# Unicode Email Addresses: grå@grå.org

What, Why, How.

ICANN

May 2024, SLAC, Arnt Gulbrandsen
UA Technology Senior Manager

# About Arnt Gulbrandsen

◉ Work for ICANN's UA team:

  Fix open source. BTW - We're hiring. Come talk to me.

  Diagnose and fix interop problems.

  Inform DNS registries etc. how to use IDNs (grå.org) and EAI ([grå@grå.org](mailto:grå@grå.org))

◉ Wrote about ten IMAP/mail/DNS RFCs.

◉ Helped work on most mail RFCs since 2000.

◉ Questions about any of those RFCs welcome today.

◉ Norwegian, live in Munich, office address ICANN Brussels.

◉ Linux user since August 1992.

# What is ICANN?

# Some of What ICANN Organization Does

**Domain Name System**

The domain name system provides addressing for the Internet so people can find websites, send email, and other tasks. The ICANN organization also supports the stability of the DNS through its work, and also its contracts and accreditations.

**POLICY**

**Policy Development**

The ICANN organization supports inclusive, open and transparent multi-stakeholder bottom-up consensus based policy development mechanisms.

**L-Root**

The ICANN organization hosts and supports one of the 13 L-Root infrastructures. At over 161 locations worldwide, L-Root is critical to infrastructure that helps reduce latency and improves performance of the DNS.

**Support and Grow the Community**

The ICANN organization engages, nurtures and supports interested stakeholders for active and meaningful participation in ICANN. ICANN connects with stakeholders through outreach and engagement, and meeting & event support.
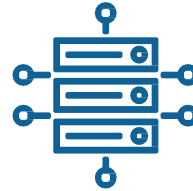
**Generic Top-Level Domains**

The ICANN organization manages the domain name system's top-level domains. ICANN helps promotes competition and choice in the gTLD marketplace.

**Country Code Top-Level Domains**

The ICANN organization delegates top-level domains identified with a country code. Management is done by national ccTLD operators.

**Protocol Parameters**

The ICANN organization, in coordination with the Internet Engineering Task Force, manages protocol parameters by maintaining many of the codes and numbers used in Internet protocols.

**Internet Protocol Addresses**

By serving as the central repository for IP addresses, the ICANN organization helps coordinate how IP addresses are supplied – preventing repetition and conflicts

**Root Zone Management**

The ICANN organization helps manage the root zone through the IANA functions, which involves assigning the operators of top-level domains, such as .bank and .com, and maintaining the technical and administrative details.

IANA functions

# Introduction to Internationalized Domain Names

- Now possible to have domain names in almost all languages and scripts.

- The German company Kärcher has kärcher.de (AKA xn--krcher-bua.de) as well as kaercher.com and other domains.

- kärcher.de is an internationalized domain name (IDN): It contains ä, which isn't ASCII.

- Some companies have many. The Italian coffee maker Illy has illy.com, illy.it, illycaffe.it, illycaffè.it (note the è), …

- IDNs are usable for web sites (with a few bugs) and email (with more bugs).

- Email needs more than just IDNs.

- IDNs can be used for other purposes in principle:
  - No uses exist today, because ordinary people only see web+email.

- In Germany, ä ≅ ae, elsewhere most people probably think ä ≅ a.

- Kärcher solved this by getting several domains:
  - karcher.com
  - kaercher.com
  - karcher.de
  - kaercher.de
  - kärcher.de
  - …

