

Network Working Group
Internet-Draft
Updates: 5890, 5891, 5894 (if approved)
Intended status: Standards Track
Expires: 10 August 2025

J.C. Klensin

A. Freytag
ASMUS, Inc.
6 February 2025

Internationalized Domain Names in Applications (IDNA): Registry
Restrictions and Recommendations
draft-klensin-idna-rfc5891bis-10

Abstract

The IDNA specifications for internationalized domain names combine rules that determine the labels that are allowed in the DNS without violating the protocol itself and an assignment of responsibility, consistent with earlier specifications, for determining the labels that are allowed in particular zones. Conformance to IDNA by registries and other implementations requires both parts. Experience strongly suggests that the language describing those responsibilities was insufficiently clear to promote safe and interoperable use of the specifications and that more details and discussion of circumstances would have been helpful. Without making any substantive changes to IDNA, this specification updates two of the core IDNA documents (RFCs 5890 and 5891) and the IDNA explanatory document (RFC 5894) to provide that guidance and to correct some technical errors in the descriptions.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 10 August 2025.

Copyright Notice

Copyright (c) 2025 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	3
2. Registry Restrictions in IDNA2008	5
3. Progressive Subsets of Allowed Characters	6
4. Considerations for Domains Operated Primarily for the Benefit of the Registry Owner, Operator, or a Related Organization	8
5. Other corrections and updates	9
5.1. Updates to RFC 5890	10
5.2. Updates to RFC 5891	11
6. Related Discussions	11
7. Security Considerations	12
8. Acknowledgments	12
9. IANA Considerations	13
10. References	13
10.1. Normative References	13
10.2. Informative References	14
Appendix A. Change Log	17
A.1. Changes from version -00 (2017-03-11) to -01	17
A.2. Changes from version -01 (2017-09-12) to -02	17
A.3. Changes from version -02 (2019-07-06) to -03	17
A.4. Changes from version -03 (2019-07-22) to -04	18
A.5. Changes from version -04 (2019-08-02) to -05	18
A.6. Changes from version -05 (2019-08-29) to -06	18
A.7. Changes from version -06 (2020-07-13) to -07	19
A.8. Changes from version -07 (2020-10-04) to -08	19
A.9. Changes from version -08 (2025-01-31) to -09	20
A.10. Changes from version -09 (2025-02-01) to -10	20
Authors' Addresses	20

1. Introduction

This document is an integral part of the specifications for Internationalized Domain Names in Applications (IDNA) [RFC5890] [RFC5891] [RFC5894] (collectively known, along with RFC 5892 [RFC5892], RFC 5893 [RFC5893] and updates to them, as "IDNA2008" or just "IDNA"). While some of the sections of those documents are reviewed below, this document assumes that readers will be familiar with the other specifications in the collection. It also assumes the terminology of those documents, e.g., that "registry", "registrar", "zone", and similar terms are used below in the DNS sense (see [RFC1034] and Section 9 of [RFC9499]) at any level of the DNS tree or any branch of the tree, even in zones that do not use that specific terminology or distinguish between the two, unless clearly specified otherwise. Those specifications impose a requirement that domain name system (DNS) registries restrict the characters they allow in domain name labels (see Section 2 below), and the contents and structure of those labels. That requirement and restriction are consistent with the "duty to serve the community" described in the original specification for DNS naming and authority [RFC1591]. The restrictions are intended to protect against security problems and confusion about and between names by limiting the permitted characters and strings to those for which the registries or their advisers have a thorough understanding and for which they are willing to take responsibility.

That understanding is key to being confident that:

- 1) Users cannot be tricked by the presence of obscure and unfamiliar characters.
- 2) Users cannot be confused by the use of strings in labels that are inconsistent with the rules in many complex scripts restricting arbitrary placement of characters.
- 3) Users cannot encounter labels that are difficult to read consistently and predictably because their appearance varies too much across common rendering systems.
- 4) Other security risks due to some feature of the writing system in use are eliminated or at least significantly mitigated.

Those provisions are centrally important because it recognized that historical relationships and variations among scripts and writing systems, the continuing evolution of those systems, differences in the uses of characters among languages (and locations) that use the same script, and so on make it impossible to generate a guideline consisting of a single list of characters and/or simple rules that would provide a completely adequate guideline with the character of "if we use these rules, we will be safe from confusion and attacks".

