Guidelines for Internationalized Email Clients
draft-ietf-eai-email-clients-01

Abstract

   This document provides some guidelines for email clients that support
   Email Address Internationalization (EAI) as outlined in [RFC6530]. A
   number of interoperability cases between different versions of email
   components are reviewed. Recommendations are made to improve
   interoperability and usability and to minimize discrepancies between
   the display of composed and received email in different language
   environments.

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 http://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 11, 2013.

Copyright Notice

Table of Contents

1. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in [RFC2119].

2. Introduction

[RFC6530] Overview and Framework for Internationalized Email
describes changes to electronic mail (email) to fully support
internationalized characters. The fundamental change is to remove
the ASCII only restriction on email addresses and allow them to
contain UTF-8 characters. Additional documents provide detailed
specifications for the extensions required to email headers [RFC6532]
and to the protocols SMTP [RFC6531], POP [I-D.ietf-eai-rfc5721bis]
and IMAP [I-D.ietf-eai-5738bis].

This document provides guidelines for email clients that support
these specifications for Email Address Internationalization (EAI).
It does not introduce any protocol extensions that are not defined in
the above documents. It highlights the extensions that are important
to the design and implementation of email clients and makes a number
of recommendations intended to improve interoperability and
usability.

3. Terminology

A number of different acronyms are typically used to describe the
major functional components of email.

Mail User Agent (MUA)
Message Submission Agent (MSA)
Message Transfer Agent (MTA)
Message Delivery Agent (MDA)
Message Store (MS)

The architecture of modern email systems can range from simple, with
all components running on one server, to very complex, with
components being distributed across multiple, geographically
dispersed machines. Nevertheless, the above terminology is generally
sufficient to represent different architectures from a functional
point of view. For a comprehensive description of email architecture
see [RFC5598].

sender -> MUA -> MSA -> MTA
                        ...
                        MTA -> MDA -> MS -> PIF -> MUA -> recipient

In this context, an "Email Client" is an MUA that has an interface to an MSA to send email and an interface to the MS to retrieve email. The interface to retrieve mail (PIF) is a POP or IMAP server or direct access to the File system. The MUA also provides a User Interface (UI) that allows an end user to read (display) and write (compose) their email.

A common email architecture includes the MSA function within the MTA. An improved architecture that better addresses security concerns is a separate MSA component as shown here [RFC6409], [RFC5068].

"SMTPUTF8" is used to indicate email address internationalization as specified by [RFC6530] and related documents.

"ASCII" refers to the strict 7-bit ASCII character set [ANSI.X3-4.1968].

"UTF-8", Unicode Transformation Format/8-bit is a character encoding scheme that can represent any character in the Unicode standard [RFC3629]. It contains ASCII as a subset.

"message/global" is an email message that contains UTF-8 characters beyond 7-bit ASCII in message headers and/or body parts [RFC6532].

"message/rfc822" is an email message that contains only 7 bit ASCII and does not use any SMTPUTF8 extensions. Note that the original message (as composed by the user) may contain non-ASCII characters that have been encoded into ASCII using IDNA [RFC5890], MIME body encoding [RFC2045] or MIME header encoding [RFC2047].

4. Interoperability

Internationalized Email is not compatible with legacy email systems, those based on prior Internet email standards [RFC5321], [RFC5322]. Non-ASCII email addresses cannot be submitted in legacy SMTP commands like MAIL FROM or RCPT TO. In addition the Internationalized Email standard does not include a method to "downgrade" message/global to message/rfc822.

An Internationalized message cannot be transmitted via SMTP if the receiving MTA does not announce SMTPUTF8 in response to EHLO. There are two failure cases that an email client may have to handle described in Section 3.2 of [RFC6531].

a) If the client is submitting a message to an MSA that does not support SMTPUTF8, the message will be rejected.

b) If the MSA does support SMTPUTF8 but a downstream MTA does not,

then the mail will bounce. That is, a delivery status notification
(DSN) that the mail could not be delivered will be sent back to the
sender.

Incompatibility between Internationalized email and legacy systems is
expected to be important initially during a transition period but
less important over time as more email systems upgrade to support the
SMTPUTF8 extensions. To the extent that this incompatibility is
deemed important at the time an implementation is undertaken, the
email client should provide methods to prevent or at least minimize
these failures.

