Network Working Group J. Klensin

Internet-Draft

Updates: <u>5892</u> (if approved) P. Faltstrom Intended status: Standards Track Netnod

Expires: March 1, 2020 August 29, 2019

IDNA Review for New Unicode Versions draft-klensin-idna-unicode-review-03

Abstract

The standards for Internationalized Domain Names in Applications (IDNA) require a review of each new version of Unicode to determine whether incompatibilities with prior version or other issues exist and, where appropriate, to allow the IETF to decide on the trade-offs between compatibility with prior IDNA versions and compatibility with Unicode going forward. That requirement, and its relationship to tables maintained by IANA, has caused significant confusion in the past. This document makes adjustments to the review procedure based on experience and updates IDNA, specifically RFC 5892, to reflect those changes and clarify the various relationships involved. It also makes other minor adjustments to align that document with experience.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of \underline{BCP} 78 and \underline{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 March 1, 2020.

Copyright Notice

Copyright (c) 2019 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 Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

$\underline{1}$. Introduction	<u>2</u>
2. Brief History of IDNA Versions, the Review Requirement, and	
<u>RFC 5982</u>	3
$\underline{3}$. The Review Model	4
3.1. Review Model Part I: Algorithmic Comparison	<u>5</u>
3.2. Review Model Part II: New Code Point Analysis	<u>5</u>
$\underline{\textbf{4}}$. IDNA Assumptions and Current Practice	<u>6</u>
$\underline{5}$. Derived Tables Published by IANA	7
$\underline{6}$. Editorial clarification to $\underline{\text{RFC }5892}$	7
$\underline{7}$. Acknowledgements	8
$\underline{8}$. IANA Considerations	8
$\underline{9}.$ Security Considerations	9
<u>10</u> . References	9
$\underline{10.1}$. Normative References	9
<u>10.2</u> . Informative References	11
<u>Appendix A</u> . Summary of Changes to <u>RFC 5892</u>	<u>12</u>
<u>Appendix B</u> . Change Log	<u>13</u>
<u>B.1</u> . Changes from version -00 (2019-06-12) to -01	<u>13</u>
B.2. Changes from version -01 (2019-07-06) to -02	<u>14</u>
<u>B.3</u> . Changes from version -02 (2019-07-22) to -03	<u>14</u>
Authors' Addresses	<u>14</u>

1. Introduction

The standards for Internationalized Domain Names in Applications (IDNA) require a review of each new version of Unicode to determine whether incompatibilities with prior version or other issues exist and, where appropriate, to allow the IETF to decide on the trade-offs between compatibility with prior IDNA versions and compatibility with Unicode [Unicode] going forward. That requirement, and its relationship to tables maintained by IANA, has caused significant confusion in the past. This document makes adjustments to the review procedure based on nearly a decade of experience and updates IDNA, specifically the document that specifies the relationship between Unicode code points and IDNA derived properties [RFC5892], to reflect those changes and clarify the various relationships involved.

This specification does not change the requirement that registries, at all levels of the DNS tree, take responsibility for the labels they are inserting in the DNS, a level of responsibility that requires allowing only a subset of the code points and strings allowed by the IDNA protocol itself. That requirement is discussed in more detail in a companion document [RegRestr].

Terminology note: In this document, "IDNA" refers to the current version as described in RFC 5890 [RFC5890] and subsequent documents and sometimes known as "IDNA2008". Distinctions between it and the earlier version are explicit only where that is necessary to understanding the relationships involved, e.g., in Section 2.

2. Brief History of IDNA Versions, the Review Requirement, and RFC 5982

The original, now-obsolete, version of IDNA, commonly known as "IDNA2003" [RFC3490] [RFC3491] was defined in terms of a profile of a collection of IETF-specific tables [RFC3454] that specified the usability of each Unicode code point with IDNA. Because the tables themselves were normative, they were intrinsically tied to a particular version of Unicode. As Unicode evolved, the standard would either have required a new version for each new version of Unicode or it would fall further and further behind.