The algorithm and rules of RFCs 5891 and 5892 eliminate many of the most dangerous and otherwise problematic cases, but they cannot eliminate the need for registries and registrars to take additional steps to mitigate security risks and confusion by suitably restricting the repertoire and structure of labels they permit. This, in turn, requires that they or their advisers have a thorough understanding of the issues associated with for a given set of characters or writing system, that they understand what they are doing and that they take responsibility for the decisions that are made.

The way in which the IDNA2008 specifications expressed these requirements may have underemphasized the intention that they actually be treated as requirements. Section 2.3.2.3 of the Definitions document [RFC5890] mentions the need for the restrictions, indicates that they are mandatory, and points the reader to section 4.3 of the Protocol document [RFC5891]. That document, in turn, points to Section 3.2 of the Rationale document [RFC5894], with each document providing further detail, discussion, and clarification.

At the same time, the Internet has evolved significantly since the management assumptions for the DNS were established with RFC 1591 and earlier. In particular, the management and use of domain names have gone through several transformations. Recounting of those changes is beyond the scope of this document but one of them has had significant practical impact on the degree to which the requirement for registry knowledge and responsibility is observed in practice. When RFC 1591 was written, the assumption was that domains at all levels of the DNS would be operated in the best interest of the registrants in the domain and of the Internet as a whole. There were no notions about domains being operated as a profitable service, much less with a business model that made them more profitable (or even allowed adequate cost recovery) the more names that could be registered (or even, under some circumstances, reserved and not registered). During that same period, there was also no notion that domains would be considered more successful based on the number of names registered and delegated from them. While rarely reflected in the DNS protocols, the distinction between domains operated primarily as a

revenue source for the organizations operating the registry and ones that are operated purely as a service to registrants and the Internet and its uses or for, e.g., use within an enterprise, have become very important today. See Section 4 for a discussion on how those issues affect this specification.

This specification is intended to unify and clarify these requirements for registry decisions and responsibility and to emphasize the importance of registry restrictions at all levels of the DNS. It also makes a specific recommendation for character repertoire subsetting that is intermediate between the code points allowed by RFCs 5891 and 5892 and those allowed by individual registries. It does not alter the basic IDNA2008 protocols and rules themselves in any way.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 RFC 2119 [RFC2119] and RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Registry Restrictions in IDNA2008

As mentioned above, IDNA2008 specifies that the registries for each zone in the DNS that supports IDN labels are required to develop and apply their own rules to restrict the allowable labels, including limiting characters they allow to be used in labels in that zone. The chosen list MUST be a subset of the collection of code points specified as "PVALID", "CONTEXTJ", and "CONTEXT0" by the rules established by the protocols themselves. Labels containing any characters from the two CONTEXT categories or any characters that are normally part of a script written right to left [RFC5893] require that additional rules, specified in the protocols and known as "contextual rules" and "bidi rules", be applied. The entire collection of rules and restrictions required by the IDNA2008 protocols themselves are known as "protocol restrictions".

Registries may apply (and generally are required to apply) additional rules to further restrict the list of permitted code points, contextual rules (perhaps applied to normally PVALID code points) that apply additional restrictions, and/or restrictions on labels as distinct from code points. In particular, safe and secure use of any of a large number of widely-used scripts or writing systems require the addition of contextual rules that go beyond the very limited restrictions implemented in by "CONTEXTJ" and "CONTEXT0" at the protocol level, but which are otherwise functionally equivalent in that they constitute limitations on where allowable code points may be placed in a label.

In contrast, protocol restrictions are by necessity somewhat generic, having to cater both to the union of the needs for all zones and to the fact that it may be entirely reasonable that some zones, especially zones deep in the DNS tree, be more permissive than others. That must be done without negative impact on the DNS.

Other restrictions that are necessary, perhaps obviously so, include provisions for restricting suggested new registrations based on conflicts with labels already registered in the zone, so as to avoid homograph attacks [Gabrilovich2002] and other issues. The specifications of what constitutes such conflicts, as well as the definition of "conflict" based on the properties of the labels in question, is the responsibility of each registry. Such specifications should further include prohibitions on code points and labels that are not consistent with the intended function of the zone, the subtree in which the zone is embedded (see Section 3), and consequent differences in the stringency of security-related measures.