4.1. Interoperability Scenarios

The following scenarios cover the different cases of sending mail
from an Internationalized server to a legacy server.

'I' indicates an Internationalized address (a non-ASCII address on an
Internationalized mail server).

'IA' indicates an ASCII address on an Internationalized server.

'LA' indicates an address on a Legacy mail server, which must be
ASCII.

Case 1. The simple compatibility case

```
From:    IA1 (or LA1)
To:      LA2
Subject: ...
Body     ...
```

The message will be successfully sent as long as the email client
sends message/rfc822 rather than message/global.

Case 2. The simple incompatibility case

```
From:    I1
To:      LA2
Subject: ...
Body     ...
```

The message will be rejected by the MSA or will bounce from a
downstream SMTP server.

If user I1 also has an ASCII email address IA1 or LA1, there may be a
simple workaround. If the email client supports multiple email
accounts, the user just has to switch the From address to an ASCII

address and it becomes Case 1.

Case 3. The general incompatibility cases

The general case is a mix of Internationalized and legacy addresses.
While many combinations are possible, the two cases below essentially
cover all possibilities.

```
From: I1
To:   LA2
Cc:   I3
```

The message will be sent to I3 but it will bounce from LA2.

Switching the From address to an ASCII address as in Case 2 is not a
solution, as the following case demonstrates.

```
From: IA1 (or LA1)
To:   LA2
Cc:   I3
```

This message will bounce from LA2 since the address in the Cc header
cannot be transmitted to a legacy server.

In these cases, users will likely send the message twice in order to
reach all intended recipients. First, to the original list and then
using an ASCII address to the bounced recipients.

If users know beforehand which addresses are on legacy servers, they
can avoid bounced messages by removing those addresses, but they
still have to send a second email to reach recipients that were
removed.

5. Compatibility Support

An email client can provide support to minimize the incompatibility
problems outlined in Section 4. There may be several ways to do
this. Following are guidelines on some of the ways that this can be
accomplished.

At the very least, to provide basic compatibility between
Internationalized and legacy systems, if all email addresses in the
SMTP envelope and the message headers are ASCII, then a message/
rfc822 should be sent (Case 1 above).

For Case 2, the email client should support multiple email accounts
and allow the user to switch the From address at any time during
composition of the message.

For Case 3, several mechanisms may be required to provide
compatibility support. These are outlined in the following sub-
sections.

5.1. Address Book

Each contact in the address book should be able to have several email
addresses, each of which is configured to be either an
Internationalized or a Legacy address.

The user may not necessarily know if an ASCII address they enter in
their address book is on a legacy server or not. If it is configured
as an Internationalized address and that turns out to be wrong, then
email sent to that contact may bounce. The user can then re-
configure the address as Legacy so the email client can provide
warnings of a possible bounce on subsequent messages.

5.2. Message Mode

Message composition should have "Message Mode" option to specify
"Internationalized Mode" or "Legacy Mode".

If the type of each address in the headers does not conform to the
message mode, then the user is given a warning about those addresses
that don't match the mode. In a graphical user interface this might
be done by setting such addresses to a different color such as red.

The user would typically first change the message mode to see if the
warnings disappear.

When the mode is switched, the email client switches addresses in
message header fields to match the mode, selecting from the list of
addresses in each contact.

There are cases where both modes provide warnings (see Example 5
below). In these cases, the user can remove the addresses that don't
conform to the mode.

For Internationalized mode, the user has an additional option to send
the message anyway, without removing flagged addresses. They would
have to handle bounced messages from Legacy servers later. The
option to send anyway cannot be provided in Legacy mode, as it is not
possible to compose a message/rfc822 if any sender or recipient
address is not ASCII.

Where both modes provide warnings, users will likely want to send the
message in each mode in order to reach all recipients. The email
client should make it easy to do this. There are many possible

designs to accomplish this. The following is one example.

An option is provided when composing email to add a second message
header section in the other mode that allows the user to move
addresses between sections. This is in addition to making individual
changes to address headers as in normal email composition. The
Subject and Body are common so the user can compose a single message
but have it sent in the two different modes to different recipients.

