

DRAFT: Translation and Transliteration of Contact Information Policy

Scope

This Policy applies to all ICANN-accredited registrars and gTLD registries.

Effective Date

TBD

All provisions detailed herein are dependent on the availability of the Registration Data Access Protocol (RDAP) as a production service in the gTLD space and corresponding Extensible Provisioning Protocol (EPP) extensions to accommodate the receipt of transformed contact information and language tag data within RDAP. As of Q4 2018, RDAP is under development, and the EPP extensions required to implement this policy have not yet been defined.

Those provisions requiring an EPP extensions have been emphasized in the policy text below.

Definitions

1. **Contact Information** is a subset of domain name registration data. It is the information that enables someone using a domain name registration data directory service (such as RDAP or the legacy WHOIS) to contact the domain name registration holder. It includes, as applicable, the name, organization, postal address, phone number, fax number, and email address of the registered name holder, as well as the technical and administrative entities associated with the domain name registration.¹

Specifically, “contact information” in the context of this policy refers to contact information contained within the “Registrant,” “Admin,” and “Tech” fields of an RDDS output.

“Contact information” thus includes the following fields:

- Registrant/Admin/Tech ID
- Registrant/Admin/Tech Name
- Registrant/Admin/Tech Organization
- Registrant/Admin/Tech Street

¹ See the [Translation and Transliteration of Contact Information Working Group’s Final Report](#), p. 12.

- Registrant/Admin/Tech City
- Registrant/Admin/Tech State/Province
- Registrant/Admin/Tech Postal Code
- Registrant/Admin/Tech Country
- Registrant/Admin/Tech Phone
- Registrant/Admin/Tech Fax
- Registrant/Admin/Tech E-mail

Not all of these contact information fields are amenable to transformations. Refer to [Section 4](#) for the model detailing which fields can be transformed (indicated by a requirement for a language tag).

2. **Language tags** are used to help identify languages, whether spoken, written, signed, or otherwise signaled, for the purpose of communication. This includes constructed and artificial languages, but excludes languages not intended primarily for human communication, such as programming languages. The requirements for and use of language tags are detailed in [RFC 5646: “Tags for Identifying Languages.”](#) [RFC 6497: “Extension T: Transformed Content”](#) defines an extension for specifying the source of content that has been transformed, including text that has been transliterated, transcribed, or translated, or in some other way influenced by the source.

In the context of this policy, language tags will accompany contact information data fields in registration data directory services that have been transformed (i.e. translated, transliterated, or transcribed). If a transformation is performed, these tags will also accompany the original contact information data fields in order to identify the source language.

3. **RDDS** is the acronym for Registration Data Directory Service(s).
4. **Transcription**, especially orthographic transcription, refers to the process of representing text making use of the standard spelling system of each target language.²

² The United Nations Group of Experts on Geographical Names (UNGEGN) provides a more formal and detailed definition of transcription: “(a) A method of phonetic names conversion between different languages, in which the sounds of a source language are recorded in terms of a specific target language and its particular script, normally without recourse to additional diacritics. (b) A result of this process. Examples: Turkish Ankara → Greek Αγκαρα; Russian Щукино → English Shchukino. Transcription is not normally a reversible process. Retranscription (e.g. by computer) might result in a form differing from the original, for example in the above cases in Turkish Agkara and Russian Щчукино. However, Pinyin romanization of Chinese, although being a conversion between scripts, but being phonetic and non-reversible, is also regarded as transcription and not as transliteration.” Orthographic transcription involves the “spelling of words according to the prescribed rules of a given linguistic tradition.” See https://unstats.un.org/unsd/geoinfo/UNGEGN/docs/pubs/Glossary_of_terms_rev.pdf

5. **Translation** refers to the process of conveying the meaning of some passage of text in one language, so that it can be expressed equivalently in another language.³
6. **Transliteration** refers to the process of representing the characters of an alphabetic or syllabic system of writing by the characters of a conversion alphabet.⁴
7. **Transformation** in the context of this policy refers to transcription OR translation OR transliteration.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119, which is available at <http://www.ietf.org/rfc/rfc2119.txt>.

