1. Abstract

MIME [RFC MIME-IMB, RFC MIME-IMT, RFC MIME-HEADERS] and various other modern Internet protocols are capable of using many different character sets. This in turn means that the ability to label different character sets is essential. This registration procedure exists solely to associate a specific name or names with a given character set and to give an indication of whether or not a given character set can be used in MIME text objects. In particular, the general applicability and appropriateness of a given registered character set is a protocol issue, not a registration issue, and is not dealt with by this registration procedure.
2. Definition of a Character Set

The term "character set" is used here to refer to a method of converting a sequence of octets into a sequence of characters. Note that unconditional and unambiguous conversion in the other direction is not required, in that not all characters may be representable by a given character set and a character set may provide more than one sequence of octets to represent a particular sequence of characters.

This definition is intended to allow various kinds of character encodings, from simple single-table mappings such as US-ASCII to complex table switching methods such as those that use ISO 2022's techniques, to be used as character sets. However, the definition associated with a character set name must fully specify the mapping to be performed. In particular, use of external profiling information to determine the exact mapping is not permitted.

NOTE: The term "character set" was originally to describe such straightforward schemes as US-ASCII and ISO-8859-1 which have a simple one-to-one mapping from single octets to single characters. Multi-octet coded character sets and switching techniques make the situation more complex. For example, some communities use the term "character encoding" for what this document calls a "character set", while using the phrase "coded character set" to denote an abstract mapping from integers (not octets) to characters.

3. Registration Requirements

Registered character sets are expected to conform to a number of requirements as described below.

3.1. Required Characteristics

Registered character sets must conform to the definition of a "character set" given above. In addition, character sets intended for use in MIME content types under the "text" top-level type must conform to the restrictions on that type
described in RFC MIME-IMB. All registered character sets must note whether or not they are suitable for such usage.

3.2. New Character Sets

This registration mechanism is not intended to be a vehicle for the definition of entirely new character sets. This is due to the fact that the registration process does NOT contain adequate review mechanisms for such undertakings.

As such, only character sets defined by other processes and standards bodies, or specific profiles of such character sets, are eligible for registration.

3.3. Naming Requirements

One or more names must be assigned to all registered character sets. Multiple names for the same character set are permitted, but if multiple names are assigned a single primary name for the character set must be identified. All other names are considered to be aliases for the primary name and use of the primary name is preferred over use of any of the aliases.

Each assigned name must uniquely identify a single character set. All character set names must be suitable for use as the value of a MIME content type charset parameter and hence must conform to MIME parameter value syntax. This applies even if the specific character set being registered is not suitable for use with "text".

3.4. Usage and Implementation Requirements

Use of a large number of character sets in a given protocol
may hamper interoperability. However, the use of a large number of undocumented and/or unlabelled character sets hampers interoperability even more.

A character set should therefore be registered ONLY if it adds significant functionality that is valuable to a large community, OR if it documents existing practice in a large community. Note that character sets registered for the second reason should be explicitly marked as being of limited or specialized use and should only be used in Internet messages with prior bilateral agreement.

3.5. Publication Requirements

Character set registrations can be published in RFCs, however, RFC publication is not required to register a new character set.

The registration of a character set does not imply endorsement, approval, or recommendation by the IANA, IESG, or IETF, or even certification that the specification is adequate. It is expected that applicability statements for particular applications will be published from time to time that recommend implementation of, and support for, character sets that have proven particularly useful in those contexts.

4. Registration Procedure

The following procedure has been implemented by the IANA for review and approval of new character sets. This is not a formal standards process, but rather an administrative procedure intended to allow community comment and sanity checking without excessive time delay.

4.1. Present the Character Set to the Community
Send the proposed character set registration to the "ietf-charsets@innosoft.com" mailing list. This mailing list has been established for the sole purpose of reviewing proposed character set registrations. Proposed character sets are not formally registered and must not be used; the "x-" prefix specified in RFC MIME-IMB can be used until registration is complete.

The intent of the public posting is to solicit comments and feedback on the definition of the character set and the name chosen for it over a two week period.

4.2. Character Set Reviewer

When the two week period has passed and the registration proposer is convinced that consensus has been achieved, the registration application should be submitted to IANA and the Character Set Reviewer. The character set reviewer, who is appointed by the IETF Applications Area Director(s), either approves the request for registration or rejects it. Rejection may occur because of significant objections raised on the list or objections raised externally. If the character set reviewer considers the registration sufficiently important and controversial, a last call for comments may be issued to the full IETF. The character set reviewer may also recommend standards track processing (before or after registration) when that appears appropriate and the level of specification of the character set is adequate.

Decisions made by the reviewer must be posted to the ietf-charsets mailing list within 14 days. Decisions made by the reviewer may be appealed to the IESG.

4.3. IANA Registration
Provided that the character set registration has either passed review or has been successfully appealed to the IESG, the IANA will register the character set and make its registration available to the community.

5. Location of Registered Character Set List

Character set registrations will be posted in the anonymous FTP file "ftp://ftp.isi.edu/in-notes/iana/assignment/character-sets/" and all registered character sets will be listed in the periodically issued "Assigned Numbers" RFC [currently RFC-1700]. The description of the character set may also be published as an Informational RFC by sending it to "rfc-editor@isi.edu" (please follow the instructions to RFC authors [RFC-1543]).

Expires October 1996            [Page 5]

Internet Draft    Character Set Registration    April 1996

6. Registration Template

To: ietf-charsets@innosoft.com
Subject: Registration of new character set

Character set name(s):

(All names must be suitable for use as the value of a MIME content-type parameter.)

Published specification(s):

(A specification for the character set must be openly available that accurately describes what is being registered.)

Person & email address to contact for further information:
7. Security Considerations

This registration procedure is not known to raise any sort of security considerations that are appreciably different from those already existing in the protocols that employ registered character sets.

8. References

[RFC-1521]

[RFC-1522]
Moore, K., "Representation of Non-ASCII Text in Internet Message Headers", RFC 1522, University of Tennessee, September 1993.
[RFC-1590]  

[RFC-1700]  

[RFC-MIME-IMB]  

[RFC-MIME-IMT]  

[RFC-MIME-HEADERS]  
Moore, K., "Multipurpose Internet Mail Extensions (MIME) Part Three: Representation of Non-Ascii Text in Internet Message Headers", RFC MIME-HEADERS, University of Tennessee, ?.

Expires October 1996

Internet Draft Character Set Registration April 1996

[US-ASCII]  
Coded Character Set -- 7-Bit American Standard Code for Information Interchange, ANSI X3.4-1986.
9. Author Address

Ned Freed
Innosoft International, Inc.
1050 East Garvey Avenue South
West Covina, CA 91790
USA
tel: +1 818 919 3600           fax: +1 818 919 3614
e-mail: ned@innosoft.com

Jon Postel
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
USA
tel: +1 310 822 1511           fax: +1 310 823 6714
e-mail: Postel@ISI.EDU
VERY IMPORTANT NOTE: This appendix is intended to communicate various editorial and procedural tasks the IANA and the RFC Editor should undertake prior to publication of this document as an RFC. This appendix should NOT appear in the actual RFC version of this document!

This document refers to the media types mailing list ietf-charsets@innosoft.com. There is no guarantee that innosoft.com will continue to be able to accommodate this list throughout the lifetime of this document. As such, this reference should be replaced by an address of the general form ietf-charsets@iana.org. The actual list can then either be moved to this location or forwarders can be installed to redirect traffic to the host that currently maintains the list.