Following is an example of this for Case 3 above.

```
---------------------------------------------
Legacy                    Internationalized
From: IA1                 From: I1
To:   LA2      <--->      To:
Cc:            <--->      Cc:   I3
---------------------------------------------
Subject: ...
Body:    ...
---------------------------------------------
```

5.3. Message Format

In Internationalized Mode, mail should be sent as message/global.
The aim of Internationalized Email is 8 bit clean messages using
UTF-8 encoding to represent Unicode characters in header fields and
the message body.

In Legacy Mode, mail must be sent as message/rfc822. This may
include non-ASCII characters that are encoded into ASCII using MIME
body encoding [RFC2045] or MIME header encoding [RFC2047]. Any
encoding should be based on UTF-8. In the interest of
interoperability, charsets other than UTF-8 are prohibited in mail
addresses and message headers described in Section 7.1 of [RFC6530].

5.4. Error Handling

If a message is rejected by the MSA with a response code that
indicates incompatibility with legacy email described in Section 3.2
of [RFC6531], the compose window should be kept open so that the user
can make changes and retry. The email client should provide guidance
to the user about switching the Message Mode, reconfiguring the type
of an address in the address book or adding an ASCII legacy address
for a contact in the address book.

Similarly, if a message bounces, the email client could parse the
delivery status notifications and message disposition notifications
[RFC6533] to determine if the failure was a compatibility problem and

if so, which addresses caused the problem.

5.5. Examples

The following examples illustrate most of the different possible cases.

Suppose the user (Sender) has set up the following email account containing two email addresses, an Internationalized address and an ASCII address on an Internationalized server.

Sender: I0, IA0

Examples are not provided for the following cases:

a) Sender: I0, LA0

If the Sender has both Internationalized and Legacy addresses, then this is equivalent to the above.

b) Sender: I0

If the Sender has only Internationalized addresses, then it cannot send Legacy messages. The email client cannot provide an option to switch the Message Mode to Legacy.

c) Sender: LA0

If the Sender has only accounts on Legacy servers, then it cannot send Internationalized messages. The email client cannot provide an option to switch the Message Mode to Internationalized.

The address book has the following contacts with email addresses.

Contact1: I1, IA1
Contact2: I2
Contact3: IA3
Contact4: LA4

Example 1:

From: Sender
To: Contact1
CC: Contact2

This message can be sent in Internationalized mode.

In Legacy mode the email client would flag Contact2, who does not

have an ASCII address.

Example 2:

From: Sender
To: Contact1
CC: Contact3

This message can be sent in either Internationalized or Legacy mode.

Example 3:

From: Sender
To: Contact1
CC: Contact4

This message cannot be sent in Internationalized mode. Contact4
would be flagged since it is not on an Internationalized server.

This message can be sent in Legacy mode.

Example 4:

From: Sender
To: Contact2
CC: Contact3

This message can be sent in either Internationalized mode or Legacy
mode.

Example 5:

From: Sender
To: Contact2
CC: Contact4

This message cannot be sent in either mode.

Internationalized mode would flag Contact4 which is on a Legacy
server. The user can remove Contact4 or use the send anyway option.

Legacy mode would flag Contact2 who does not have an ASCII address.
The user would have to remove Contact2 in order to send this message.

5.6. Limitations

In summary, the guidelines outlines in Section 4 and Section 5 will
provide the following compatibility solutions:

1. When there is an ASCII address for all contacts in the message,
then a single legacy compatible message can be sent to all
recipients.

2. When some contacts in the message do not have an ASCII address
and some have only ASCII addresses on legacy servers, then the
message can be split into two. One message is sent as an
Internationalized message to recipients on Internationalized servers.
The other is sent as a legacy compatible message to recipients on
legacy servers.

These guidelines have a number of limitations.

a) Unknown Address Types

Message Mode is effective only if users are fairly disciplined about
keeping addresses in their address book and configuring the type
correctly as Internationalized or Legacy.

When replying to an email, the message may have addresses that are
not in the address book. The user may also enter addresses directly
during message composition that are not in the address book.