- Most web sites redirect to karcher.com

- Good strategy for many customers
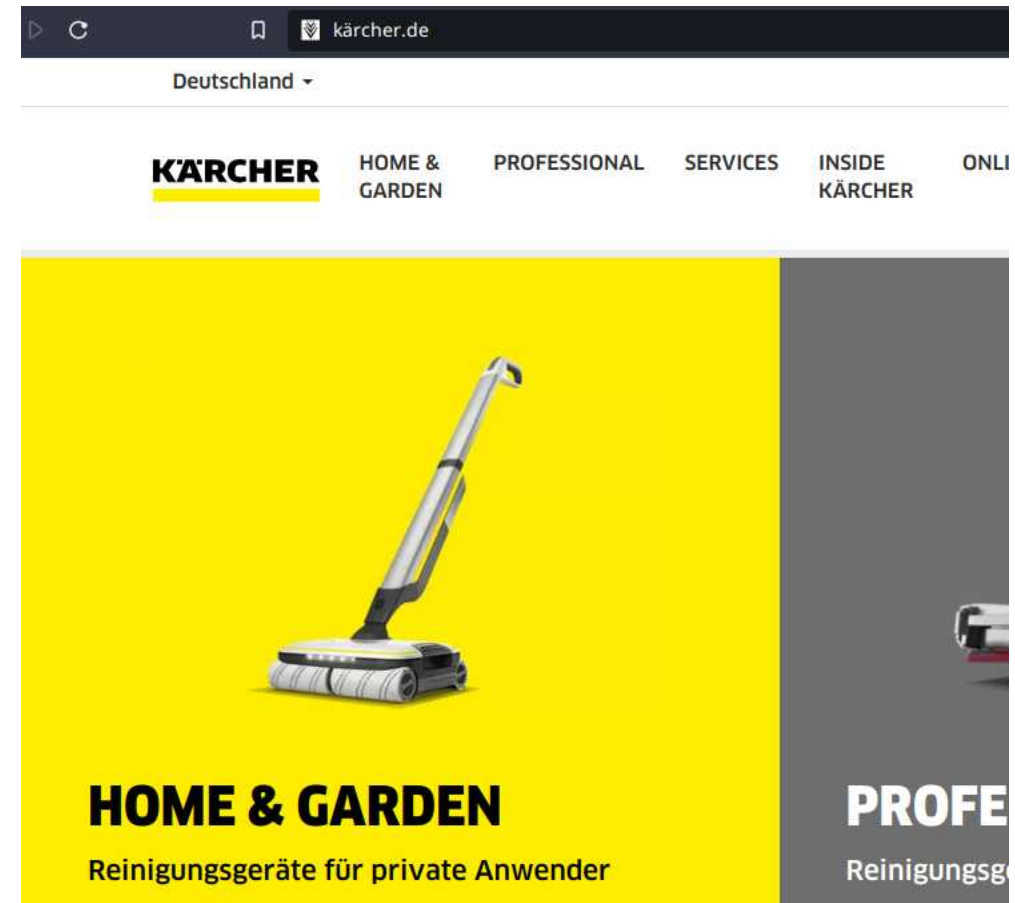
- Small business that targets Munich, which is called München in München

- Locals prefer the spelling München, also know the spelling muenchen.

- This business has the same content on two sites
  - One with –muenchen
  - One with –münchen