Policy Requirements

1. The following provisions apply to both gTLD Registries and ICANN-accredited Registrars:

1.1 It is OPTIONAL for registries and registrars to perform transformations of contact information in RDDS data fields.

1.2 It is OPTIONAL for registries and registrars to enter transformations of contact information into RDDS data fields.

1.3 If a transformation of contact information is performed, each of the original and transformed contact data elements MUST be accompanied by a language

³ See the [Translation and Transliteration of Contact Information Working Group's Final Report](#), p. 27. The UNGEGN provides a more formal and detailed definition of translation: "(a) The process of expressing meaning, presented in a source language, in the words of a target language. (b) A result of this process. In toponymy it is sometimes applied only to the generic element of a name. Examples: Mer Noire (French for Russian Čornoje More); Casablanca (Spanish for Arabic Dār al-Bay,dā'); Lake Como (English for Italian Lago di Como); Mount Fuji (English for Japanese Fuji San)." See https://unstats.un.org/unsd/geoinfo/UNGEEN/docs/pubs/Glossary_of_terms_rev.pdf

⁴ Ibid., p. 27. The UNGEGN provides a more formal and detailed definition of transliteration: "(a) A method of names conversion between different alphabetic scripts and syllabic scripts, in which each character or di-, tri- and tetragraph of the source script is represented in the target script in principle by one character or di-, tri- or tetragraph, or a diacritic, or a combination of these. Transliteration, as distinct from transcription, aims at (but does not necessarily achieve) complete reversibility, and must be accompanied by a transliteration key. (b) A result of this process. Examples (with English exonyms in parentheses): هرة → al-Qāhirah (Cairo); Владивосток → Vladivostok; חיפה → Hefa (Haifa); → Adis Abeba (Addis Ababa)." See https://unstats.un.org/unsd/geoinfo/UNGEEN/docs/pubs/Glossary_of_terms_rev.pdf

tag per the requirements set out in [Section 3](#) and according to the data model provided in [Section 4](#).

1.4 Registries and registrars MAY support any language and script for registrants to input registration data. All data input MUST conform to standards in the Registrar Accreditation Agreement (RAA), Registry Agreement (RA), relevant Consensus Policy, Additional WHOIS Information Policy (AWIP) and any other applicable policies.

1.5 Registrars who opt to gather language and script data from registrants with the intention to provide transformed contact information and language tag data within an RDDS SHOULD reach an agreement with relevant registry operators (i.e., those registry operators that opt in pursuant to [Provisions 1.1](#) and [1.2](#) above) to populate and validate the underlying data for language tags in accordance with the provisions outlined in [Section 3](#) and those Policies and Agreements referenced in [Provision 1.4](#).

1.6 Registrars who opt to gather language and script data from registrants with the intention to provide transformed contact information and language tag data within an RDDS MUST validate and verify original and transformed registration data for correct format and accuracy in accordance with the Policies and Agreements referenced in [Provision 1.4](#) (see [Provision 3.4](#) on script validation).

1.7 Registries who have reached an agreement with any registrar(s) to populate and validate the underlying data in accordance with [Provision 1.5](#) MUST validate original and transformed contact information for correct syntax in accordance with the Policies and Agreements referenced in [Provision 1.4](#) (see [Provision 3.4](#) on script validation).

1.8 **[EPP extension required]** If a registrar opts to gather language and script data from registrants with the intention to provide transformed contact information and language tag data within an RDDS, and has reached an agreement with relevant registry operators (i.e., those registry operators that opt in pursuant to [Provisions 1.1](#) and [1.2](#) above) to populate and validate the underlying data for language tags in accordance with [Provision 1.5](#), then the registrar MUST provide the registry or registries with the language and script data for each data field requiring a language tag according to the data model provided in [Section 4](#).

1.9 If multiple transformations are performed, each version of the transformed data must include language tags in accordance with the data model provided in [Section 4](#).

1.10 Any coded character associated with contact information entered into an RDDS MUST be supported by the latest version of Unicode (see [Provision 3.4](#) on script validation) so that registries and registrars MAY use automated means to identify the entered script.

1.11 Any transformed contact information data must be escrowed along with the original contact information data provided by the registrant.

2. The following provisions apply to Registrars only:

2.1 Registrars MAY allow registrants to provide language data for the registration data elements requiring a language tag in the data model provided in [Section 4](#).

2.2 Registrars MAY allow — but MUST NOT require — registrants to provide transformations of registration data elements requiring a language tag in the data model provided in [Section 4](#).

