks, i.e. additional information that may include population breakdowns by *dialect*, *gender*, or *ethnic groups*; the *population of the deaf community*; or other comments related to population;

— population in all countries;

— ethnic population, i.e. population of those who identify themselves as part of the ethnic group, regardless of whether they speak the language.

Reference material information includes:

— source (“literature”);

— dictionary;

— grammar;

— broadcast media;

— braille literature.

Geographical information includes:

— geological and topographical information, e.g. altitude range;

— ecological information.

Sociocultural information includes:

— religion, i.e. religious affiliation of people.

Linguistic features information includes:

— linguistic typology, e.g. constituent order, syllable patterns, or other features of linguistic interest;

— constituent order, e.g. “Subject, Object, Verb = SOV”;

— syllable patterns, e.g. “Consonant, Vowel, Consonant = CVC”.

7 Combining language identifiers with other standards and codes

[Combine this section with the Applications section in 5.]

7.1 Combining ISO 639 with ISO 3166

The language identifiers of ISO 639 may be combined with country and country subdivision identifiers of ISO 3166 (all parts) to denote the area in which a word, term, phrase, or language variant is (or has been) used.

NOTE 1 In ISO 3166 the term “code element” is used to refer to the concept of “identifier” according to ISO 639 terminology.

NOTE 2 Some applications may not allow the use of the country subdivision code of ISO 3166-2, because of the variable format of that code.

Examples:

— “eng US” (or “en US”, “eng USA”, “en USA”, “eng 840”, “en 840”) indicates English of the United States of America;

— “eng US-NY” indicates English of the state of New York;

— “fra FR” (or “fre FR”, “fr FR”, “fra FRA”, “fre FRA”, “fr FRA” ”, “fra 250”, “fre 250”, “fr 250”) indicates French of France;

— “fra FR-75” indicates French of Paris.

NOTE 3 Applications may define a default region for each language and use ISO 3166 identifiers to specify usage outside this region only.

7.2 Combining ISO 639 with ISO 19111 and ISO 19112

The language identifiers of ISO 639 may also be combined with spatial referencing information in accordance with ISO 19111 and ISO 19112 to denote the area in which a word, term, phrase, language, or language variant is used.

ISO 19111 and ISO 19112 complement ISO 3166 in that they allow spatial referencing independent of political and administrative considerations.

7.3 Combining ISO 639 with ISO 15924

The language identifiers of ISO 639 may be combined with script identifiers of ISO 15924 to indicate which script is used in a document, text segment, language, or language variant.

Examples:

— “deu Latf” (or “ger Latf” or “de Latf”) indicates German in Latin Fraktur script;

— “kur Cyrl” (or “ku Cyrl”) indicates Kurdish in Cyrillic script.

NOTE Applications may define a default script for each language and use ISO 15924 identifiers to specify the use of scripts other than the default.

7.4 Other code combinations

The language identifiers of ISO 639 may be combined with any other standardized or user-defined code to establish combined identifiers suitable for given purposes.

The usage of such combined identifiers and combination codes shall be documented in each individual case. The intension of the individual language identifiers and the ISO 639 language code shall remain unchanged by such combinations.

7.5 Formats of combined identifiers

This International Standard does not require a specific format of combined identifiers. The format used in ~~8~~7.1 and ~~8~~7.2 is intended as an example only.

Each application shall specify the format of combined identifiers. The specification may include one or more of the following:

— the order of the elements,

— a separation character between elements,

— prefixes or other indicators to some or all of the elements,

— structuring features, such as XML tagging.

Depending on the specification of the application, a combined language identifier for a text in German in Latin Fraktur script, originating from Austria, could for instance be encoded in one of the following ways:

— de Latf AT

— de_Latf_AT

— de_AT_Latf

— de-C:AT-S:Latf

— language="de" script="Latf" area="AT"

Annex A
(~~informative~~normative)

Governance and maintenance of ISO 639

A.1 General principles

ISO 639 is maintained according to the rules and principles of “Standards as databases” as specified in Annex ST (normative) to the ISO Supplement to the ISO/IEC Directives: “Procedure for the development and maintenance of standards in database format”.

[Expand on this annex. Compare with ISO/TC 184/SC 4 text.]

A.2 ISO 639 Validation team

Each ISO/TC 37 National Member Body and liaison organization (according to the rules that are specified in Annex ST to the Directives) can nominate one member of the ISO 639 Validation team (VT).

A.3 ISO 639 Secretariat

ISO/TC 37 appoints a secretariat.

A.4 ISO 639 Maintenance team

An ISO 639 Maintenance team (MT) is appointed.

[Who actually appoints? How many? Rules and principles?]

A.5 Change requests

Anyone may submit change requests (CR) to ISO 639 Secretariat. Change requests may relate to the addition, deletion or change of any language identifier, language name, or any other information in the ISO 639 database. Change requests shall be supported by justifications.

ISO 639 Secretariat shall consult members of the ISO 639 Maintenance team as relevant depending on the nature of the Change request. Following such consultation the Secretariat shall submit to the ISO 639 Validation team a proposed decision.

The ISO 639 Validation team decides on the Change request.

[More specific rules? Timeliness rules (beyond what is specified in Annex ST)?]

A.6 Creation of language identifiers

The creation of language identifiers is subject to the rules in 4.1.

A.7 Stability of the code table

Every effort shall be made to maintain stability of the ISO 639 code table. Language identifiers that have been assigned shall not be withdrawn, but may in exceptional cases be deprecated.