These per-registry (or per-zone) rules are commonly known as "registry restrictions" to distinguish them from the protocol restrictions described above. Such registry restrictions are essential to provide for the necessary security in the face of the tremendous variations and differences in writing systems and their ongoing evolution and development, as well as the human ability (or inability) to recognize and distinguish characters in different scripts around the world and under different circumstances.

3. Progressive Subsets of Allowed Characters

Three existing rules could easily be misunderstood and are repeated and rephrased here to provide additional clarity:

- i. RFCs 5891 and 5892 determine the set of code points that are possible for inclusion in domain name labels; registries MUST reject all other code points.
- ii. In addition, labels MUST NOT contain code points in positions where they violate the "CONTEXTJ" or "CONTEXT0" rules or other restrictions defined in RFCs 5891 and 5892.
- iii. Labels that contain code points that are normally written from right to left MUST also conform to the requirements of RFC 5893.

Many registries have decided to restrict each label to a single script, even if the registry permits different labels to use different scripts. The Unicode script property is defined in UAX24 [UAX24], and the relevant data tables are published in a supplemental document [UAX24-scripts-text]. However, while suitable for many purposes, experts on particular languages or usages sometimes define scripts and their contents differently and there are no requirements that registries use the Unicode definitions.

Unicode includes many code points that are unfamiliar to most current Internet users. Several groups have developed code point repertoires and rules that include what they have concluded is needed, while excluding, e.g., historic code points that might be used to confuse 21st century readers. Examples include:

- i. The RFC Series includes specific recommendations about some scripts or languages, including RFCs for Cyrillic script RFC 4992 [RFC5992] and Arabic language RFC 5564 [RFC5564].
- ii. ICANN maintains a set of Label Generation Rules [SL-REF-LGR], each of which contains a repertoire and rules for one script or language. At the time of writing there are 23 script-based LGRs and 31 language-based LGRs.

While registries may allow code points that the RFC contributors or ICANN committees decided to exclude, registries are advised to be careful and allow such code points only after they are confident they understand the reasons for the exclusions.

Basing the permitted repertoire and defining constraints on one or more of the specifications mentioned above enables members of the general public to register domain names based on contemporary spellings and character choices, while limiting the scope for mischief as far as practically possible.

Registries that cater to a world-wide audience may wish to allow scripts for which a suitable repertoire has not yet been developed or reached broad consensus. Basing the permitted repertoire on the general Unicode Identifier specification [UAX31] may be a good option in such cases but those considering doing so should keep in mind that there are substantive reasons, some of them consequences of the design of the DNS itself, why domain names are not governed by the rules of that Unicode specification.

4. Considerations for Domains Operated Primarily for the Benefit of the Registry Owner, Operator, or a Related Organization

As discussed in the Introduction (Section 1), the distributed administrative structure of the DNS today can be described by dividing zones into two categories depending on how they are administered and for whom. These categories are not precise -- some zones may not fall neatly into one category or the other -- but are useful in understanding the practical applicability of this specification. They are:

- * Zones operating primarily or exclusively within a organization, enterprise, or country or region and responsible to the management of the organization or enterprise or the Internet users in that country or region. DNS operations, including registrations and delegations, will typically occur in support of the purpose of that country, organization or enterprise rather than being its primary purpose. This category also includes zones operated as a service to a relevant community and to the Internet user community as a whole. If there are charges associated with name registrations, they will typically be set on a cost recovery basis.
- * Zones operating primarily as all or part of a business of selling names for the financial benefit of entities responsible for the registry. For these domains, most delegations of subdomains are to entities with little or no affiliation with the registry operator other than contractual agreements about operation of those subdomains. Another common characteristic of these zones is that they are usually considered more successful the more names they can register. These zones are often known as "public domains" or with similar terms, but those terms typically have other semantics related to the level in the tree and may not cover all cases.

Rules requiring strict registry responsibility typically come naturally to registries for zones of the first type. Such rules include thorough understanding of scripts and related issues in domain name labels being considered for registration or local naming rules that have the same effect. Registration of labels that would prove problematic for any reason hurts the relevant organization or enterprise or its customers or users within the relevant country or region and more broadly. More generally, there are strong incentives to be extremely conservative about labels that might be registered and few, if any, incentives favoring adventures into labels that might be considered clever, much less ones that are hard to type, render, or, where it is relevant to users, remember correctly.

