

Subject: Re: [UA-Measurement] How does UTS #46 differ from IDNA 2008? [was: Re: Background on "Characterize how much Android platform limits acceptance of IDNs in web browsing" (FY23 M4)]
Date: 15 August 2022 Monday 05:47:52 GMT+03:00
From: UA-Measurement on behalf of Marc Blanchet via UA-Measurement
To: Jim DeLaHunt
CC: ua-measurement@icann.org

> Le 14 août 2022 à 19:50, Jim DeLaHunt via UA-Measurement <ua-measurement@icann.org> a écrit :

>

>

> On 2022-08-14 08:21, Marc Blanchet wrote:

>>> Le 14 août 2022 à 01:04, Jim DeLaHunt via UA-Measurement <ua-measurement@icann.org> a écrit :

>>>

>>> Marc:

>>>

>>> On 2022-08-08 07:06, Marc Blanchet via UA-Measurement wrote:

>>>> The end result is that, especially in 2022, we must not use UTS#46 and stick with pure IDNA2008

>>> Interesting.

>>>

>>> Does this statement apply just to software developers, or just to domain name registries, or both?

>> Both.

> I am realising one thing that bothers me about this topic. At UASG we spend effort measuring software which does not conform to the practices we support, and mitigating software, but I don't notice us measuring domain names and mitigating registries nearly as much. If the answer is "both", then maybe we should be doing more about domain names.

>>> For instance, it seems that the primary attraction to software developers of UTS#46 over IDNA2008 is if registries are not compliant with IDNA2008 —

>> Recipe for failure. Owners of these domains do not have global interoperability therefore their domains do not work all the time with all software. There are standards for a purpose: to insure global interoperability.

> I wish for that, and you wish for that, but does it reflect the way the world actually is? Remember, the starting point for this conversation is an observation that the leading consumer of domain names on a platform — Chrome on Android — conforms to IDNA2003 not IDNA2008. Maybe from the point of view of a domain owner looking at this situation, it is IDNA2008 which is the recipe for failure, and they make a rational choice to register domain names based on IDNA2003.

>>> especially if registries use the four so-called "Deviation Characters". If registries completely stopped using those characters in name labels, does software which follows UTS#46 behave the same as software which follows IDNA2008? And if registries do not comply with IDNA2008, but software does, is that a better experience for end users than if the software complied with UTS#46? And/or, does the conflict lead to the registries moving to IDNA2008 sooner?

>> Best experience for end users is predictability: when I use a domain name, it works for any software I'm using. Do we agree? Therefore, anything else is not good user experience. Simple as that.

>

> Do these responses really answer my questions? It seems they ignore the questions and assert a dogma.

No dogma. It is just how Internet has been built and has been successful. If we can exchange email between you and me, it is because all the software and service providers involved between the two endpoints conform to the same SMTP standard.

>

> The premise of the question is a flawed world which follows IDNA2003 in some cases, even though we prefer that they would follow IDNA 2008. Given that flawed world, what are the interests of software developers and website operators, taking into account the world as it actually is?

I'm one of the co-author of the IDNA2003 RFCs and I was the co-chair of the IETF idn working group which

defined IDNA (which was later named as IDNA2003). Then, after some years, community realized many issues with IDNA2003 and defined a new version, a better version, later named IDNA2008. That second working group was chaired by Vint Cerf, known as the father of the Internet... The group at that time was well aware of the current state of deployment and support of then IDNA2003, including the fact that the new version (IDNA2008) had (very very very few) incompatibilities with IDNA2003. They took a significant amount of time discussing those with the pros and cons, under the guidance of Vint. At the end, IDNA2008 was defined and IDNA2003 was deprecated, with the idea that no "transition" method was the best path forward. Almost 99,99...% of the domains at that time were compatible with IDNA2008 so no issue whatsoever. Unicode consortium did, by their own, without working together with IETF, define a "transition" scheme which as I wrote, essentially end results in creating a third version of IDNs. The community did agree to move to IDNA2008, including ICANN, but not UTS46. So it is not dogma. The only way we can make this technology working properly is to only support IDNA2008. Anything else will cause/is causing harm. User Acceptance, User experience, ... is not about to support stuff that is causing issues but to let users have the best experience by using the only standard that works.

The right thing to do for the UA community is to work as hard as possible to have everybody moves to IDNA2008 (only). One way with open source software is to work hard by developing code/fixing bugs and submit them to the maintainers with appropriate documentation. This is how it works in open-source world. Lobbying to corporations may work, but it is usually much much more expensive in time and money, and much less chance of success.

