

Initial Language Subtag Registry

Status of This Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2006).

Abstract

This memo defined the initial contents of the IANA Language Subtag Registry for use in forming tags for the identification of languages. Since the contents of this memo only served as a starting point for the registry, its actual contents have been removed before publication to avoid confusion.

Table of Contents

| | |
|---|---|
| 1. Introduction | 2 |
| 2. Initialization of the Registry | 2 |
| 3. Initial Registry Contents | 5 |
| 4. Omitted Code Elements | 5 |
| 5. Security Considerations | 5 |
| 6. IANA Considerations | 5 |
| 7. References | 6 |
| 7.1. Normative References | 6 |
| 7.2. Informative References | 6 |

1. Introduction

[RFC4646] provides for a Language Subtag Registry and describes its format. This memo defines the initial contents of the IANA Language Subtag Registry, using the criteria described in Section 2.

The Language Subtag Registry is formatted in a modified record-jar text format, as described in [record-jar]. The specific format of the registry, and the definition and intended purpose of each of the fields, are described in [RFC4646].

The registry is expected to change over time, as new subtags are registered and existing subtags are modified or deprecated. The process of updating the registry is described in Section 3 of [RFC4646]. This memo does not define the permanent contents of the registry and should not be represented as doing so.

Many of the subtags defined in this registry are based on code elements defined in [ISO639-1], [ISO639-2], [ISO15924], [ISO3166-1], and [UN_M.49]. This registry is not a mirror of the code lists defined by these standards and should not be used as one.

2. Initialization of the Registry

Section 3.7 of [RFC4646] requires that the LTRU working group create an initial version of the Language Subtag Registry and populate it with the initial set of subtags. This involves converting the entries from the existing IANA language tag registry defined by [RFC3066] to the new format, as well as defining valid subtags from various source standards. This section describes the process that was used to create the initial registry entries.

The initial set of records was based on the following standards: [ISO639-1], [ISO639-2], [ISO15924], and [ISO3166-1]. The following criteria were used to select and format the records of the subtags included in the initial Language Subtag Registry (hereafter "ILSR"):

1. For each source standard, the date of the standard referenced in [RFC1766] was selected as the starting date. Code elements that were valid on that date in the selected standard were added to the ILSR. Code elements that were previously assigned, but that were vacated or withdrawn before that date, were not added to the ILSR.
2. For each successive change to the standard, any additional assignments up to the date of the adoption of [RFC4646] were added to the ILSR. Values that have been withdrawn are marked as deprecated, but not removed. Changes in meaning or

assignment of a subtag were permitted during this process (for example, the [ISO3166-1] code element 'CS' was originally assigned to Czechoslovakia and is now assigned to Serbia and Montenegro).

Code elements from [UN_M.49] were also included in the ILSR using the criteria above, with the following additional rules:

3. UN numeric code elements assigned to "macro-geographical (continental)" regions as of the date of adoption of [RFC4646] were added to the ILSR and thereby made valid for use in language tags.
4. The UN numeric code elements for "economic groupings" or "other groupings," and the alphanumeric code elements in Appendix X of the UN document, were not added to the ILSR.
5. The UN numeric code elements for countries or areas not associated with an assigned [ISO3166-1] alpha-2 code element were not added to the ILSR. These values are listed in Section 4 and may be requested for registration by individuals using the process defined in [RFC4646] and according to the rules described therein. Listing of these code elements in this section is not a guarantee of future registration.
6. Code elements that were withdrawn, vacated, or deprecated from [UN_M.49] as of the date of adoption of [RFC4646] were not added to the ILSR.

Using the initial set of subtags described above, the tags in the [RFC3066] registry were evaluated as follows:

7. Tags in the [RFC3066] registry that were not deprecated, consisted entirely of subtags already in this document, and have the correct form and format for tags defined by [RFC4646] were converted to records of type "redundant" in the ILSR. For example, "zh-Hant" is now defined by [RFC4646] because 'zh' is an [ISO639-1] code element and 'Hant' is an [ISO15924] code element, and both are defined as subtags in the ILSR.
8. Tags in the [RFC3066] registry that contained one or more subtags that either did not match the valid registration pattern or were not otherwise defined by [RFC4646] were converted to corresponding records of type "grandfathered" in the ILSR. These records cannot become type "redundant" except by revision of [RFC4646], but may have a "Deprecated" and

"Preferred-Value" field added to them if a subsequent subtag assignment or combination of assignments renders the tag obsolete.