By contrast, in a zone in which the revenues are derived exclusively, or almost exclusively, from selling or reserving (including "blocking") as many names as possible, there may be perceived incentives to register whatever names would-be registrants "want" or fears that any restrictions will cut into the available namespace. In such situations, restrictions are unlikely to be applied unless they meet at least one of two criteria: (i) they are easy to apply and can be applied algorithmically or otherwise automatically and/or (ii) there is clear evidence that the particular label would cause harm.

As suggested above, the two categories above are not precise. In particular, there may be domains that, despite being set up to operate to produce revenue above actual costs, are sufficiently conservative about their operations to more closely resemble the first group in practice than the second one.

The requirement of IDNA that is discussed at length elsewhere in this specification stands: IDNA (and IDNs generally) would work better and Internet users would be better protected and more secure if registries and registrars (of any type) confined the labels they were willing to accept for registration to scripts and code point sequences that they understood thoroughly. Irrespective of in which of the two categories a registrar operates, it is the position of the IETF that significant conservatism in what is allowed to be registered, even for reservation purposes, and even more conservatism about what labels are actually entered into zones and delegated, is the best option for the Internet and its users. If practical considerations do not allow that much conservatism, then it is desirable to consult and utilize the many lists and tables that have been, and continue to be, developed to advise on what might be sensible for particular scripts and languages. Some of those lists, tables, and recommendations are described in Section 6 below.

5. Other corrections and updates

After the initial IDNA2008 documents were published (and RFC 5892 was updated for Unicode 6.0 by RFC 6452 [RFC6452] and for Unicode 12.0 RFC 9233 [RFC9233]) several errors or instances of confusing text were noted. For the convenience of the community, the relevant corrections for RFCs 5890 and 5891 are noted below and update the corresponding documents. There are no errata for RFC 5893 or 5894 as of the date this document was published. Because further updates to RFC 5892 would require addressing other pending issues, the outstanding erratum for that document is not considered here. For consistency with the original documents, references to Unicode 5.0 are preserved in this document.

5.1. Updates to RFC 5890

All but two of the outstanding errata against RFC 5890 [RFC-Editor-5890Errata] (Errata IDs 4695, 4696, 4823, and 4824 are associated with the same issue, the number of Unicode characters that can be associated with a maximum-length (63 octet) A-label. Some comments in the errata suggest expressing the value in octets. However, RFC 5890 and the other IDNA2008 documents are generally careful to work exclusively with Unicode code points. Consequently, the relevant rules in RFC 5890 are amended as follows:

Section 2.3.2.1 Old: expansion of the A-label form to a U-label may produce strings that are much longer than the normal 63 octet DNS limit (potentially up to 252 characters).

New: expansion of the A-label form to a U-label may produce strings that are much longer than the normal 63 octet DNS limit (See Section 4.2).

Comment: If the length limit is going to be a source of confusion or careful calculations, it should appear in only one place.

Section 4.2 Old: Because A-labels (the form actually used in the DNS) are potentially much more compressed than UTF-8 (and UTF-8 is, in general, more compressed than UTF-16 or UTF-32), U-labels that obey all of the relevant symmetry (and other) constraints of these documents may be quite a bit longer, potentially up to 252 characters (Unicode code points).

New: A-labels (the form actually used in the DNS) and the Punycode algorithm used as part of the process to produce them [RFC3492] are strings that are potentially much more compressed than any standard Unicode Encoding Form. A 63 octet A-label cannot represent more than 58 Unicode code points (four octet overhead and the requirement that at least one character lie outside the ASCII range) but implementations allocating buffer space for the conversion should allow significantly more space (i.e., extra octets) depending on the encoding form they are using.

Errata ID 5484 suggests modifying the paragraph of Section 2.3.2.1 that begins "For IDNA-aware applications..." to improve the pointer to definitions, adding "and in Section 2.3.1" at the end of the sentence.