The email client may determine by inspection that some addresses are
Internationalized. If an address contains any non-ASCII character,
then it must be Internationalized. However, an ASCII address may be
on either an Internationalized server or a Legacy server and there is
no way software can determine this automatically.

In such cases, it may be useful for the email client to flag unknown
address types in a message so that the user is not lead to believe
that the message will not bounce just because there were no
incompatibility warnings.

b) Address Removal

When email addresses are removed from a message to meet compatibility
requirements, recipients do not see everyone who was intended to be
part of the conversation. The email client can provide the address
of removed recipients by using an empty group. This technique is
described in Section 3.1.8 of [I-D.ietf-eai-popimap-downgrade].

This is not an ideal solution, since replies to the message will not
reach everyone intended. But at least it provides the necessary
contact information to recipients who may be able to use other
methods to reply to all intended.

6. **Mailbox Integration**

   If more than one email address is used for the sender user, emails
   may arrive at different email accounts. There are several ways to
   provide mailbox integration so the user is able to view all mail in
   one location, such as a single 'Inbox' folder.

   If integration is done on the server, through the use of aliases,
   then the email client does not need to do anything. All mail will be
   received at the client from one address.

   The email client should provide mailbox integration for cases where
   server side integration is not available and for more flexibility on
   the part of the user. Many email clients already provide a
   convenient way to manage multiple email accounts.

   An option to view all mail from a group of accounts in one integrated
   folder should also be provided.

7. **Character Encoding**

   Email message bodies may be composed and displayed using many
   different character encoding schemes. Numerous character encodings
   have been developed over time in order to best represent different
   language scripts. In recent years there has been a trend to prefer
   Unicode as a "universal" character set and UTF-8 as the preferred
   encoding method.

   A good general principle to follow is to minimize character
   conversions. This will reduce the chance that the received message
   is displayed differently from how it was composed. Displaying
   received mail SHOULD use the character encoding of the received mail.

   Since older MUAs may not be able to parse UTF-8, the MUA SHOULD try
   to reply to mail using the character encoding of the received mail.
   This may not be possible if the sender adds new characters that
   cannot be encoded in the original encoding. For example, if the
   received message is encoded in ISO-2022-JP and characters in ISO-
   8859-1 are added to the message, the text cannot be carried in ISO-
   2022-JP and conversion to UTF-8 may be the best solution.

   For new mail, A SMTPUTF8 compliant MUA SHOULD use UTF-8 as the
   default encoding if the message type is global or if the envelope
   contains non-ASCII addresses. If email clients utilize this default,
   character conversions will be minimized and there will be less chance
   that someone will receive mail in an unrecognized encoding.

   If the message type is rfc822, other considerations may apply, such

as using the system locale/language.

Notwithstanding the above, there may be cases where the default does
not work well. There SHOULD be options for the user to reset the
default character encoding. There SHOULD also be options to change
the encoding when reading or writing individual email messages.

8. Normalization

Different sequences of UTF-8 characters may represent the same thing.
Normalization is a process that converts all canonically equivalent
sequences to a single unique form.

Normalization of email headers is specified in Section 3.1 of
 [RFC6532]. The MUA SHOULD normalize all email addresses in the
envelope and message headers.

For message bodies that contain UTF-8 characters (message/global),
the "Net-Unicode" standardized text transmission format specified in
[RFC5198] SHOULD be followed. It covers both normalization and
control characters that may affect display of text.

If the MUA saves email addresses (such as in an address book), they
SHOULD be stored in normalized form.

Other normalizations may be needed in specific language environments.
For example, in the Japanese environment, special considerations are
needed for the "@" and "." symbols. Most Japanese input methods
convert "@" to FULLWIDTH COMMERCIAL AT (U+FF20) and "." to either
IDEOGRAPHIC FULL STOP (U+3002) or FILLWIDTH FULL STOP (U+FF0E). In
email addresses, "@" is needed to separate the local name from the
domain name and "." to separate domain name labels. Normalization is
necessary to replace FULLWIDTH COMMERCIAL AT (U+FF20) with ASCII "@",
IDEOGRAPHIC FULL STOP (U+3002) with ASCII "." and FILLWIDTH FULL STOP
(U+FF0E) with ASCII ".".