When that version of IDNA was superseded by the current one, known as IDNA2008 [RFC5890], a different strategy, one that was property-based rather than table-based, was adopted for a number of reasons of which the reliance on normative tables was not dominant [RFC4690]. In the IDNA2008 model, the use of normative tables was replaced by a set of procedures and rules that operated on Unicode properties [Unicode-properties] and a few internal definitions to determine the category and status, and hence an IDNA-specific "derived property", for any given code point. Those rules are, in principle, independent of Unicode versions. They can be applied to any version of Unicode, at least from approximately version 5.0 forward, to yield an appropriate set of derived properties. However, the working group that defined IDNA2008 recognized that not all of the Unicode properties were completely stable and that, because the criteria for new code points and property assignment used by the Unicode Consortium might not precisely align with the needs of IDNA, there were possibilities of incompatible changes to the derived property value. More specifically, there could be changes that would make previously-disallowed labels valid, previously-valid labels disallowed, or that would be disruptive to IDNA's defining rule structure. Consequently, IDNA2008 provided for an expert review of each new version of Unicode with the possibility of providing exceptions to the rules for particular new code points, code points whose properties had changed, and newly-discovered issues with the

IDNA2008 collection of rules. When problems were identified, the reviewer was expected to notify the IESG. The assumption was that the IETF would review the situation and modify IDNA2008 as needed, most likely by adding exceptions to preserve backward compatibility (see Section 3.1 below).

For the convenience of the community, IDNA2008 also provided that IANA would maintain copies of calculated tables resulting from each review, showing the derived properties for each code point. Those tables were expected to be helpful, especially to those without the facilities to easily compute derived properties themselves. Experience with the community and those tables has shown that they have been confused with the normative tables of IDNA2003: the IDNA2008 tables published by IANA have never been normative and statements about IDNA2008 being out of date with regard to some Unicode version because the IANA tables have not been updated are incorrect or meaningless.

3. The Review Model

While the text has sometimes been interpreted differently, IDNA2008 actually calls for two types of review when a new Unicode version is introduced. One is an algorithmic comparison of the set of derived properties calculated from the new version of Unicode to the derived properties calculated from the previous one to determine whether incompatible changes have occurred. The other is a review of newly-assigned code points to determine whether any of them require special treatment (e.g., assignment of what IDNA2008 calls contextual rules) and whether any of them violate any of the assumptions underlying the IDNA2008 derived property calculations. Any of the cases of either review might require either per-code point exceptions or other adjustments to the rules for deriving properties that are part of RFC 5892. The subsections below provide a revised specification for the review procedure.

Unless the IESG or the Designated Expert conclude that there are special problems or unusual circumstances, these reviews will be performed only for full Unicode versions (those numbered NN.0, e.g., 12.0) and not for minor updates (e.g., 12.1).

As can be seen in the detailed descriptions in the following sections, proper review will require a team of experts that has both broad and specific skills in reviewing Unicode characters and their properties in relation to both the written standards and operational needs. The IESG will need to appoint experts who can draw on the broader community to obtain the necessary skills for particular situations. See the IANA Considerations Section 8 details.

3.1. Review Model Part I: Algorithmic Comparison

Section 5.1 of RFC 5892 is the description of the process for creating the initial IANA tables. It is noteworthy that, while it can be read as strongly implying new reviews and new tables for versions of Unicode after 5.2, it does not explicitly specify those reviews or, e.g., the timetable for completing them. It also indicates that incompatibilities are to be "flagged for the IESG" but does not specify exactly what the IESG is to do about them and when. For reasons related to the other type of review and discussed below, only one review was completed, documented [RFC6452], and a set of corresponding new tables installed. That review, for Unicode 6.0, found only three incompatibilities; the consensus was to ignore them (not create exceptions in IDNA2008) and remain consistent with computations based on current (Unicode 6.0) properties rather than preserving backward compatibility within IDNA. The 2018 review (for Unicode 11.0 and versions in between it and 6.0) [IDNA-Unicode12] also concluded that maintain Unicode compatibility, rather than IDNA backward compatibility, should be maintained. That decision was partially driven by the long period between reviews and the concern that table calculations by others in the interim could result in unexpected incompatibilities if derived property definitions where then changed. See Section 4 for further discussion of these preferences.

3.2. Review Model Part II: New Code Point Analysis

The second type of review is not clearly explained in RFC 5892, but is intended to identify cases in which newly-added code points, or perhaps even newly-discovered problematic older ones, violate design assumptions of IDNA, identify defects in those assumptions, or are inconsistent (from an IDNA perspective) with Unicode commitments about assignment, properties, and stability of newly-added code points. The discovery after Unicode 7.0 was released that new code points were being added that were potentially visually equivalent, in the same script, to previously-available code point sequences was one example of the type of situation the review was expected to discover (and did so [IAB-Unicode7-2015] [IDNA-Unicode7]).