- Web browsers stay on the site they entered, there's no redirect

- Good strategy for some local audiences
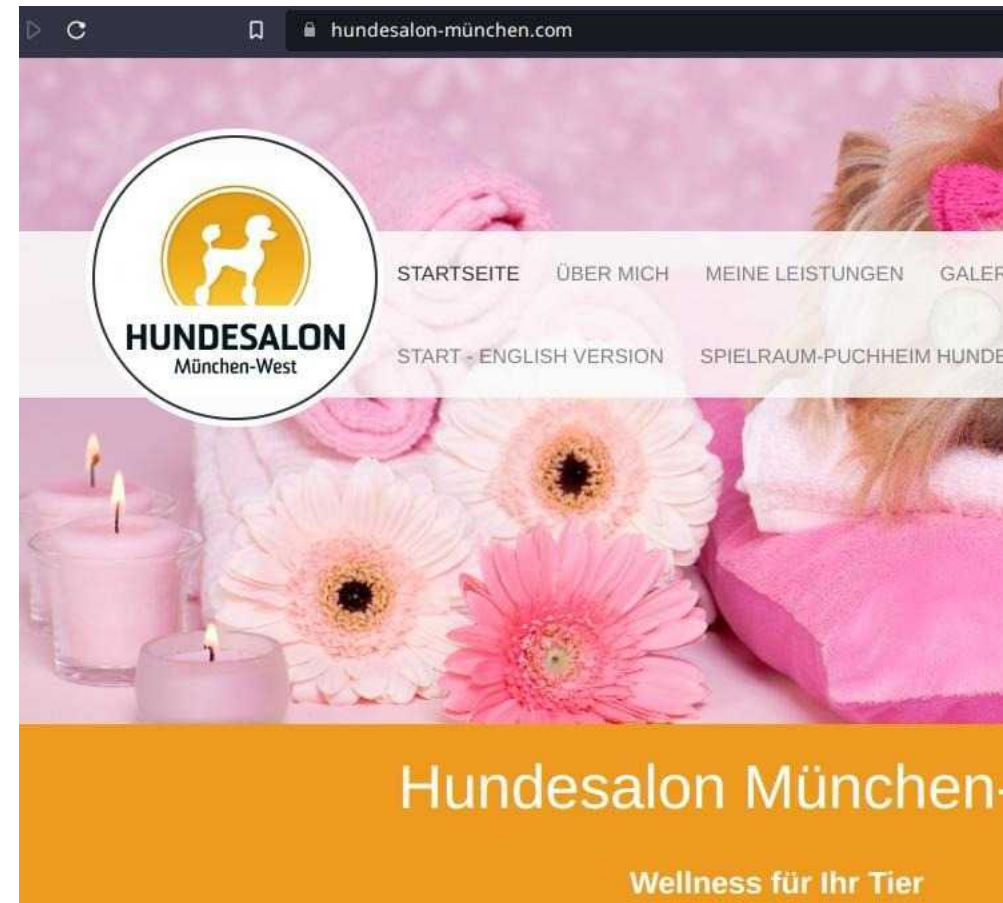
# IDN Example: example of browser bug

- The address bar shows -münchen

- This adblocker (shown on the right) shows –mnchen-e3b instead

- Most people will not understand –e3b

- Best regarded as a browser bug

- -e3b is part of an "A-label", which we'll explain later, and seeing A-labels in unexpected places is the most common bug

- Annoying and confusing for end-users
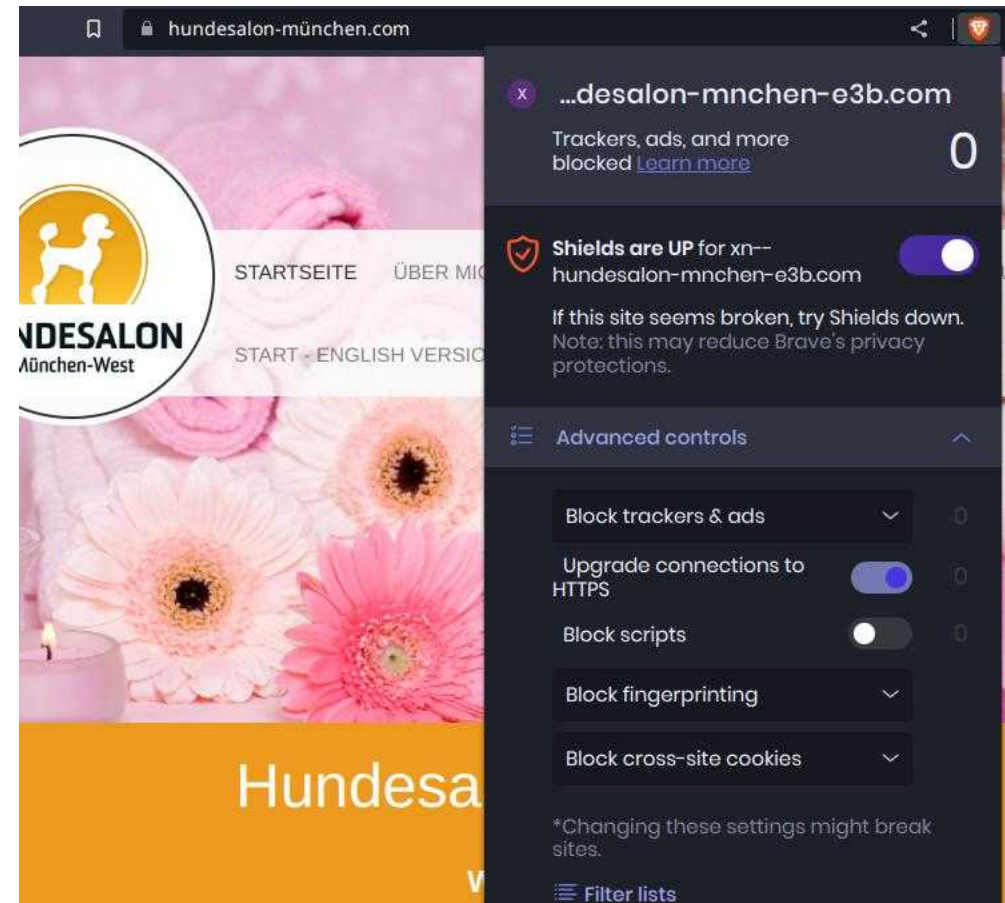
- Not just companies can have more than one domain, countries can too.