Errata ID 7291 identifies RFC 5890 as updating RFC 4343. The RFC Editor's metadata has been updated to make that correction. Readers of RFC 5890 should note the correction and any replacement for RFC 5890 should address it as appropriate.

5.2. Updates to RFC 5891

There is only one outstanding erratum for RFC 5891, Errata ID 3969 [RFC5891Erratum] on improving the reference for combining marks. Combining marks are explained in the cited section, but not, as the text indicates, exactly defined.

Old: The Unicode string MUST NOT begin with a combining mark or combining character (see The Unicode Standard, Section 2.11 [Unicode] for an exact definition) .

New: The Unicode string MUST NOT begin with a combining mark or combining character (see The Unicode Standard, Section 2.11 [UnicodeA] for an explanation and Section 3.6, definition D52 [UnicodeB] for an exact definition).

Comment: When RFC 5891 is replaced, the references in the text should be updated to the current version of Unicode and the section numbers checked. This document provides more recent references than appeared in RFC 5891 but they may change again by the time it is further revised or replaced.

6. Related Discussions

This document is one of a series of measures that have been suggested to address IDNA issues raised in other documents and discussions but the only one that actually updates the IDNA Standard. Those other discussions and associated documents include suggested mechanisms for dealing with combining sequences and single-code point characters with the same appearance, including ones that Unicode normalization neither combines nor decomposed as IDNA2008 assumed. That topic was discussed further in an Internet-Draft that was never completed and published [IDNA-Unicode] and in the IAB response to that issue [IAB-2015].

RFC 8753 [RFC8753] discusses some of these issues and updates RFC 5892 to clarify and improve the review process in ways that should improve the issues discussed in Section 3. Even if applied carefully, it will not fundamentally change those issues: it is impossible for those reviews to catch all possible problematic code points. RFC 9233 [RFC9233] reflects a partial implementation of RFC 8753.

Those and other documents also discuss issues with IDNA and character graphemes for which abstractions exist in Unicode in precomposed form but that can be generated from combining sequences. Another approach has been to create a listing of code points known to be problematic [Freytag-troublesome], but that work was never carried forward either. In combination, the various discussions of combining sequences and non-decomposing characters may lay the foundation for an actual update to the IDNA code points document [RFC5892]. Such an update would presumably also address the existing errata against that document discussed at the end of Section 5.1.

Perhaps the most important contemporary efforts are ones coordinated by ICANN to develop rules for specific scripts and writing systems. They including the twin efforts of creating per-script Root Zone Label Generation Rules [ICANN-RZLGR-5] and Second Level Reference Label Generation Rules [SL-REF-LGR] (the latter of which may be per language). They also include other lists of code points or code point relationships that may be particularly problematic and that should be treated with extra caution or prohibited entirely.

At a much higher level, discussions are ongoing to consider issues, demands, and proposals for new uses of the DNS.

7. Security Considerations

As discussed in IAB recommendations about internationalized domain names [RFC4690], [RFC6912], and elsewhere, poor choices of strings for DNS labels can lead to opportunities for attacks, user confusion, and other issues less directly related to security. This document clarifies the importance of registries carefully establishing design policies for the labels they will allow and that having such policies and taking responsibility for them is a requirement, not an option. If that clarification is useful in practice, the result should be an improvement in security.

8. Acknowledgments

Many thanks to Patrik Fältström who provided an important review on the initial version, to Jaap Akkerhuis, Don Eastlake, Barry Leiba, John Levine, and Alessandro Vesely who did reviews that improved the text and to Pete Resnick who acted as document shepherd and did an additional careful review of the 2020 version of this specification.

Thanks also to Murray Kucherawy and Orie Steele who managed to get it moving again in 2024 after an extended delay starting after the initial IETF Last Call was completed in August 2019 without problems being identified by the community. Arnt Gulbrandsen provided multiple extensive and careful reviews and suggested some improved

text to address issues with, and possible misunderstandings from, earlier versions. The reviews in response to IETF Last Call that started in October 2024 were very helpful in identifying areas where clarifications or other changes were needed.

9. IANA Considerations

This memo does not contain any provisions that would alter any IDNA-related registries or tables in fundamental ways. However, any tables that reference RFC 5891 should be updated to reference this document as well.