9. Tags in the [RFC3066] registry that had a notation that they were deprecated were converted to records of type "grandfathered" in the ILSR. The record for the grandfathered entry contains a "Deprecated" field with the most appropriate date that can be determined for when the [RFC3066] record was deprecated. The "Comments" field may optionally contain a reason for the deprecation. The "Preferred-Value" field contains a tag that replaces the value. For example, the [RFC3066] tag "art-lojban" is deprecated and thus appears as a grandfathered tag in the ILSR. Its "Deprecated" field contains the deprecation date (in this case "2003-09-02") and the "Preferred-Value" field the value "jbo".
10. The remaining tags in the [RFC3066] registry are not deprecated and have a format consistent with language tags as defined by [RFC4646] but contain subtags that are not defined in the ILSR. These subtags are eligible for registration as variants. The ILSR contains appropriate variant records for the following list of subtags, and the registered [RFC3066] tags containing these subtags were entered into the ILSR as type "redundant":
 - 1901 (use with Prefix: de)
 - 1996 (use with Prefix: de)
 - nedis (use with Prefix: sl)
 - rozaj (use with Prefix: sl)
11. All remaining [RFC3066] registered tags were converted to records of type "grandfathered" in the ILSR. Interested parties may use the registration process in [RFC4646] to attempt to register the variant subtags not already present in the Language Subtag Registry. If all of the subtags in the original tag become fully defined by the resulting registrations, then the original tag is superseded. Such tags will have their record changed from type "grandfathered" to type "redundant" in the registry. Note that previous approval of a tag under [RFC3066] is not a guarantee of approval of a variant subtag under [RFC4646]. The existing [RFC3066] tag maintains its validity, but the original reason for its registration might have become obsolete.

3. Initial Registry Contents

The remainder of this section specified the initial set of records for the registry. This material was deleted on publication of this memo, to avoid any potential confusion with the registry itself. The IANA language subtag registry can be found at <http://www.iana.org/numbers.html> under "Language Tags".

4. Omitted Code Elements

The following code elements from [UN_M.49] were not associated with [ISO3166-1] alpha-2 code elements. Consequently, they were not assigned as subtags in the initial Language Subtag Registry, but were valid candidates for registration as region subtags, using the process in [RFC4646]:

| | |
|-----|-----------------|
| 830 | Channel Islands |
| 831 | Guernsey |
| 832 | Jersey |
| 833 | Isle of Man |

The last three became ineligible for registration in April, 2006, when the [ISO3166-1] code elements GG, JE, and IM were assigned as region subtags.

5. Security Considerations

This document specifies the initial contents to be used by IANA in populating the Language Subtag Registry. For security considerations relevant to that registry and the use of language tags, see [RFC4646].

6. IANA Considerations

This document points to the initial content for the Language Subtag Registry which is maintained by the IANA. The IANA language subtag registry can be found at <http://www.iana.org/numbers.html> under "Language Tags". For details on the procedures for the format and ongoing maintenance of this registry, see [RFC4646].

7. References

7.1. Normative References

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

7.2. Informative References

- [ISO15924] International Organization for Standardization, "ISO 15924:2004. Information and documentation -- Codes for the representation of names of scripts", January 2004.
- [ISO3166-1] International Organization for Standardization, "ISO 3166:1988. Codes for the representation of names of countries, 3rd edition", August 1988.
- [ISO639-1] International Organization for Standardization, "ISO 639-1:2002. Codes for the representation of names of languages -- Part 1: Alpha-2 code", 2002.
- [ISO639-2] International Organization for Standardization, "ISO 639-2:1998. Codes for the representation of names of languages -- Part 2: Alpha-3 code, first edition", 1998.
- [RFC1766] Alvestrand, H., "Tags for the Identification of Languages", RFC 1766, March 1995.
- [RFC3066] Alvestrand, H., "Tags for the Identification of Languages", BCP 47, RFC 3066, January 2001.
- [UN_M.49] Statistics Division, United Nations, "Standard Country or Area Codes for Statistical Use", UN Standard Country or Area Codes for Statistical Use, Revision 4 (United Nations publication, Sales No. 98.XVII.9, June 1999.
- [record-jar] Raymond, E., "The Art of Unix Programming", 2003.

Author's Address

Doug Ewell (Editor)
Consultant

EMail: dewell@adelphia.net
URI: <http://users.adelphia.net/~dewell>

Full Copyright Statement

Copyright (C) The Internet Society (2006).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).