Because multiple perspectives on Unicode and writing systems are required, this review will not be successful unless done by a team -- a single, all-knowing, Designated Expert is not feasible or likely to produce an adequate analysis. It should also be recognized that, if this review identifies a problem, that problem is likely to be complex and/or involve multiple trade-offs. Actions to deal with it are likely to be disruptive (although perhaps not to large communities of users) or to leave either security risks (opportunities for attacks and inadvertent confusion as expected

matches do not occur) or excessive reliance on registries understanding and taking responsibility for what they are registering [RFC5894] [RegRestr]. The latter, while a requirement of IDNA, has often not worked out well in the past.

Because resolution of problems identified by this part of the review may take some time even if that resolution is to add additional contextual rules or disallow one or more code points, there will be cases in which it will be appropriate to publish the results of the algorithmic review and provide IANA with corresponding tables, with warnings about code points whose status is uncertain until there are IETF consensus conclusions about how to proceed. The affected code points should be considered unsafe and identified as "under review" in the IANA tables until final derived properties are assigned.

4. IDNA Assumptions and Current Practice

At the time the IDNA2008 documents were written, the assumption was that, if new versions of Unicode introduced incompatible changes, the Standard would be updated to preserve backward compatibility for users of IDNA. For most purposes, this would be done by adding to the table of exceptions associated with Rule G [RFC5892a].

This has not been the practice in the reviews completed subsequent to Unicode 5.2, as discussed in <u>Section 3</u>. Incompatibilities were identified in Unicode 6.0 [RFC6452], in the cumulative review leading to tables for Unicode 11.0 [ID.draft-faltstrom-unicode11] and the subsequent one for Unicode 12.0 [IDNA-Unicode12]. In all of those cases, the decision was made to maintain compatibility with Unicode properties rather than with prior versions of IDNA.

Subsequent to the publication of this document, changes in Unicode detected by algorithmic reviews that would break compatibility with derived properties associated with prior versions of Unicode or that preserve such compatibility within IDNA at the price of departing from current Unicode specifications must be documented (in documents expected to be published as standards track RFCs), explained to, and reviewed by the IETF.

The community has now made decisions and updated tables for Unicode 6.0 [RFC6452], done catch-up work between it and Unicode 11.0 [ID.draft-faltstrom-unicode11], and completed the review and tables for Unicode 12.0 [IDNA-Unicode12]. The decisions made in those cases were driven by preserving consistency with Unicode and Unicode property changes for reasons most clearly explained by the IAB [IAB-Unicode-2018]. Doing things that way is not only at variance with the language in RFC 5892 but is also inconsistent with the intent of commitments to the registry and user communities to ensure

that IDN labels, once valid under IDNA2008, would remain valid and, excepting those that were invalid because they contained unassigned code points, those that were invalid remained invalid.

This document restores and clarifies that original language and intent: absent extremely strong evidence on a per-code point basis that preserving the validity status of possible existing (or prohibited) labels would cause significant harm, Unicode changes that would affect IDNA derived properties are to be reflected in IDNA exceptions that preserves the status of those labels. There is one partial exception to this principle. If the new code point analysis (see Section 3.2) concludes that some code points or collections of code points should be further analyzed, those code points, and labels including them, should be considered unsafe and used only with extreme caution because the conclusions of the analysis may change their derived property values and status.

5. Derived Tables Published by IANA

As discussed above, RFC 5892 specified that derived property tables be provided via an IANA registry. Perhaps because most IANA registries are considered normative and authoritative, that registry has been the source of considerable confusion, including the incorrect assumption that the absence of published tables for versions of Unicode later than 6.0 meant that IDNA could not be used with later versions. That position was raised in multiple ways, not all of them consistent, especially in the ICANN context [ICANN-LGR-SLA].

If the changes specified in this document are not successful in significantly mitigating the confusion about the status of the tables published by IANA, serious consideration should be given to eliminating those tables entirely.

6. Editorial clarification to RFC 5892

In order to avoid this document going forward with remaining known errors or omissions in RFC 5892, this section updates that document to provide fixes to known applicable errata. In particular, verified RFC Editor Erratum 3312 [RFC5892Erratum] provides a clarification to Appendix A and Section A.1 of RFC 5892. That clarification is resolved below.

1. In Appendix A, add a new paragraph after the paragraph that begins "The code point...". The new paragraph should read:

"For the rule to be evaluated to True for the label, it MUST be evaluated separately for every occurrence of the Code point in the label; each of those evaluations must result in True."