10. References

10.1. Normative References

- [RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, DOI 10.17487/RFC1034, November 1987, <<https://www.rfc-editor.org/info/rfc1034>>.
- [RFC1591] Postel, J., "Domain Name System Structure and Delegation", RFC 1591, DOI 10.17487/RFC1591, March 1994, <<https://www.rfc-editor.org/info/rfc1591>>.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC5890] Klensin, J., "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework", RFC 5890, DOI 10.17487/RFC5890, August 2010, <<https://www.rfc-editor.org/info/rfc5890>>.
- [RFC5891] Klensin, J., "Internationalized Domain Names in Applications (IDNA): Protocol", RFC 5891, DOI 10.17487/RFC5891, August 2010, <<https://www.rfc-editor.org/info/rfc5891>>.
- [RFC5891Erratum] "RFC 5891, "Internationalized Domain Names in Applications (IDNA): Protocol", Errata ID 3969, 19 April 2014, <http://www.rfc-editor.org/errata_search.php?rfc=5891>.
- [RFC5893] Alvestrand, H., Ed. and C. Karp, "Right-to-Left Scripts for Internationalized Domain Names for Applications (IDNA)", RFC 5893, DOI 10.17487/RFC5893, August 2010, <<https://www.rfc-editor.org/info/rfc5893>>.

- [RFC5894] Klensin, J., "Internationalized Domain Names for Applications (IDNA): Background, Explanation, and Rationale", RFC 5894, DOI 10.17487/RFC5894, August 2010, <<https://www.rfc-editor.org/info/rfc5894>>.
- [RFC6912] Sullivan, A., Thaler, D., Klensin, J., and O. Kolkman, "Principles for Unicode Code Point Inclusion in Labels in the DNS", RFC 6912, DOI 10.17487/RFC6912, April 2013, <<https://www.rfc-editor.org/info/rfc6912>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.
- [RFC9499] Hoffman, P. and K. Fujiwara, "DNS Terminology", BCP 219, RFC 9499, DOI 10.17487/RFC9499, March 2024, <<https://www.rfc-editor.org/info/rfc9499>>.

10.2. Informative References

- [Freytag-troublesome]
Freytag, A., Klensin, J., and A. Sullivan, "Those Troublesome Characters: A Registry of Unicode Code Points Needing Special Consideration When Used in Network Identifiers", 31 December 2018, <[draft-freytag-troublesome-characters-02](#)>. Reference supplied for historical purposes. This document is no longer under development.
- [Gabrilovich2002]
Gabrilovich, E. and A. Gontmakher, "The Homograph Attack", Communications of the ACM 45(2):128, February 2002.
- [IAB-2015] Internet Architecture Board (IAB), "IAB Statement on Identifiers and Unicode 7.0.0", 11 February 2015, <<https://www.iab.org/documents/correspondence-reports-documents/2015-2/iab-statement-on-identifiers-and-unicode-7-0-0/>>.
- [ICANN-RZLGR-5]
Internet Corporation for Assigned Names and Numbers (ICANN): Integration Panel, "Root Zone Label Generation Rules (RZ LGR-5) Overview and Summary", 26 May 2022, <<https://www.icann.org/news/announcement-2-2019-04-25-en>>.

[IDNA-Unicode]

Klensin, J. and P. Faltstrom, "IDNA Update for Unicode 7.0.0", 8 October 2017, <draft-klensin-idna-5892upd-unicode70-05>. Reference supplied for historical purposes. This document is no longer under development.

[RFC-Editor-5890Errata]

RFC Editor, "RFC Errata: RFC 5890, "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework", August 2010", 2016, <https://www.rfc-editor.org/errata_search.php?rfc=5890>.

[RFC3492] Costello, A., "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", RFC 3492, DOI 10.17487/RFC3492, March 2003, <<https://www.rfc-editor.org/info/rfc3492>>.

[RFC4690] Klensin, J., Faltstrom, P., Karp, C., and IAB, "Review and Recommendations for Internationalized Domain Names (IDNs)", RFC 4690, DOI 10.17487/RFC4690, September 2006, <<https://www.rfc-editor.org/info/rfc4690>>.

[RFC5564] El-Sherbiny, A., Farah, M., Oueichek, I., and A. Al-Zoman, "Linguistic Guidelines for the Use of the Arabic Language in Internet Domains", RFC 5564, DOI 10.17487/RFC5564, February 2010, <<https://www.rfc-editor.org/info/rfc5564>>.