- On the right, CloudIT in Bahrain has an English-language .bh site and an Arabic-language البحرين site.

- Mostly, the two are independent.

- Registering a .bh domain does not give an automatic right to the corresponding البحرين domain.
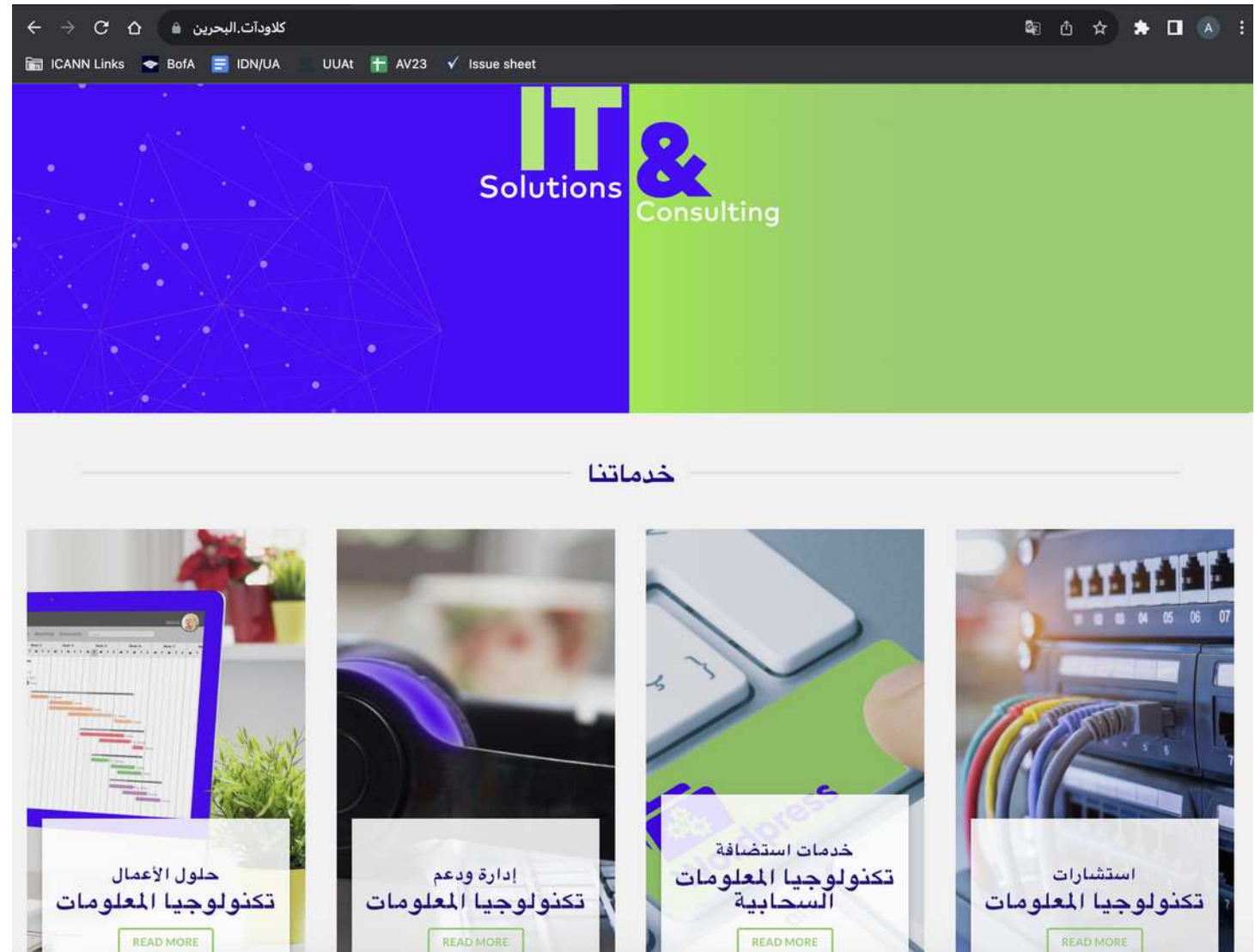
# Unicode Addresses in Email