9. Security Considerations

This document does not introduce any security considerations beyond
those already covered by the normative references for Email Address
Internationalization (EAI).

10. IANA Considerations

IANA changes are covered by the normative references for Email
Address Internationalization (EAI).

11. Acknowledgments

12. Normative References

[ANSI.X3-4.1968]   American National Standards
                            Institute, "USA Code for
                            Information Interchange",
                            ANSI X3.4, 1968.

[I-D.ietf-eai-5738bis]   Resnick, P., Newman, C., and S.
                            Shen, "IMAP Support for UTF-8",
                            draft-ietf-eai-5738bis-09 (work in
                            progress), August 2012.

[I-D.ietf-eai-popimap-downgrade] Fujiwara, K., "Post-delivery
                            Message Downgrading for
                            Internationalized Email Messages",
                            draft-ietf-eai-popimap-downgrade-07
                            (work in progress), August 2012.

[I-D.ietf-eai-rfc5721bis]   Gellens, R., Newman, C., Yao, J.,
                            and K. Fujiwara, "POP3 Support for
                            UTF-8",
                            draft-ietf-eai-rfc5721bis-07 (work
                            in progress), July 2012.

[RFC2045]   Freed, N. and N. Borenstein,
                            "Multipurpose Internet Mail
                            Extensions (MIME) Part One: Format
                            of Internet Message Bodies",
                            RFC 2045, November 1996.

[RFC2047]   Moore, K., "MIME (Multipurpose
                            Internet Mail Extensions) Part
                            Three: Message Header Extensions
                            for Non-ASCII Text", RFC 2047,
                            November 1996.

[RFC2119]   Bradner, S., "Key words for use in
                            RFCs to Indicate Requirement
                            Levels", BCP 14, RFC 2119,
                            March 1997.

[RFC3629]   Yergeau, F., "UTF-8, a
                            transformation format of ISO
                            10646", STD 63, RFC 3629,
                            November 2003.

[RFC5068]   Hutzler, C., Crocker, D., Resnick,
                        P., Allman, E., and T. Finch,
                        "Email Submission Operations:
                        Access and Accountability
                        Requirements", BCP 134, RFC 5068,
                        November 2007.

[RFC5198]   Klensin, J. and M. Padlipsky,
                        "Unicode Format for Network
                        Interchange", RFC 5198, March 2008.

[RFC5321]   Klensin, J., "Simple Mail Transfer
                        Protocol", RFC 5321, October 2008.

[RFC5322]   Resnick, P., Ed., "Internet Message
                        Format", RFC 5322, October 2008.

[RFC5598]   Crocker, D., "Internet Mail
                        Architecture", RFC 5598, July 2009.

[RFC5890]   Klensin, J., "Internationalized
                        Domain Names for Applications
                        (IDNA): Definitions and Document
                        Framework", RFC 5890, August 2010.

[RFC6409]   Gellens, R. and J. Klensin,
                        "Message Submission for Mail",
                        STD 72, RFC 6409, November 2011.

[RFC6530]   Klensin, J. and Y. Ko, "Overview
                        and Framework for Internationalized
                        Email", RFC 6530, February 2012.

[RFC6531]   Yao, J. and W. Mao, "SMTP Extension
                        for Internationalized Email",
                        RFC 6531, February 2012.

[RFC6532]   Yang, A., Steele, S., and N. Freed,
                        "Internationalized Email Headers",
                        RFC 6532, February 2012.

[RFC6533]   Hansen, T., Newman, C., and A.
                        Melnikov, "Internationalized
                        Delivery Status and Disposition
                        Notifications", RFC 6533,
                        February 2012.

Authors' Addresses

    Ernie Dainow
    Afilias Canada
    4141 Yonge Street
    Toronto, Ontario M2P 2A8
    Canada

    Phone:
    EMail: edainow@afilias.info

    Kazunori Fujiwara
    Japan Registry Services Co., Ltd.
    Chiyoda First Bldg. East 13F, 3-8-1 Nishi-Kanda
    Chiyoda-ku, Tokyo 101-0065
    Japan

    Phone: +81 3 5215 8451
    EMail: fujiwara@jprs.co.jp