[RFC5892] Faltstrom, P., Ed., "The Unicode Code Points and Internationalized Domain Names for Applications (IDNA)", RFC 5892, DOI 10.17487/RFC5892, August 2010, <<https://www.rfc-editor.org/info/rfc5892>>.

[RFC5992] Sharikov, S., Miloshevic, D., and J. Klensin, "Internationalized Domain Names Registration and Administration Guidelines for European Languages Using Cyrillic", RFC 5992, DOI 10.17487/RFC5992, October 2010, <<https://www.rfc-editor.org/info/rfc5992>>.

[RFC6452] Faltstrom, P., Ed. and P. Hoffman, Ed., "The Unicode Code Points and Internationalized Domain Names for Applications (IDNA) - Unicode 6.0", RFC 6452, DOI 10.17487/RFC6452, November 2011, <<https://www.rfc-editor.org/info/rfc6452>>.

[RFC8753] Klensin, J. and P. Faltstrom, "Internationalized Domain Names for Applications (IDNA) Review for New Unicode Versions", RFC 8753, DOI 10.17487/RFC8753, April 2020, <<https://www.rfc-editor.org/info/rfc8753>>.

- [RFC9233] Feltstrom, P., "Internationalized Domain Names for Applications 2008 (IDNA2008) and Unicode 12.0.0", RFC 9233, DOI 10.17487/RFC9233, March 2022, <<https://www.rfc-editor.org/info/rfc9233>>.
- [SL-REF-LGR]
Internet Corporation for Assigned Names and Numbers (ICANN), "Second Level Label Generation Rules", 2019, <<https://www.icann.org/resources/pages/second-level-lgr-2015-06-21-en>>.
- [UAX24] The Unicode Consortium, "Unicode Script Property", 30 April 2024, <<https://www.unicode.org/reports/tr24/>>. The date given is as of the publication of this document; the link given points to the most current version.
- [UAX24-scripts-text]
The Unicode Consortium, "Scripts.txt", 10 September 2024, <<https://www.unicode.org/Public/UCD/latest/ucd/Scripts.txt>>. The date given is as of the publication of this document; the link given points to the most current version.
- [UAX31] The Unicode Consortium, "Unicode Identifiers and Syntax", 2 September 2024, <<https://www.unicode.org/reports/tr31/>>. The date given is as of the publication of this document; the link given points to the most current version.
- [Unicode] The Unicode Consortium, "The Unicode Standard, Version 5.0", Section 2.11, 2007. This printed reference has now been updated online to reflect additional code points. For code points, the reference at the time this document was published is to Unicode 5.2. (Note that this is the reference exactly at it appeared in RFC 5891. The best handling for this reference and the next two have been posed as questions to the RFC Production Center.)
- [UnicodeA] The Unicode Consortium, "The Unicode Standard, Version 16.0", Section 2.11, 10 September 2024, <<https://www.unicode.org/versions/Unicode16.0.0/core-spec/chapter-2/#G1708>>. Note that this can be adjusted for newer Unicode version by adjusting the version portion of the URL.

[UnicodeB] The Unicode Consortium, "The Unicode Standard, Version 16.0.0", Section 3.6, definition D52, 10 September 2024, <<https://www.unicode.org/versions/Unicode16.0.0/core-spec/chapter-3/#G30602>>. Note that this can be adjusted for newer Unicode version by adjusting the version portion of the URL.

Appendix A. Change Log

RFC Editor: Please remove this appendix before publication.

A.1. Changes from version -00 (2017-03-11) to -01

- * Added Acknowledgments and adjusted references.
- * Filled in Section 5 with updates to respond to errata.
- * Added Section 6 to discuss relationships to other documents.
- * Modified the Abstract to note specifically updated documents.
- * Several small editorial changes and corrections.

A.2. Changes from version -01 (2017-09-12) to -02

After a pause of nearly 34 months due to inability to get this draft processed, including nearly a year waiting for a new directorate to actually do anything of substance about fundamental IDNA issues, the -02 version was posted in the hope of getting a new start. Specific changes include:

- * Added a new section, Section 4, and some introductory material to address the very practical issue that domains run on a for-profit basis are unlikely to follow the very strict "understand what you are registering" requirement if they support IDNs at all and expect to profit from them.
- * Added a pointer to draft-klensin-idna-unicode-review to the discussion of other work.
- * Editorial corrections and changes.