Given the fact that languages and geopolitical situations change over time it may occasionally be advisable to modify language names and the exact extension of an encoded item. All such modifications shall be done with great care so as to ensure continuous validity of data repositories.

New items will be added to the code table, as well as to defined subsets of the code table. All such additions shall be well documented. Every effort shall be made to minimize the impact that additions to the code table have on other items in the table.

Annex B
(informative)

Transition rules

B.1 General

This annex is expected to be removed from this document prior to its finalization as an International Standard.

The annex explains the previous governance structure and specifies how the transition from this structure to the new governance structure as specified in Annex A shall be put in place.

B.2 Registration Authorities and responsibilities for parts of the previous ISO 639

B.2.1 Current situation

Separate Registration Authorities (RA) have been nominated for parts 1, 2, and 3 of the previous version of ISO 639:

- ISO 639-1:2002 RA: The International Information Centre for Terminology (Infoterm), Vienna, Austria.
- ISO 639-2:1998 RA: The Library of Congress, Washington DC, USA.
- ISO 639-3:2007 RA: The Summer Institute of Linguistics, Inc. (SIL International), Dallas, Texas, USA.
- ISO 639-5:2008 RA: ISO 639-2 Registration Authority has also been responsible for the maintenance of ISO 639-5:2008.
- ISO 639-6:2009 RA: GeoLang Ltd, Wales, United Kingdom.

B.2.2 Transition rules

All the designated ISO 639 Registration Authorities cease to function as Registration Authorities at the latest at the time of publication of this International Standard.

B.3 Joint Advisory Committee ISO 639/RA-JAC

B.3.1 Current situation

A Joint Advisory Committee (ISO 639/RA-JAC) is established to advise the ISO 639 Registration Authorities. It shall guide the application of the coding rules as laid down in ISO 639-1, ISO 639-2, and ISO 639-3, respectively.

ISO 639/RA-JAC has been composed of

— one representative of ISO 639-1 Registration Authority,

— one representative of ISO 639-2 Registration Authority,

— one representative of ISO 639-3 Registration Authority,

— three representatives of ISO/TC 37 (nominated by ISO/TC 37), and

— three representatives of ISO/TC 46 (nominated by ISO/TC 46).

Both TCs may nominate substitute representatives.

The representatives of ISO 639-1/RA, ISO 639-2/RA, and ISO 639-3/RA will hold chair on a two-year rotating basis.

Up to five technical experts may be asked to participate as non-voting observers. The observers will be entitled to receive documents sent to the ISO 639/RA-JAC membership and are expected to provide comments in return.

B.3.2 Transition rules

The Joint Advisory Committee ceases to exist at the latest at the time of publication of this International Standard. All members of the ISO 639/RA-JAC are nominated to be members of the ISO 639 Maintenance team.

B.4 Secretariat

B.4.1 Current situation

The ISO 639/RA-JAC has appointed a secretary to prepare matters to be decided by the committee, to organize ballots, and to maintain documentation of decisions relating to the code tables of ISO 639.

B.4.2 Transition rules

The current secretariat transfers all responsibilities to the new ISO 639 at the latest at the time of publication of this International Standard. The current secretary is nominated to be a member of the ISO 639 Maintenance team.

Bibliography

[1] ISO 9:1995, *Information and documentation — Transliteration of Cyrillic characters into Latin characters — Slavic and non-Slavic languages.*

[2] ISO 233:1984, *Documentation — Transliteration of Arabic characters into Latin characters.*

[3] ISO 233-2:1993, *Information and documentation — Transliteration of Arabic characters into Latin characters — Part 2: Arabic language — Simplified transliteration.*

[4] ISO 233-3:1999, *Information and documentation — Transliteration of Arabic characters into Latin characters — Part 3: Persian language — Simplified transliteration.*

[5] ISO 259:1984, *Documentation — Transliteration of Hebrew characters into Latin characters.*

[6] ISO 259-2:1994, *Information and documentation — Transliteration of Hebrew characters into Latin characters — Part 2: Simplified transliteration.*

[7] ISO 843:1997, *Information and documentation — Conversion of Greek characters into Latin characters.*

[8] ISO 3602:1989, *Documentation — Romanization of Japanese (kana script).*

[9] ISO 7098:1991, *Information and documentation — Romanization of Chinese.*

[10] ISO 9984:1996, *Information and documentation — Transliteration of Georgian characters into Latin characters.*

[11] ISO 9985:1996, *Information and documentation — Transliteration of Armenian characters into Latin characters.*

[12] ISO/IEC 10646:2003, *Information technology — Universal multiple-octet coded character set (UCS).*

[13] ISO 11940:1998, *Information and documentation — Transliteration of Thai.*

[14] ISO/TR 11941:1996, *Information and documentation — Transliteration of Korean script into Latin characters.*

[15] ISO 19108:2002, *Geographic information — Temporal schema.*

[16] ISO 19115:2003, *Geographic information — Metadata.*

[17] ISO 19135:2005, *Geographic information — Procedures for item registration.*

[18] Peter Constable. 2002. “Toward a model for language identification: defining an ontology of language-related categories.” (SIL electronic working papers, 2002-003.) Dallas: Summer Institute of Linguistics. http://www.sil.org/silewp/abstract.asp?ref=2002-003.

[19] David Dalby. 1999/2000. *The Linguasphere Register of the World’s Languages and Speech Communities.* Linguasphere Press, Hebron, Wales. First edition.

[20] *Ethnologue. Languages of the World.* SIL International, Dallas, Texas, USA. Fifteenth edition, 2005. The content is also accessible at http://www.ethnologue.com/.