# Overview of EAI (Email Address Internationalization)

◉ UTF8 is permitted everywhere.

◉ Quoted-printable, RFC2047 etc. are optional.

◉ Sending to someone who doesn't support it causes a bounce.

◉ 75+% support it now.

◉ Microsoft and Google improve the average, open source decreases average.

# Message Example

◉ 100% UTF8 in this message:

To: Jøran Øygårdvær <jøran@blåbærsyltetøy.gulbrandsen.priv.no>
Subject: Høy på pæra
Content-Type: text/plain; charset=utf8

Gørrlei av eksempler.

◉ There are two syntax errors in the message above.

*Can you find them?*

# Message Example (continued)

◉ This screenshot is from Gmail and shows a message from Datamail.

◉ Datamail is an Indian company, has forked various open source, added EAI support and sells to Indian customers.

◉ Note Gmail's support level:

   Gmail can receive mail from Indian addresses.

   Gmail can send mail to Indian addresses.

   Gmail cannot host Indian domains.
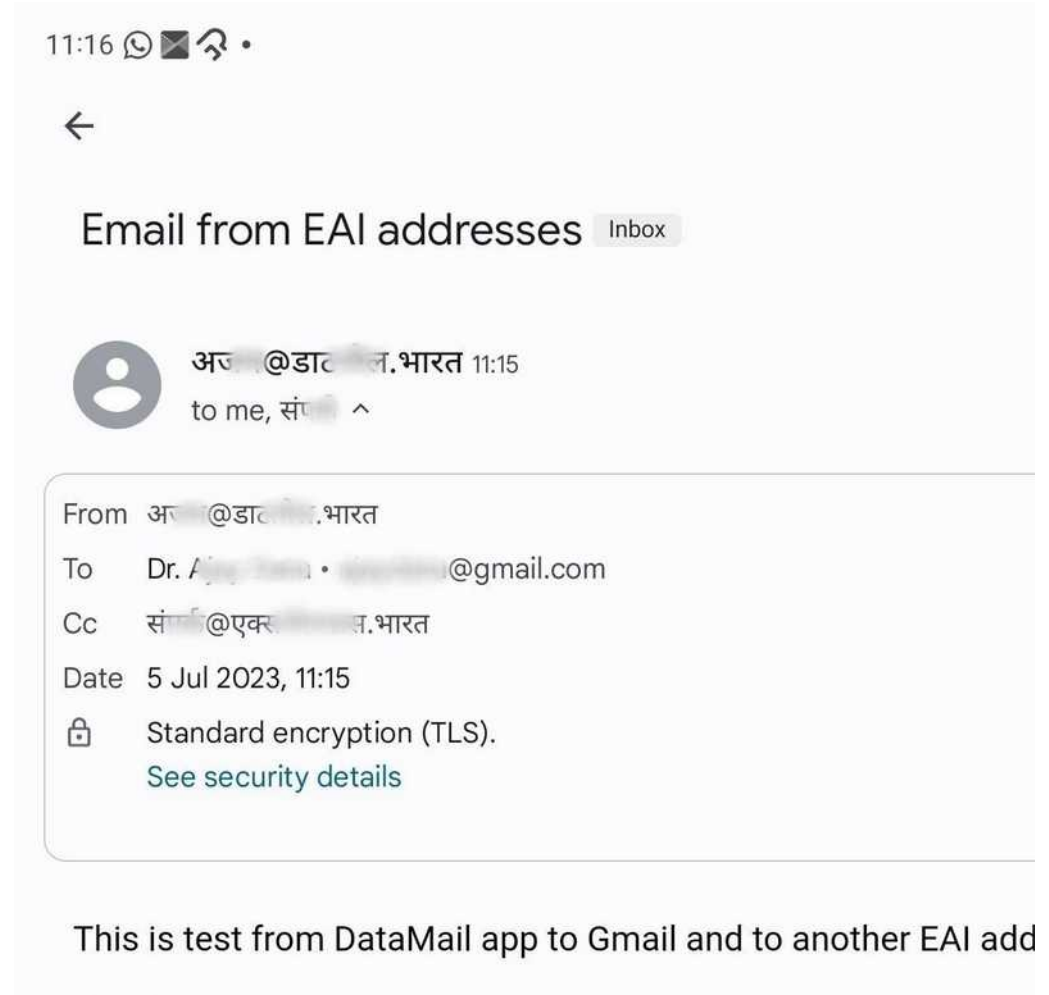
# An SMTP Transaction, 1/2



S                   R
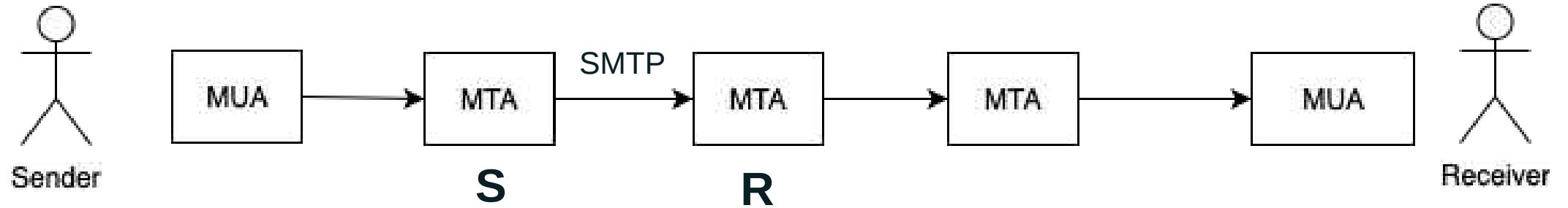
Server S forwarding an email to server R

S: <connect>

R: 220 receive.net ESMTP

Specific SMTPUTF8 Signaling (EAI support)

S: EHLO sender.org

R: 250-8BITMIME

R: 250-**SMTPUTF8**

R: 250 PIPELINING

S: MAIL FROM:< 猫王 @ 普遍接受 - 测试 . 世界 > **SMTPUTF8**

R: 250 Sender accepted

S:RCPT TO:<ray@receive.net>

R:250 Recipient accepted

S:DATA
R:354 Send your message
S:From: 猫王 ＜猫王 @ 普遍接受 - 测试 . 世界 ＞
S:To: ray@receive.net
S:Subject: 我们要吃午饭吗？
S:
S:How about lunch at 12:30?
S:.
R:250 Message accepted 389dck343fg34
S:QUIT
R:221 Sayonara

Email itself

# An IMAP Connection

S: <connect>
R: * OK [CAPABILITY ... ENABLE ...
S: a login arnt pils
R: a OK [CAPABILITY ... ENABLE ... UTF8=ACCEPT ...
S: b enable **utf8=accept**
R: * ENABLED **UTF8=ACCEPT**
R: b OK done
S: c select"آرنت"
R: ...