2. In Appendix A, Section A.1, replace the "Rule Set" by

7. Acknowledgements

This document was inspired by extensive discussions within the I18N Directorate of the IETF Applications and Real Time (ART) area in the first quarter of 2019 about sorting out the reviews for Unicode 11.0 and 12.0. Careful reviews by Joel Halpern and text suggestions from Barry Leiba resulted in some clarifications.

8. IANA Considerations

For the algorithmic review described in <u>Section 3.1</u>, the IESG is to appoint a Designated Expert [RFC8126] with appropriate expertise to conduct the review and supply derived property tables to IANA. For the code point review, the expertise will be supplied by an IESG-designated expert team as discussed in <u>Section 3.2</u>. In both cases, the experts should draw on the expertise of other members of the community as needed. In particular, and especially if there is no overlap in the people holding the various roles, coordination with the IAB-appointed liaison to the Unicode Consortium will be essential to mitigate possible errors due to confusion.

As discussed in <u>Section 5</u>, and if they have not already done so, IANA is requested to modify the IDNA tables collection [<u>IANA-IDNA-Tables</u>] to identify them clearly as non-normative and in a way that drops the idea of a "current" or "correct" version of those tables, pointing to this document for an explanation. That includes publishing and retaining tables, as supplied by the IETF's Designated Expert, for each new version of Unicode after this document is published, keeping all older versions available. IANA is also requested to change the current title of that registry from "IDNA Parameters", which is misleading, to "IDNA Rules and Derived Property Values".

The "Note" in that registry should also be revised to be consistent with the above, perhaps to say:

"IDNA does not require that applications and libraries, either for registration/storage or lookup, support any particular version of Unicode. Instead, they are required to use derived property values based on calculations associated with whatever version of Unicode they are using elsewhere in the application or library. For the convenience of application and library developers and others, the IETF has supplied, and IANA maintains, derived property tables for several version of Unicode as listed below. It should be stressed that these are not normative that, in principle, an application can do its own calculations and these tables can change as IETF understanding evolves. By contrast, the list of code points requiring contextual rules and the associated rules are normative and should be treated as updates to the list in RFC 5892."

As long as the intent is preserved, the specific text is at IANA's discretion.

IANA's attention is called to the introduction, in <u>Section 3.2</u>, of a temporary "under review" category to the PVALID, DISALLOWED, etc., entries in the tables.

Security Considerations

Application of the procedures described in this document and understanding of the clarifications it provides should reduce confusion about IDNA requirements. Because past confusion has provided opportunities for bad behavior, the effect of these changes should improve Internet security to at least some small extent.

Because of the preference to keep the derived property value stable (as specified in RFC 5892 and discussed in Section 4), the algorithm used to calculate those derived properties does change as explained in Section 3. If these changes are not taken into account, the derived property value will change and the implications might have negative consequences, in some cases with security implications. For example, changes in the calculated derived property value for a code point from either DISALLOWED to PVALID or from PVALID to DISALLOWED can cause changes in label interpretation that would be visible and confusing to end users and might enable attacks.

10. References

10.1. Normative References

[IANA-IDNA-Tables]

Internet Assigned Numbers Authority (IDNA), "IDNA Parameters", March 2019,

<https://www.iana.org/assignments/idna-tables-11.0.0/idna-tables-11.0.0.xhtml>.

This documents make changes to this registry and a way that could change the title, the URL, or both. This citation is to be version published on 2019-03-31. It may be appropriate to supply a citation to the finished version when this document is published.

[RFC5892a]

Faltstrom, P., Ed., "The Unicode Code Points and Internationalized Domain Names for Applications (IDNA)", RFC 5892, August 2010, http://www.rfc-editor.org/rfc/rfc5892.txt.

Section 2.7

[RFC5892Erratum]

"RFC5892, "The Unicode Code Points and Internationalized
Domain Names for Applications (IDNA)", August 2010, Errata
ID: 3312", Errata ID 3312, August 2012,
http://www.rfc-editor.org/errata_search.php/doc/html/rfc5892.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, https://www.rfc-editor.org/info/rfc8126.

[Unicode] The Unicode Consortium, "The Unicode Standard (Current Version)", 2019, http://www.unicode.org/versions/latest/>.

The link given will always access the current version of the Unicode Standard, independent of its version number or date.

[Unicode-properties]

The Unicode Consortium, "The Unicode Standard Version 11.0", 2018,

<https://www.unicode.org/versions/Unicode11.0.0/>.

Section 3.5.

10.2. Informative References

[IAB-Unicode-2018]

Internet Architecture Board (IAB), "IAB Statement on
Identifiers and Unicode", March 2018,
< https://www.iab.org/documents/
correspondence-reports-documents/2018-2/
iab-statement-on-identifiers-and-unicode/>.