Marc.

>

>>> Bringing the Measurement aspect into this, do we have any measure of how many domain name labels have contents that provoke different behaviour from software which implements IDNA2003, UTS#46, and IDNA2008? What are the most significant of those labels? How much traffic do they see?

>> I'm sure some people have gathered some data about this. But it could also be an interesting study for UA Measurement.

>>

>> Marc.

>>

>>> Best regards,

>>> —Jim DeLaHunt

>>>

>>> On 2022-08-08 07:06, Marc Blanchet via UA-Measurement wrote:

>>>> Le 5 août 2022 à 15:45, Jim DeLaHunt via UA-Measurement <ua-measurement@icann.org> a écrit :

>>>>>

>>>>> Thank you, Julien. That link [1] to Table 4 is very helpful. It in turn links to the Unicode FAQ, "Q: Are the local mappings in IDNA2008 just a UI issue?" [2], which is also useful.

>>>>>

>>>>> What I take away from the UTS#46, especially Table 4, is not that implementing IDNA 2008 is better for UA and implementing UTS#46 is worse for UA. It is that implementing IDNA 2008 is good for UA, and, UTS#46 has solutions to some hard problems and complexities involved in implementing IDNA 2008, so implementing IDNA 2008 according to UTS#46 is also good for UA.

>>>> In a nutshell, IDNA2003, IDNA2008 and UTS#46 are three different versions of IDN. They have a common set of code points and some specific ones for each of them. The IETF and ICANN have stated clearly that the only standard is IDNA2008. While Unicode people could had good intentions for the "transition between IDNA2003 and IDNA2008", that transition was studied carefully by IETF and the decision was that it was worst to have transitioning method than just deprecate IDNA2003, given that IDN2003 was still not too deployed. Now the more people are implementing UTS#46, the more we are in trouble.

>>>>

>>>> The end result is that, especially in 2022, we must not use UTS#46 and stick with pure IDNA2008. This is the only way to guarantee conformance and interoperability.

>>>>

>>>> Regards, MARc.

>>>>

>>>>

>>>>

>>>> UTS#46 also points out two things I find interesting.

>>>>

>>>> 1. UA and IDNA 2008 depend on domain name registries to make wise choices. UTS #46 points this out in detail. Some of the problems involve the four so-called "Deviation Characters": U+200C ZERO WIDTH NON-JOINER, U+200D ZERO WIDTH JOINER, U+00DF (ß) LATIN SMALL LETTER SHARP S, and U+03C2 (ς) GREEK SMALL LETTER FINAL SIGMA. It matters a lot that managers of domain name registries make wise policies about how they handle names in their domain which involve these characters. But domain name registries can make unwise choices about other characters also.

>>>>

>>>> 2. There are a lot more domain name registries than we might think of at first. It is not just the registries of top-level domains, those whom ICANN has influence over. It is also third-level domain names and below. Consider all the blogs registered with domain names foo.blogspot.com, or foo.blogspot.de. The managers of Blogspot (Google employees?) have domain name registries. I manage a domain name registry under jdlh.com, and another under jdlh.palo-alto.ca.us. And so on. Maybe we should add "operators of lower-level domain name registries" as another stakeholder group to whom we pay attention for Universal Acceptance?

>>>>

>>>> Best regards,

>>>> —Jim DeLaHunt

>>>>

>>>> [1]

[https://urldefense.com/v3/__https://unicode.org/reports/tr46/*Table_IDNA_Comparisons__;lw!!PtGJab4!8E0UdxWjx_ik12q8II7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7oIWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PYUdMhDAQ\\$](https://urldefense.com/v3/__https://unicode.org/reports/tr46/*Table_IDNA_Comparisons__;lw!!PtGJab4!8E0UdxWjx_ik12q8II7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7oIWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PYUdMhDAQ$) [unicode[.]org]

>>>> [2]

[https://urldefense.com/v3/__https://www.unicode.org/faq/idn.html*20__;lw!!PtGJab4!8E0UdxWjx_ik12q8II7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7oIWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PYxczbZrQ\\$](https://urldefense.com/v3/__https://www.unicode.org/faq/idn.html*20__;lw!!PtGJab4!8E0UdxWjx_ik12q8II7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7oIWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PYxczbZrQ$) [unicode[.]org]

>>>>

>>>> On 2022-08-05 06:43, Julien Bernard wrote:

>>>>> I don't have the full background but it seems that UTS#46 was created to ensure the maximum compatibility with IDNA 2003 and IDNA 2008 to ease the transition. Table 4 in [1] summaries the differences between those 3 specifications. The full post contains many information.