A.3. Changes from version -02 (2019-07-06) to -03

- * Minor editorial changes in response to shepherd review.
- * Additional references.

A.4. Changes from version -03 (2019-07-22) to -04

- * Editorial changes after AD review and some additional changes to improve clarity.

A.5. Changes from version -04 (2019-08-02) to -05

- * Small editorial corrections, many to correct glitches found during IETF Last Call.
- * Updated acknowledgments, particularly to reflect reviews in Last Call.

A.6. Changes from version -05 (2019-08-29) to -06

Other than some small editorial adjustments, these changes made after, and reflect, IESG post-last-call review and comments. To the extent it was possible to do so without making this document inconsistent with the other IDNA documents, established IETF, Unicode, and ICANN community il8n terminology, or well-established IDNA or il8n practices, the first author believes that the document responds to all previously-outstanding IESG substantive comments.

- * Fixed a remaining citation issue with a Unicode document. This version has not been updated to reflect Unicode 13, but the document should be adjusted so that all references are contemporary at the time of publication.
- * Added reference to homograph attacks, and slightly adjusted discussion of them, per discussion with IESG post-last-call.
- * Removed pointer to RFC 5890 from discussion of mixed-script labels in Section 3.
- * Rewrote parts of Section 4 to eliminate the term "for-profit" and clarify the issues.
- * Removed pointer to draft-klensin-idna-unicode-review because RFC 8753 has been published and is therefore no longer pending / parallel work.
- * Rewrote Section 6 to make the relationships among various documents and efforts somewhat more clear.
- * References to RFCs 5893 and 6912 moved from Informative to Normative.

A.7. Changes from version -06 (2020-07-13) to -07

- * Significant parts of this draft have been rewritten, and text rearranged, to reflect discussions subsequent to the end of the original IETF Last Call in late August 2019 and the posting of version -06 nearly a year later to resolve IESG comments and objections that appeared to be consistent with the purpose of the document and the Last Call comments. The items below reflect the most significant changes. Note of these changes are believed to be substantive rather than improvements of clarity and explanation.
- * Multiple small editorial corrections including one more change from "profits" to "revenues" to make it clear that the motivation problem might exist even for a registry that was loss-making.
- * Extensive changes to clarify the intent of, and need for, the document and improve the explanation of its context and relationship to define additional restrictions for particular scripts or writing systems.
- * Added reference to RFC 8174 to the 2119 boilerplate.
- * In Section 5, updated the errata description for RFC 5890 and verified the absence of errata for RFCs 5893 and 5894 as of 2024-09-08.
- * Updated references including those associated with the errata list in Section 5.1.
- * Clarified the Unicode references (those in RFC 5891 were to Unicode 5.0).
- * Updated source to RFCXMLv3.

A.8. Changes from version -07 (2020-10-04) to -08

- * Adjusted BCP 14 statement to match RFC 8174.
- * Replaced Section 3 with a slightly modified version of suggested text from Arnt Gulbrandsen.
- * Modified the Introduction (Section 1) to make the audience more clear and clarify terminology. The latter includes additional language to reinforce the "all levels of the DNS" comment that has been present for many version of the document.

- * Made a small improvement in the RFC 5890 errata discussion per October Last Call comment.
- * Removed unused references "[ICANN-MSR5]", "[LGR-Procedure]", and "[RFC4713]".
- * Acknowledgments updated.
- * Many editorial corrections and improvements.

A.9. Changes from version -08 (2025-01-31) to -09

- * Replaced the "While the IETF rarely gives advice..." sentence that some found problematic.

A.10. Changes from version -09 (2025-02-01) to -10

- * Removed incomplete reference caught right after IETF LC was initiated.
- * Corrected a typo.

Authors' Addresses

John C Klensin
1770 Massachusetts Ave, Ste 322
Cambridge, MA 02140
United States of America
Phone: +1 617 245 1457
Email: john-ietf@jck.com

Asmus Freytag
ASMUS, Inc.
Email: asmus@unicode.org