3. Requirements for language tags within the Registration Data Access Protocol (RDAP) (for those parties opting to provide or receive transformed contact information and language tag data):

3.1 **[EPP extension required]** If a transformation of contact information is performed, an RFC 5646-compliant language tag MUST be provided with the original contact information fields provided by the registrant and the transformed fields according to the data model provided in [Section 4](#).

3.2 **[EPP extension required]** An RFC 6497-compliant language tag MUST be provided with all transformed data fields to indicate:

1. that a transformation has been made.
2. the transliteration or transcription standard, or conventional scheme (if no formal standard exists), that was used in cases of transliteration or transcription.⁵
3. the source of the transformation, the values for which are limited to: “registrant”, “registrar”, “registry”, “reseller”, or “other”.

3.3 Language tags for each data element requiring them per the data model in [Section 4](#) MUST contain at least the primary language and script subtags per RFC 5646.

3.4 A script subtag MUST be valid for all Unicode code points detected (see [Provision 1.10](#)). Registries that opt in pursuant to [Provisions 1.1](#) and [1.2](#) above and that have reached an agreement with registrars to populate and validate the underlying data for language tags in accordance with [Provision 1.5](#) MUST validate that the Unicode code points in the entered data match with the corresponding script subtag.

⁵ A “conventional scheme” is one that is not laid down in an official or formal “standard”. For example, an ISO standard may specify “Gorbačëv” as a standard transliteration of the Russian leader’s name from the Cyrillic “Горбачёв,” but the German press often uses “Gorbatschow” as part of its conventional transliteration scheme.

3.5 If the language is not known, the primary language subtag "und" MUST be used in the language tag.

3.6 If the script cannot be determined (e.g., in cases of inputs that do not conform to mixed-script usage within a language), the script subtag "Zyyy" MUST be used in the language tag.

3.7 Private use language tags as described in [RFC 5646, Section 2.2.7](#) MUST NOT be used to generate language tags.

4. Data Model for RFC 5646- and 6497-Compliant Language Tagging for Original and Transformed RDDS Outputs⁶

NB: The table below is to be utilized by those parties who opt to exchange transformed contact information and language tag data within RDAP. Language tags need to be present with the data fields indicated below.

Data Element	Format	Min length	Max length	Cardinality	RFC 5646 and 6497 Language Tag Requirements
Registrant/Admin/Tech ID	Freeform text	1	255	1	n/a
Registrant/Admin/Tech Name	Freeform text	1	255	{0,1}	n/a ⁷
Registrant/Admin/Tech Organization	Freeform text	1	255	{0,1}	required
Registrant/Admin/Tech	Freeform text in a language	1	255	{1,3}	required

⁶ This data model is derived from the one included in the [Internationalized Registration Data Working Group's Final Report](#), and has been incorporated into the Translation and Transliteration of Contact Information implementation per [Board Resolution 2016.03.10.07](#): "Resolved (2016.03.10.07), the President and CEO, or his designee(s), is directed to work with the implementation review team for the new consensus policy on translation and transliteration to consider the IRD Working Group's data model and requirements and incorporate them, where appropriate, to the extent that the IRD's recommendations are consistent with, and facilitate the implementation of the new consensus policy on translation and transliteration."

⁷ While a language tag for "Registrant/Admin/Tech Name" was proposed as a requirement by the Internationalized Registration Data Working Group, the Implementation Review Team has determined that providing a tag for a proper name would be unduly onerous given the difficulty of ascribing a language to it. Many proper names are based on a combination of languages.

Street	or script appropriate for its region.				
Registrant/Admin/Tech City	Freeform text in a language or script appropriate for its region	1	255	1	required
Registrant/Admin/Tech State / Province	Freeform text in a language or script appropriate for its region.	1	255	{0,1}	required
Registrant/Admin/Tech Country / Territory	ISO 3166 part 2 code list	2	3	1	n/a
Registrant/Admin/Tech Postal Code	Freeform text	1	255	{0,1}	n/a
Registrant/Admin/Tech Phone	RFC 5733		64	1	n/a
Registrant/Admin/Tech Phone Ext	RFC 5733		64	{0,1}	n/a
Registrant/Admin/Tech Fax	RFC 5733		64	{0,1}	n/a
Registrant/Admin/Tech Fax Ext	RFC 5733		64	{0,1}	n/a
Registrant/Admin/Tech Email	RFC 5322 / 6532		255	1	n/a