[IAB-Unicode7-2015]

Internet Architecture Board (IAB), "IAB Statement on Identifiers and Unicode 7.0.0", February 2015, https://www.iab.org/documents/
correspondence-reports-documents/2015-2/
identifiers-and-unicode-7-0-0/>.

[ICANN-LGR-SLA]

Internet Corporation for Assigned Names and Numbers
(ICANN), "Proposed IANA SLAs for Publishing LGRs/IDN
Tables", June 2019, https://www.icann.org/public-comments/proposed-iana-sla-lgr-idn-tables-2019-06-10-en.

Captured 2019-06-12. In public comment until 2019-07-26.

[ID.draft-faltstrom-unicode11]

Faltstrom, P., "IDNA2008 and Unicode 11.0.0", March 2019, https://datatracker.ietf.org/doc/draft-faltstrom-unicode11/.

[IDNA-Unicode12]

Faltstrom, P., "IDNA2008 and Unicode 12.0.0", March 2019, https://datatracker.ietf.org/doc/draft-faltstrom-unicode12/.

This document is in the RFC Editor queue at of 2019-06-09. Update to RFC reference if/when appropriate.

[IDNA-Unicode7]

Klensin, J. and P. Falstrom, "IDNA Update for Unicode 7.0.0", January 2015, https://datatracker.ietf.org/doc/draft-klensin-idna-5892upd-unicode70/03/>.

Note that this is an historical reference to a superseded document. There is nothing "in progress" about it.

[RegRestr]

Klensin, J. and A. Freytag, "Internationalized Domain Names in Applications (IDNA): Registry Restrictions and Recommendations", July 2019, https://datatracker.ietf.org/doc/draft-klensin-idna-rfc5891bis/>.

- [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.
- [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>.
- [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.

Appendix A. Summary of Changes to RFC 5892

Other than the editorial correction specified in <u>Section 6</u> all of the changes in this document are concerned with the reviews for new versions of Unicode and with the IANA Considerations in <u>Section 5</u>, particularly <u>Section 5.1</u>, of <u>RFC 5982</u>. Whether the changes are

substantive or merely clarifications may be somewhat in the eye of the beholder so the list below should not be assumed to be comprehensive. At a very high level, this document clarifies that two types of review were intended and separates them for clarity and restores the original (but so far unobserved) default for actions when code point derived properties change. For this reason, this document effectively provides a replacement for Section 5.1 of RFC 5892 and adds or changes some material needed to have the replacement be clear or make better sense.

Changes or clarifications that may be considered important include:

- o Separated the new Unicode version review into two explicit parts and provided for different review methods and, potentially, asynchronous outcomes.
- o Specified a review team, not a single expert, for the code point review.
- o Eliminated the de facto requirement for the (formerly single)
 Designated Expert to be the same person as the IAB's Liaison to
 the Unicode Consortium but called out the importance of
 coordination.
- o Created an explicit provision for an "under review" entry in the IANA tables so that, if there is ever again a need to tell the community to wait until the IETF sorts things out, that will be about selected potentially problematic code points and not whole Unicode versions.
- o In part because Unicode is now on a regular one-year cycle rather than producing major and minor versions as needed, to avoid overloading the IETF's i18n resources, and to avoid generating and storing IANA tables for trivial changes (e.g., the single new code point in Unicode 12.1), the review procedure is applied only to major versions of Unicode unless exceptional circumstances arise and are identified.

Appendix B. Change Log

RFC Editor: Please remove this appendix before publication.

B.1. Changes from version -00 (2019-06-12) to -01

- o Added a note about the relationship to <u>draft-klensin-idna-rfc5891bis</u>.
- o Adjusted references per discussion with RFC Editor.

o Minor editorial corrections and improvements.

B.2. Changes from version -01 (2019-07-06) to -02

o Removed an unnecessary reference and a duplicate one.

B.3. Changes from version -02 (2019-07-22) to -03

- o Addition of text to <u>Section 3</u> to clarify IESG responsibilities.
- o Very small editorial changes in response to AD review.

Authors' Addresses

John C Klensin 1770 Massachusetts Ave, Ste 322 Cambridge, MA 02140 USA

Phone: +1 617 245 1457 Email: john-ietf@jck.com

Patrik Faltstrom Netnod Franzengatan 5 Stockholm 112 51 Sweden

Phone: +46 70 6059051 Email: paf@netnod.se