# JMAP and Other IMAP-like Protocols

◉ JMAP supports EAI fully as part of the code.

◉ Three proprietary protocols exist, all with EAI support in the core:

   Gmail API used with Gmail, open to all, more or less.

   EWS, used with Exchange and O365, now slightly deprecated.

   Microsoft Graph, replaces EWS.

◉ IMAP4rev2 supports EAI in the core.

◉ POP3 needs an extension, like IMAP4rev1.

# Supporting EAI in Software

# Typical Changes to Support EAI

◉ IMAP clients generally need to add ENABLE support.

◉ The actual IMAP parser often needs no change:

    Example: Code often matches strings using /"[^"]*"/ and just works.

    EAI uses UTF8 heavily, most code today uses UTF8 heavily and just works.

◉ A new function is needed to test whether email addresses use non-ASCII characters.

◉ SMTP/sending support needs to use that new function.

◉ Ruby's standard library is not unusual: a 100-line diff for IMAP ENABLE and tests, a little less for IMAP UTF8=ACCEPT, 200 for SMTP.

◉ Record holders: Procmail and Postfix.

◉ Users of Ruby, Python, Java already have support in the standard libraries.

# Testing

◉ Quick way to test: send mail to my autoresponder grå@grå.org.

◉ Unit tests: git clone github.com:arnt/eai-test-messages

◉ Interop testing: Gmail and O365 have some support.

◉ Open source interop: Courier-IMAP and Thunderbird recommended.

◉ Or Mox (xmox.nl), a new all-in-one server with default-right DMARC etc.

◉ Infrastructure needed generally works:

> Postfix, Exim
> Halon, Momentum
> Amavis, ClamAV

# Typical Bugs

◉ Software often does e.g. RFC2047 encoding on demand: The demand check must be modified.

◉ Case sensitivity checks sometimes need extension to make grå == GRÅ.

◉ Gmail will send to To: grå@grå.org, Exchange will send To: grå@xn--gr-zia.org, software often needs extra code to consider those equal.

◉ Making a search for grå match a message containing info@xn--gr-zia.org can be challenging.

◉ Some code does punycode on demand and accidentally does it on a localpart.

     grå@grå.org is **not** the same as xn--gr-zia@xn--gr-zia.org.

◉ Composing or decomposing other people's addresses is legal, but passing on exactly the bytes you receive triggers fewer bugs.

# Downgrading

◉ Common question:
  If the recipient does not support EAI, how do you downgrade for compatibility?

◉ Three simple answers: You generally can't, in two cases you can, and it seldom matters.

    If you're the composing software and have two addresses, you can choose which address to use as sender by looking at the recipients.

    In some cases it makes sense to fall back to a noreply address.

◉ A testbed (RFC5738, around 2010) included downgrades, and it was a net negative for overall utility.

◉ The current RFCs say: If you can't deliver, return the message

# Some Relevant Materials

◉ See https://uasg.tech for list of reports.

◉ RFC 6530 is the overview RFC covering EAI

◉ RFCs 6531-3, 6855 and 6858 are relevant

◉ Questions: arnt.gulbrandsen@icann.org or uaprogram@icann.org

# Engage with ICANN – Thank You and Questions

One World, One Internet

**ICANN**

Visit us at **icann.org**

@icann

facebook.com/icannorg

youtube.com/icannnews

flickr.com/icann

linkedin/company/icann

slideshare/icannpresentations

soundcloud/icann