>>>>>

>>>>> [1]

[https://urldefense.com/v3/__https://unicode.org/reports/tr46/*Table_IDNA_Comparisons__;lw!!PtGJab4!8E0UdxWjx_ik12q8II7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7oIWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PYUdMhDAQ\\$](https://urldefense.com/v3/__https://unicode.org/reports/tr46/*Table_IDNA_Comparisons__;lw!!PtGJab4!8E0UdxWjx_ik12q8II7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7oIWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PYUdMhDAQ$) [unicode[.]org]

>>>>>

>>>>>

>>>>> Best regards,

>>>>> Julien

>>>>>

>>>>> On 04/08/2022 22:56, Jim DeLaHunt wrote:

>>>>>> Julien:

>>>>>>

>>>>>>> Thank you for your comments on the thread about "Characterize how much Android platform limits acceptance of IDNs in web browsing" (FY23 M4).

>>>>>>>

>>>>>>> I have a question about one of your side comments:

>>>>>>>

>>>>>>> On 2022-08-04 13:07, Julien Bernard wrote:

>>>>>>>> One relevant point is that ICU ... has implemented IDNA 2008 according to ... UTS #46 rather than IDNA 2008.

>>>>>>>> Is there an explanation somewhere about the difference between IDNA 2008 and UTS #46? In reading UTS #46, I got the impression it was saying something like, "everyone should comply with IDNA 2008 not 2003. However, IDNA 2008 is not fully specified and does not give enough help to a developer moving from IDNA 2003 to IDNA 2008. Here are some instructions for making that transition. You end up with correct IDNA 2008."

>>>>>>>>

>>>>>>>> I might well be misunderstanding what UTS #46 says. I could spend some time reading UTS #46 and

IDNA 2008 side by side. But if there is a paper which explains what the difference is, that would save me time in seeing how my understanding is wrong.

>>>>>>

>>>>>> Can you suggest some reading for me?

>>>>>>

>>>>>> Best regards,

>>>>>> —Jim DeLaHunt

>>>>>>

>>>>> --

>>>>> . --Jim DeLaHunt, jdlh@jdlh.com

[https://urldefense.com/v3/_http://blog.jdlh.com/_;!!PtGJab4!8E0UdxWjx_lk12q8ll7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7olWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PYv-0Cobw\\$](https://urldefense.com/v3/_http://blog.jdlh.com/_;!!PtGJab4!8E0UdxWjx_lk12q8ll7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7olWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PYv-0Cobw$) [blog[.]jdlh[.]com]

([https://urldefense.com/v3/_http://jdlh.com/_;!!PtGJab4!8E0UdxWjx_lk12q8ll7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7olWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PZO4Wzig\\$](https://urldefense.com/v3/_http://jdlh.com/_;!!PtGJab4!8E0UdxWjx_lk12q8ll7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7olWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PZO4Wzig$) [jdlh[.]com])

>>>>> multilingual websites consultant

>>>>>

>>>>> 2201-1000 Beach Ave, Vancouver BC V6E 4M2, Canada

>>>>> Canada mobile +1-604-376-8953

>>>>>

>>>>>

>>>>> UA-Measurement mailing list

>>>>> UA-Measurement@icann.org

>>>>> <https://mm.icann.org/mailman/listinfo/ua-measurement>

>>>>>

>>>>> By submitting your personal data, you consent to the processing of your personal data for purposes of subscribing to this mailing list accordance with the ICANN Privacy Policy (<https://www.icann.org/privacy/policy>) and the website Terms of Service (<https://www.icann.org/privacy/tos>). You can visit the Mailman link above to change your membership status or configuration, including unsubscribing, setting digest-style delivery or disabling delivery altogether (e.g., for a vacation), and so on.

>>>>>

>>>>> UA-Measurement mailing list

>>>>> UA-Measurement@icann.org

>>>>> <https://mm.icann.org/mailman/listinfo/ua-measurement>

>>>>>

>>>>> By submitting your personal data, you consent to the processing of your personal data for purposes of subscribing to this mailing list accordance with the ICANN Privacy Policy (<https://www.icann.org/privacy/policy>) and the website Terms of Service (<https://www.icann.org/privacy/tos>). You can visit the Mailman link above to change your membership status or configuration, including unsubscribing, setting digest-style delivery or disabling delivery altogether (e.g., for a vacation), and so on.

>>>>> --

>>>>> . --Jim DeLaHunt, jdlh@jdlh.com

[https://urldefense.com/v3/_http://blog.jdlh.com/_;!!PtGJab4!8E0UdxWjx_lk12q8ll7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7olWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PYv-0Cobw\\$](https://urldefense.com/v3/_http://blog.jdlh.com/_;!!PtGJab4!8E0UdxWjx_lk12q8ll7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7olWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PYv-0Cobw$) [blog[.]jdlh[.]com]

([https://urldefense.com/v3/_http://jdlh.com/_;!!PtGJab4!8E0UdxWjx_lk12q8ll7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7olWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PZO4Wzig\\$](https://urldefense.com/v3/_http://jdlh.com/_;!!PtGJab4!8E0UdxWjx_lk12q8ll7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChtwOhbv7olWC0yJeq67dsa2MAy_sJ-A_dmiXl8_PZO4Wzig$) [jdlh[.]com])

>>>>> multilingual websites consultant

>>>>>

>>>>> 2201-1000 Beach Ave, Vancouver BC V6E 4M2, Canada

>>>>> Canada mobile +1-604-376-8953

>>>>>

>>>>>

>>>>> UA-Measurement mailing list

>>>>> UA-Measurement@icann.org

>>>>> <https://mm.icann.org/mailman/listinfo/ua-measurement>

>>>>>

>>>>> By submitting your personal data, you consent to the processing of your personal data for purposes of subscribing to this mailing list accordance with the ICANN Privacy Policy (<https://www.icann.org/privacy/policy>) and the website Terms of Service (<https://www.icann.org/privacy/tos>). You can visit the Mailman link above to

change your membership status or configuration, including unsubscribing, setting digest-style delivery or disabling delivery altogether (e.g., for a vacation), and so on.

>

> --

> . --Jim DeLaHunt, jdlh@jdlh.com

[https://urldefense.com/v3/__http://blog.jdlh.com/__;!!PtGJab4!8E0UdxWjx_lk12q8II7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChwOhbv7oIWCOyJeq67dsa2MAy_sJ-A_dmiXl8_PYv-0Cobw\\$](https://urldefense.com/v3/__http://blog.jdlh.com/__;!!PtGJab4!8E0UdxWjx_lk12q8II7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChwOhbv7oIWCOyJeq67dsa2MAy_sJ-A_dmiXl8_PYv-0Cobw$) [blog[.]jdlh[.]com]

([https://urldefense.com/v3/__http://jdlh.com/__;!!PtGJab4!8E0UdxWjx_lk12q8II7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChwOhbv7oIWCOyJeq67dsa2MAy_sJ-A_dmiXl8_PZO4Wzig\\$](https://urldefense.com/v3/__http://jdlh.com/__;!!PtGJab4!8E0UdxWjx_lk12q8II7IG2P8vPW5_aC18c9Dwc3BUO-Sk2beUChwOhbv7oIWCOyJeq67dsa2MAy_sJ-A_dmiXl8_PZO4Wzig$) [jdlh[.]com])

> multilingual websites consultant

>

> 2201-1000 Beach Ave, Vancouver BC V6E 4M2, Canada

> Canada mobile +1-604-376-8953

>

>

> UA-Measurement mailing list

> UA-Measurement@icann.org

> <https://mm.icann.org/mailman/listinfo/ua-measurement>

>

> By submitting your personal data, you consent to the processing of your personal data for purposes of subscribing to this mailing list accordance with the ICANN Privacy Policy (<https://www.icann.org/privacy/policy>) and the website Terms of Service (<https://www.icann.org/privacy/tos>). You can visit the Mailman link above to change your membership status or configuration, including unsubscribing, setting digest-style delivery or disabling delivery altogether (e.g., for a vacation), and so on.

UA-Measurement mailing list

UA-Measurement@icann.org

<https://mm.icann.org/mailman/listinfo/ua-measurement>

By submitting your personal data, you consent to the processing of your personal data for purposes of subscribing to this mailing list accordance with the ICANN Privacy Policy (<https://www.icann.org/privacy/policy>) and the website Terms of Service (<https://www.icann.org/privacy/tos>). You can visit the Mailman link above to change your membership status or configuration, including unsubscribing, setting digest-style delivery or disabling delivery altogether (e.g., for a vacation), and so on.